Factor Zoo Meets the News Library

Antoine Didisheim, Bryan Kelly, Mo Pourmohammadi, Hanqing Tian*

November 8, 2026

Abstract

Financial news contains the information behind accounting ratios and price trends, alongside a much broader set of firm developments. To show this, we propose a new AI-based method for constructing interpretable news-based firm characteristics that capture intuitive economic concepts and connect to macroeconomic conditions. Factors built from these characteristics span the traditional factor zoo and expand the efficient frontier, with much of the spanning power concentrated in three of thirteen news themes. Beyond factor construction, we show that accounting for each article's loadings on multiple topics reveals predominantly market underreaction.

*Antoine Didisheim is at University of Melbourne. Bryan Kelly is at Yale School of Management, AQR Capital Management, and NBER. Mo Pourmohammadi is at Yale School of Management. Hanqing Tian is at University of Melbourne. AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR.

"in the social sciences often that is treated as important which happens to be accessible to measurement."

F. A. Hayek, *The Pretence of Knowledge*, Nobel Lecture (1974)

1 Introduction

Academic research and investment practice differ in the information they can use: empirical research requires systematic measurement, while practitioners can also act on information that resists quantification. This contrast is particularly pronounced in asset pricing. Over the past fifty years, researchers have assembled a large catalog of firm characteristics, drawn predominantly from accounting and price data, that underpins the *factor zoo* (Harvey et al., 2016; Jensen et al., 2022, hereafter JKP). These characteristics provide useful representations of firms, but accounting measures arrive with a lag, and both accounting and price variables offer a limited description of firms' circumstances and prospects. Their numerical structure nevertheless makes them readily comparable across firms and over time, helping them occupy a central place in empirical research. Practitioners, by contrast, supplement these measures with news, company reports, and other textual sources that convey timely information on a much broader range of developments.

This gap between investment practice and academic research directly reflects the distinction between structured and unstructured data. Early methods, such as bag-of-words models and financial dictionaries, quantified text but largely captured word counts and sentiment rather than context and nuance (Tetlock, 2007; Loughran and McDonald, 2011). Recent advances in AI enable large language models to capture this richer meaning in numerical representations, with an ability to process financial text that can match or even exceed that of practitioners. These advances give researchers an opportunity to bridge the gap between what academics measure and what investors read.

Our earlier work takes an important step in this direction (Didisheim et al., 2026). We use language-model embeddings of financial news to construct asset pricing factors that dwarf those in the *factor zoo*. In this paper, we develop a novel approach that transforms firm-level news into interpretable firm loadings on a small set of news topics. These loadings allow us to build a set of interpretable news-based factors that connects differences in returns across firms to important developments in the news and the macroeconomy. Interpretability comes without sacrificing predictive power. Combined into a portfolio, our factors deliver higher

Sharpe ratios with lower turnover than characteristic-based factors. They capture substantial variation in the traditional *factor zoo*, while the zoo captures much less of theirs. The news themes track broader macroeconomic developments and reveal systematic differences in how investors underreact or overreact to different topics.

To build our factors, we start from embeddings of individual news articles about firms produced by a large language model (Chen et al., 2024; Didisheim et al., 2026). Each embedding is a numerical vector that summarizes the model's internal representation of the text.[1] Our goal is to map these coordinates into a smaller set of interpretable firm exposures. We draw on sparse autoencoders (SAEs), developed in the computer-science literature on mechanistic interpretability, to construct this representation (Cunningham et al., 2023; Gao et al., 2024). An SAE is a small auxiliary network trained to reconstruct an embedding through an intermediate representation in which only a few features activate for each text. These SAEs are trained to minimize reconstruction error, with an L_1 penalty encouraging sparsity in the lower-dimensional representation.

Chen et al. (2026a) show how these tools can reveal the concepts language models use to process financial text. However, consistent with conventional practice in the SAE literature, their general-purpose SAE expands the representation to a larger feature space.[2] Our objective is more focused: we seek a small set of interpretable firm exposures that retains economically relevant information. We therefore train our own SAE directly on financial news to compress the representation into a compact set of economic themes.

Training our own SAE has several advantages. First, we can pair it with a point-in-time language model whose training ends before the evaluation period, eliminating data leakage from subsequent information in the LLM's training corpus (Sarkar and Vafa, 2024; Didisheim et al., 2025).[3] Second, we can control both the sparsity and the dimension of the resulting representation. These choices create a tradeoff between interpretability and predictive performance: greater sparsity and fewer dimensions constrain how much of the original information the SAE can retain, but yield a smaller set of more distinct concepts. To identify the best balance, we develop four measures of feature interpretability and evaluate them alongside performance. *Label–headline alignment* measures whether a feature's label

[1] The dimension depends on the LLM architecture and generally increases with model size, ranging from roughly 1,000 coordinates for small models to 16,384 for Llama 3.1 405B. In our application, the embedding has 4,096 coordinates.

[2] The general-purpose SAE used by Chen et al. (2026a) produces approximately 131,000 features.

[3] The portfolio evaluation covers signal months January 2016 to November 2025, with returns realized in the following month. Throughout, portfolio sample dates refer to signal months. Both the language model and the SAE are trained exclusively on data preceding January 2016.

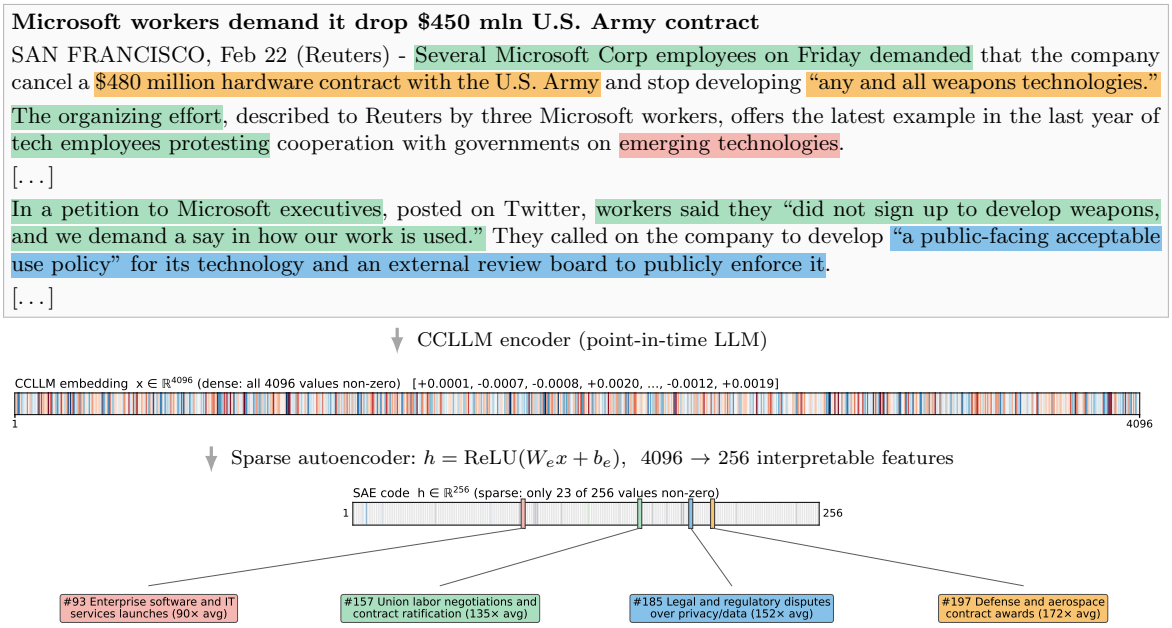


Figure 1: From a News Article to Interpretable News Features

Note. A Reuters article about Microsoft, its dense LLM embedding, and its sparse SAE representation. Callouts identify four strongly activated topics. Parentheses express activations relative to each feature's article-level average. Highlights are added by the authors.

faithfully describes the news that activates it. *Distinctiveness* measures whether different features capture different concepts rather than duplicate one another. *Label stability* assesses whether a feature retains its meaning between the training and evaluation periods. Finally, the *number of active features per article* measures how concisely the representation expresses each news item.

Figure 1 illustrates the benefits of the approach with an article in our sample about Microsoft employees opposing a U.S. Army contract, its dense LLM representation, and its sparse SAE representation using our methodology. Our SAE transforms the dense embedding into 23 nonzero features out of 256, including the clearly labeled main topics. One article thus gives the firm several distinct, interpretable exposures that capture different aspects of the same event.

Combined, our dense news embeddings deliver annualized out-of-sample Sharpe ratios of 1.68 for the value-weighted portfolio and 2.31 for the equally weighted portfolio. An SAE with only 256 dimensions retains all of this performance: at a moderate sparsity penalty, the corresponding Sharpe ratios reach 1.74 and 2.41. Increasing the penalty improves label–

headline alignment, distinctiveness, and label stability while reducing the number of features active in each article. Beyond this range, excessive penalization eliminates most features, reducing performance and leaving too few active concepts for the interpretability measures to remain informative.

We label each SAE coordinate using an LLM to identify the common topic among the 20 news items that activate it most strongly. We retain coordinates that capture economic content, excluding those that identify individual firms, reflect textual patterns, or lack a clear interpretation. This yields 228 interpretable news topics, each associated with a tradable factor. We manually group these topics into 13 broader themes based on their labels: earnings results; guidance and outlook; analyst actions and estimates; credit ratings and debt financing; equity issuance and payouts; management and governance; mergers, acquisitions and restructuring; labor and workforce; legal, regulatory and misconduct news; healthcare and biotech approvals and trials; energy, utilities and commodities; markets, the financial sector and the macroeconomy; and industry operations and products.

The factors match economic intuition at both the theme and individual-factor levels. Theme coverage tracks familiar developments: *guidance* news falls during the pandemic as firms withdraw forecasts, while *credit and debt financing* rises amid downgrades and emergency borrowing. The portfolio's Sharpe ratio draws on this breadth of economic content: a Shapley decomposition identifies *management*, *earnings*, and *analyst news* as the largest contributors under equal weighting, while *M&A*, *analyst news*, and *industry operations* lead under value weighting. Individual factors also behave as their labels suggest. For example, the *oil and gas drilling* factor has a monthly correlation of 0.29 with oil-price changes, while the *investment bank and capital markets* factor correlates 0.64 with intermediary capital risk.

Our broader claim is that text offers a more natural and informative basis for factors than traditional characteristics. This requires news-based factors to (i) add information beyond traditional characteristics and (ii) span the information captured by traditional factors. Our asset-pricing results support both claims. A stochastic discount factor (SDF) constructed from news expands the efficient frontier relative to JKP factors: its portfolio achieves out-of-sample Sharpe ratios of 2.35 under equal weighting and 1.83 under value weighting, compared with 1.54 and 1.12 for JKP. Even an infeasible combination chosen with hindsight improves on news alone by only 6% and 4%, respectively.

The news SDF also produces lower pricing errors for both sets of factors. It prices traditional factors better overall, with lower errors in eight of the thirteen JKP themes

under equal weighting and six under value weighting. The reverse comparison is more one-sided: news prices twelve of its thirteen themes better under both weightings. The JKP SDF's only advantage is in *equity issuance and payouts*, where the difference is modest.

These gains come with substantially lower turnover. Monthly turnover of the news portfolio is 20.8% versus 29.6% for JKP under equal weighting, and 15.9% versus 26.7% under value weighting. Across individual factors, mean monthly turnover ranges from 7.3% to 34.8% for news versus 6.9% to 67.6% for JKP under equal weighting, and from 4.7% to 34.0% versus 4.9% to 84.7% under value weighting. Text-based exposures capture persistent firm narratives, offering a more stable basis for factor positions than characteristics that change mechanically with prices.

Much of the explanatory power comes from three intuitive news themes: *earnings*, *analyst actions*, and *guidance*, which describe the information underlying accounting measures and momentum. JKP factors require a broader mix of themes to explain a smaller share of news-factor variation. Other news themes, including *M&A* and *industry operations*, contribute substantially to portfolio performance while explaining relatively little of JKP returns.

Finally, our results suggest that some evidence of overreaction in previous research may reflect measurement error in assigning news to topics. Kwon and Tang (2025) document both underreaction and overreaction, with positive and negative response slopes in 13 and 11 categories, respectively, while assigning each news item to a single topic. Figure 1 illustrates the limitation: financial articles tend to concern several relevant economic concepts. Our SAE measures continuous exposures to these concepts, allowing each article to contribute to several topics with different weights.

We estimate a separate weighted least squares regression for each topic, relating subsequent returns to the news-day return and weighting observations by their relative SAE activation. This lets the same news contribute to several topics according to its relevance. Underreaction predominates: 90 of the 228 factors exhibit statistically significant continuation, compared with only 13 showing significant reversal.

The delayed response is economically meaningful. For *earnings news*, returns over the following 120 trading days extend the initial reaction by 9.3%, so 8.5% of the combined response occurs after the news day. Significant continuation also extends the initial reaction by 7.7% for *guidance*, 7.8% for *analyst news*, 6.6% for *credit and debt financing*, and 7.5% for *equity issuance and payouts* over the same horizon. These patterns show that a meaningful part of the market response unfolds over several months across distinct kinds of financial news.

We examine four behavioral channels behind these differences. In line with Kwon and Tang (2025), more extreme news exhibits less underreaction. Extremeness alone explains 33% of the variation in response slopes, and a one-standard-deviation increase reduces continuation by slightly more than its average level. News about heavily covered firms also exhibits less continuation, consistent with greater investor attention amplifying the initial response (Hou et al., 2025). Quantitative content predicts more underreaction, consistent with evidence that statistics are remembered less durably than stories (Graeber et al., 2024). Finally, harder-to-read text predicts greater continuation, consistent with processing difficulty delaying the market response (You and Zhang, 2009).

Related Literature. Our paper contributes to research using text to study asset prices, surveyed by Gentzkow et al. (2019), Loughran and McDonald (2020), and Hoberg and Manela (2025). Early work extracts sentiment and fundamental information from word counts and financial dictionaries (Tetlock, 2007; Tetlock et al., 2008; Loughran and McDonald, 2011). More recent studies use general news, rather than firm-level news, to construct interpretable economic topics and systematic risk factors (Bybee et al., 2023, 2024), while large language models provide richer return-predictive representations (Lopez-Lira and Tang, 2024; Chen et al., 2024, 2026b). Our earlier work shows that removing the component of news predictable from stock characteristics greatly strengthens return predictability (Didisheim et al., 2026). Our contribution is to add interpretability through a new method for constructing news-based firm characteristics that span and expand the traditional set of firm characteristics.

We also contribute to research on the traditional factor zoo and its aggregation into pricing models. Comprehensive studies document the breadth, replicability, and implementation costs of characteristic-based factors (Harvey et al., 2016; Hou et al., 2020; Chen and Zimmermann, 2022; Jensen et al., 2022; Novy-Marx and Velikov, 2016). Flexible models combine this information to improve return prediction and SDF estimation (Gu et al., 2020, 2021; Kelly et al., 2024). Related work studies how shrinkage shapes portfolio efficiency in high-dimensional settings (Kelly et al., 2026a; Ke and Pourmohammadi, 2026). We study a different source of firm characteristics: interpretable concepts extracted from news. Our comparison examines what these exposures add to the efficient frontier, how well the resulting SDF prices traditional factors, and which economic themes explain the overlap.

Methodologically, we draw on mechanistic interpretability. SAEs separate concepts that overlap in dense language-model representations (Elhage et al., 2022; Cunningham et al., 2023; Bricken et al., 2023; Gao et al., 2024). Chen et al. (2026a) apply these tools to identify

financial concepts, interpret model predictions, and steer model behavior. To our knowledge, we are the first to train an SAE on embeddings from a point-in-time LLM. We also introduce a small sparse layer that compresses the representation for our financial application, rather than expanding it into a much larger feature space.

Finally, we contribute to research on market reactions to information and their behavioral foundations (Daniel et al., 1998; Hong and Stein, 1999). Kwon and Tang (2025) connect differences in underreaction and overreaction across news categories to the extremeness of their outcomes. Our continuous exposures accommodate several concepts within an article and let us measure topic-specific reactions without assigning the article to a single category. We relate the resulting differences in continuation to extremeness, investor attention, memory for quantitative information, and textual processing difficulty (Hou et al., 2025; Graeber et al., 2024; Hong, 2026; You and Zhang, 2009).

2 Sparse Autoencoders for Financial News

LLM embeddings are dense and polysemantic: each concept is spread across many coordinates and each coordinate mixes many concepts ("superposition," Elhage et al., 2022). The news portfolio of Didisheim et al. (2026) is built directly from these raw embeddings: it combines the return predictability of all embedding coordinates (4,096 in this case), but the underlying news-managed portfolios have no clear economic interpretation. An SAE (Cunningham et al., 2023; Bricken et al., 2023) maps each dense embedding into a sparse representation of interpretable features, only a few of which activate for any given article. When a feature fires on a coherent set of articles, it can be given a label, and the news-managed portfolio built from it becomes an interpretable news factor. Appendix A details the complete pipeline, from article embeddings and SAE training to feature labeling and tradable news factors.

2.1 From News Articles to Interpretable Factors

Figure 2 illustrates the problem and our approach. We begin with a news article, which may discuss several subjects at once. The LLM encoder turns the text into a numerical embedding that captures these subjects and their context. But the embedding does not assign each subject its own coordinate. Instead, each concept can contribute to many coordinates, and each coordinate can combine several concepts. The information is present, but its organization does not tell us which economic developments a firm is exposed to.

Compression alone does not solve this problem. Principal component analysis (PCA) or an ordinary autoencoder can summarize the embedding with fewer coordinates while retaining much of its information. We obtain a smaller vector without necessarily obtaining a clearer economic interpretation.

The SAE, by contrast, learns to reconstruct the original embedding from a representation in which most features are zero for each article. Its reconstruction objective encourages it to retain information, while the sparsity penalty discourages it from spreading that information across many active features. The goal is for a passage about, for example, *quarterly results* to activate an *earnings* feature rather than contribute to a diffuse pattern across the vector. An article can still activate several features when it covers several subjects. Sparsity therefore does not assign each article to a single topic. It encourages a concise representation in terms of key relevant concepts.

These article-level features are the inputs to our asset-pricing analysis. We aggregate them across the news about each firm to measure its exposure to each topic. We then ask whether a compact set of interpretable features preserves the relevant economic information.

2.2 News Data and Embeddings

We train our own SAEs directly on the embeddings of financial news articles. Our news data are articles from the Thomson Reuters Real-time News Feed tagged as associated with a single U.S. stock, over the period January 1996 to December 2025. We embed each article with the point-in-time language model of Kelly et al. (2026b), a 4-billion-parameter model trained only on text available up to a given date. Following Chen et al. (2024) and Didisheim et al. (2026), we define an article's embedding as the equally weighted average of its token embeddings, a $D = 4{,}096$-dimensional vector.[4]

We use the model's 2015 checkpoint, whose training data end in November 2015, for the entire sample. The encoder therefore cannot have seen any text, event or return from our out-of-sample period, which starts in January 2016.[5] Using a point-in-time model comes at little cost in performance. Didisheim et al. (2026) show that the news portfolio built

[4]In computer science, SAEs are typically trained on token-level activations to understand how a language model processes text one token at a time. Our interest is in the economic content of an entire news article, so we train on article-level embeddings to obtain interpretable exposures at the document level.

[5]This is what fixes the start of the out-of-sample period. A language model frozen at a given date can only be evaluated on news published after that date; otherwise the embedding of an article may already reflect information that an investor did not have at the time. Such look-ahead enters the signal itself, and no out-of-sample split on returns can remove it.

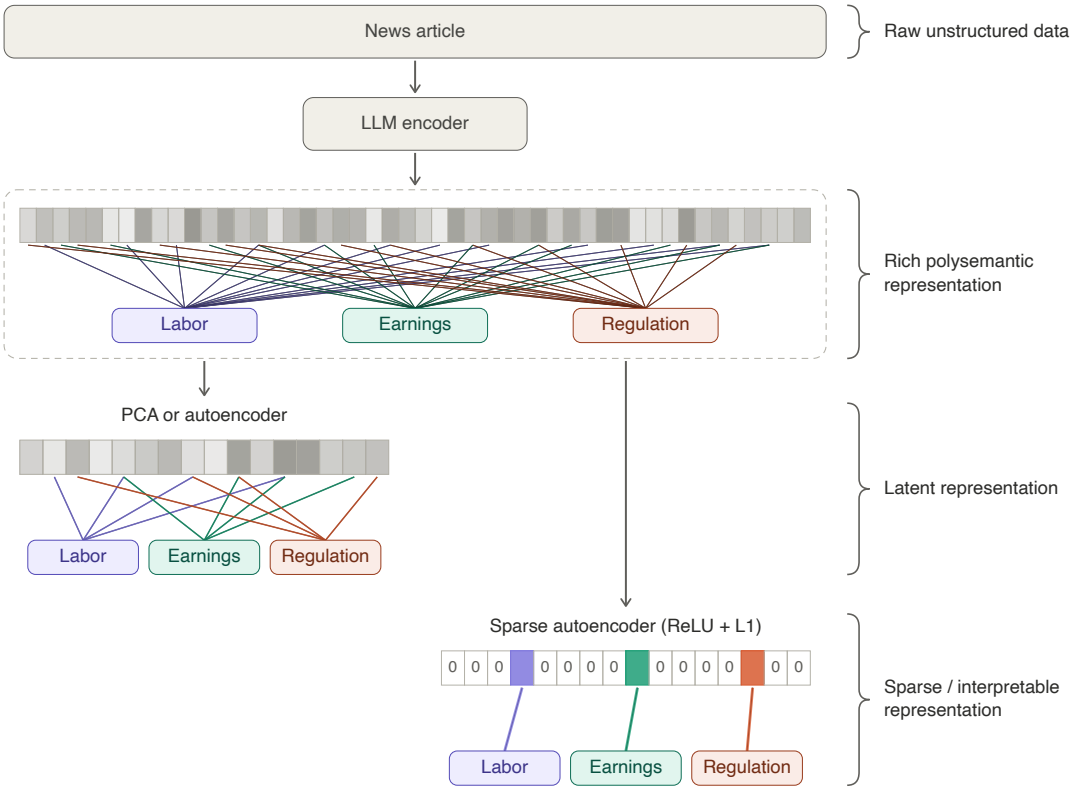


Figure 2: From Dense News Embeddings to Sparse Economic Features

Note. This figure illustrates how a news article becomes a numerical representation. In the dense LLM embedding, concepts such as labor, earnings, and regulation contribute to overlapping coordinates. The left branch illustrates compression by PCA or an ordinary autoencoder, which need not separate these concepts. The right branch illustrates an SAE with ReLU activations and an L_1 penalty, which encourages a representation with few active features. Colors and connections are schematic illustrations of the intended concept separation, not estimated model weights or a guarantee that every learned feature captures a single concept.

from this checkpoint recovers about 96% of the out-of-sample Sharpe ratio earned with the commercial *E5-Mistral-7B* embeddings of their main analysis over the same months.[6]

This section describes how our SAEs differ from the standard approach, the two design choices that govern them, and how we measure whether an SAE does a good job.

[6] Didisheim et al. (2026) show that the first two point-in-time checkpoints have too few training tokens for a model of this size, motivating the use of the 2015 checkpoint as the starting point.

2.3 Compressing Rather Than Expanding

Let $e_a \in \mathbb{R}^D$ be the embedding of news article a. An SAE encodes e_a into a non-negative feature vector $\tilde{e}_a \in \mathbb{R}^H_{\geq 0}$ and linearly decodes it back:

$$\tilde{e}_a = \mathrm{ReLU}\Big(W_{\mathrm{enc}}(e_a - b_{\mathrm{dec}}) + b_{\mathrm{enc}}\Big), \qquad \hat{e}_a = W_{\mathrm{dec}}\,\tilde{e}_a + b_{\mathrm{dec}}, \tag{1}$$

with encoder and decoder weights $W_{\mathrm{enc}} \in \mathbb{R}^{H\times D}$ and $W_{\mathrm{dec}} \in \mathbb{R}^{D\times H}$, and bias vectors $b_{\mathrm{enc}} \in \mathbb{R}^H$ and $b_{\mathrm{dec}} \in \mathbb{R}^D$. Each column of W_{dec} is a dictionary vector, a direction in embedding space, and $\tilde{e}_{a,j}$ measures how strongly feature j is present in article a. By sparsity, we mean that each article's representation should be mostly zero, with only a few economically meaningful features taking nonzero values. The SAE is trained to reconstruct the embedding while encouraging this sparsity,

$$\min_{W,b} \sum_a \Big(\big\| e_a - \hat{e}_a \big\|_2^2 + \kappa \sum_{j=1}^{H} \tilde{e}_{a,j} \Big), \tag{2}$$

where the L_1 penalty κ pushes most activations to exactly zero.[7]

The standard SAE in the machine learning literature is *overcomplete*, with $H \gg D$. It is built to interpret an LLM across everything the LLM can express, so it needs a feature for every concept in human language. The SAE of Lieberum et al. (2024), for example, has 131,000 features, and the dictionaries of Templeton et al. (2024) and Gao et al. (2024) run into the millions. Expansion is the natural choice in that setting: there are far more concepts than embedding dimensions, and a large dictionary gives each of them its own coordinate.

Our focus on financial news restricts attention, by definition, to a much smaller set of topics and concepts than the full range expressible in the English language. We seek to capture recurring subjects such as earnings, guidance, financing and litigation. This narrower scope motivates SAEs with $H \leq D$, which compress the embedding into a smaller set of interpretable features suited to asset pricing. A general-purpose dictionary devotes much of its capacity to concepts outside finance: of the 131,000 features of Lieberum et al. (2024), Chen et al. (2026a) identify only 7,000 as financially relevant. Training directly on financial news lets a much smaller dictionary concentrate on the concepts relevant to our application.

Without expansion, interpretability can still come from sparsity. Even with $H \leq D$, the penalty in (2) forces each article to be described by a handful of active features, so each

[7] Because the ReLU makes every activation non-negative, the sum in (2) is the L_1 norm of $\tilde{e}_a$.

feature must specialize in a distinct subject. Without the penalty, such an SAE essentially solves the same problem as principal component analysis: it finds the best H-dimensional reconstruction of the embedding, and its features are as dense and as hard to read as the principal components. It is the penalty that moves the features away from these dense directions and makes them interpretable. Appendix B gives the details of the training.

2.4 From SAE Features to News Portfolios

We turn SAE features into tradeable portfolios following the construction of Didisheim et al. (2026), with the SAE features in place of the dense embedding. The SAE is trained on articles published before 2016 and then held fixed. Each article a is passed through the SAE to obtain its features $\tilde{e}_a$ from equation (1). Following Didisheim et al. (2026), we define $\tilde{E}_{i,t}$ as the average of these article embeddings for firm i over the last six months. This window cuts turnover by about three quarters relative to a one-month window and raises performance after costs (Appendix C, Table 5). Each coordinate of $\tilde{E}_{i,t}$ is then Z-scored with a mean and standard deviation estimated on an expanding window through month $t-1$.

Next, we orthogonalize the embedding with respect to stock characteristics. Let $\tilde{E}_t$ be the $N_t \times H$ matrix stacking the individual embeddings $\tilde{E}_{i,t}$ of the N_t stocks with news in month t, and S_t the $N_t \times L$ matrix of their characteristics, including a constant. Each month we run the cross-sectional regression

$$\tilde{E}_t = S_t \beta_t + \tilde{\varepsilon}_t \tag{3}$$

and keep the residual $\tilde{\varepsilon}_t$, the part of the news that is not predictable from S_t. We estimate (3) by ordinary least squares for equal-weighted portfolios and by weighted least squares, with weights $v_{i,t}$, for value-weighted ones. Here $v_{i,t}$ is stock i's market equity in month t, capped at the NYSE 80th percentile and normalized by the sum of capped market equities across stocks in the portfolio universe; each portfolio therefore uses the residual that is orthogonal to S_t under its own weights, which makes it exactly neutral to the characteristics. Appendix D details both regressions and their implementation. We Z-score the residuals cross-sectionally each month and retain the notation $\tilde{\varepsilon}_{ij,t}$ for convenience. Each feature is then turned into a long-short *news factor*,

$$F_{j,t+1} = \sum_i \tilde{\varepsilon}_{ij,t} R_{i,t+1} \qquad \text{or} \qquad F_{j,t+1} = \sum_i v_{i,t}\Big(\tilde{\varepsilon}_{ij,t} - \bar{\tilde{\varepsilon}}_{j,t}\Big) R_{i,t+1}, \tag{4}$$

equal-weighted and value-weighted respectively, where $R_{i,t+1}$ is the stock's excess return in month $t+1$ and $\bar{\tilde{\varepsilon}}_{j,t} = \sum_i v_{i,t}\tilde{\varepsilon}_{ij,t}$. Stacking the H news factors in F_{t+1}, we aggregate them by maximum Sharpe ratio regression (MSRR; Kelly and Xiu, 2023),

$$\min_b \sum_t \left(1 - b'F_{t+1}\right)^2 + \lambda\, b'b, \tag{5}$$

which estimates the tangency portfolio of the news factors with ridge shrinkage. We estimate $\hat{b}_t$ on an expanding window, selecting λ by leave-one-out cross-validation, and form the out-of-sample *news portfolio* $F^{\star}_{t+1} = \hat{b}_t' F_{t+1}$.

Characteristics. We orthogonalize the news only with respect to market capitalization and 25 GICS industry indicators. Orthogonalizing with respect to a richer set of characteristics, such as the 132 characteristics of JKP, would raise performance further as in Didisheim et al. (2026) (Appendix C, Figure 18). But the purpose of this paper is to compare the news factors with the characteristic *factor zoo*, and news factors built to be orthogonal to the zoo's characteristics would be uninformative about how the two relate. We therefore strip away only the minimum: size, which drives how much a firm is covered in the news, and industry, which drives the vocabulary in which it is covered.

2.5 The Width–Penalty Tradeoff

An SAE has two design choices: the number of features H, which sets how many news factors there are, and the penalty κ, which sets how many of them each article may draw on. Each choice trades off the interpretability of the features against the information they retain.

The number of features. A smaller number of features forces each feature to cover a broad, distinct subject rather than splitting one subject into many near-duplicate variants. This helps produce a news factor library in which every factor is distinct and can be interpreted. Too few features, however, cannot represent everything in the news that matters for returns and may merge concepts that are better kept apart.

The penalty. A stronger penalty encourages fewer active features per article, promoting specialization in recognizable concepts and greater coherence among the articles that activate each feature. An excessive penalty, however, can prevent features from activating altogether, causing the sparse representation to lose information contained in the original news.

Neither choice therefore has a direction that is uniformly better. We seek a balance

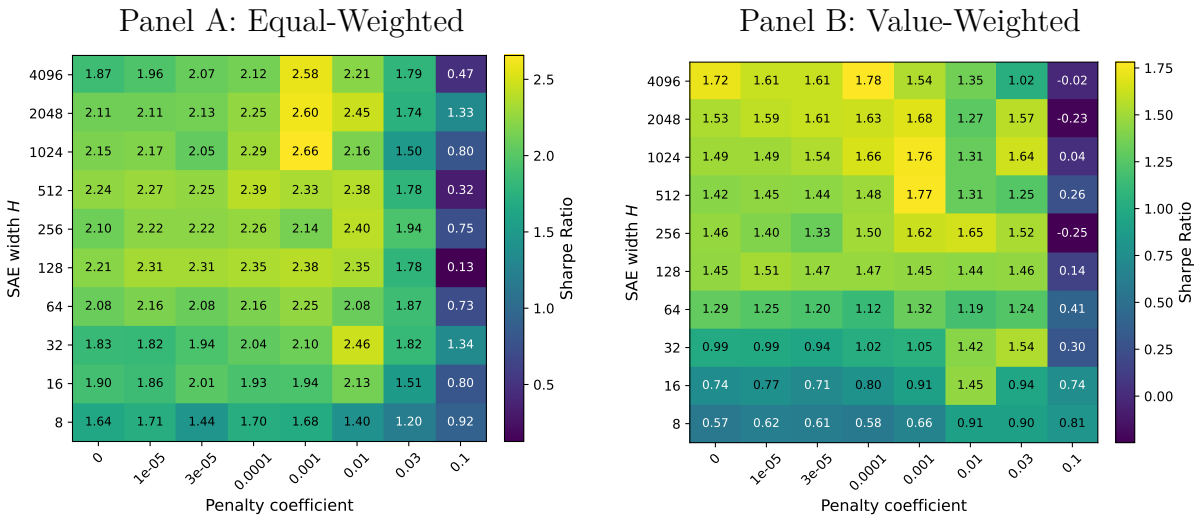


Figure 3: News Portfolio Sharpe Ratio by SAE Width and Penalty

Note. Annualized out-of-sample Sharpe ratios of MSRR portfolios across SAE widths H and penalties κ. Each cell uses the median combined EW/VW Sharpe ratio among three runs. Panels A and B use equal and value weighting. January 2016–November 2025.

between the two: a small set of well-separated news factors that together preserve the economically relevant information in the news

Figure 3 maps this tradeoff. We train SAEs over a grid of ten widths, from $H = 8$ to $H = 4{,}096$, and eight penalty levels, from $\kappa = 0$ to $\kappa = 0.1$, and report the Sharpe ratio of the news portfolio built from each. Three patterns emerge. First, without a penalty the SAE does not improve on the dense embedding, which earns Sharpe ratios of 2.3 equal-weighted and 1.7 value-weighted with the same six-month signal. This is expected: without a penalty, an SAE with $H \leq D$ is close to principal component analysis. Second, a moderate penalty helps. Around $\kappa = 10^{-3}$ to 10^{-2} the Sharpe ratio rises to as much as 2.7 equal-weighted and 1.8 value-weighted, above the dense embedding. Sparsity therefore costs nothing in performance. Only at the strongest penalty, $\kappa = 0.1$, where most features die, does performance collapse.

Third, performance generally increases with the number of features. Taking the best penalty for each width, the equal-weighted Sharpe ratio rises from 1.7 at $H = 8$ and 2.1 at $H = 16$ to about 2.4 for widths between 32 and 512, and to 2.6–2.7 for the widest SAEs. The value-weighted Sharpe ratio rises from 0.9 at $H = 8$ to about 1.5 for widths between 16 and 128, and to 1.7–1.8 from $H = 256$ up.

The main takeaway from Figure 3 is that a moderate penalty costs nothing in performance, and that most of the gains from increasing the number of features are realized with a

few hundred features. In the rest of the paper we use $H = 256$ as our main specification. At its best penalty it earns Sharpe ratios of 2.4 equal-weighted and 1.7 value-weighted, about 90% of the best SAE in the grid, while keeping the number of news factors small enough that each one can be labelled, inspected and compared with the others. With the width fixed, the remaining choice is the penalty, which we make on the basis of interpretability.

2.6 Labeling and Interpretability

For each feature j, we collect the $\underline{K} = 20$ in-sample headlines (published before 2016) that activate it most strongly. This follows common practice in the automated interpretability literature, which labels a feature from its 10 to 20 most strongly activating examples (Bills et al., 2023; Bricken et al., 2023; Cunningham et al., 2023; Choi et al., 2024). We use only alerts, the headline-only articles of the news feed. A headline typically reports a single piece of news, whereas a full article often covers several subjects at once. Headlines usually capture an article's key theme, helping us isolate the common thread among the news items that activate a given feature.[8] A feature is *labellable* in a period if it has at least $\underline{K}$ activating headlines there. We prompt an LLM with a feature's headlines and ask it for (i) a short label ℓ_j naming the concept the headlines share and (ii) a type τ_j, which classifies the feature as an economic theme, firm-specific, a textual pattern, or unclear. We call a feature *clean* if its type is an economic theme and *junk* otherwise.[9] Appendix E reports the full prompt and the model details.

Solely to evaluate interpretability, we repeat this procedure independently on out-of-sample headlines (2016 onward). This second labeling allows us to assess whether features retain their meaning over time, as described in the next subsection. Throughout the paper, we use only the in-sample labels to name and select news factors and group them into themes. The two independent labelings give each feature an in-sample label ℓ_j^{IS} and type τ_j^{IS}, and an out-of-sample label ℓ_j^{OOS} and type τ_j^{OOS}. Let $\mathcal{K}^{\text{IS}}$ and $\mathcal{K}^{\text{OOS}}$ be the sets of features labellable in each period, and $\mathcal{K} = \mathcal{K}^{\text{IS}} \cap \mathcal{K}^{\text{OOS}}$ the features labelled in both.

[8] The restriction to headlines applies only to labeling. All other stages use the full available article text, including the LLM embedding and the SAE activations that determine each article's exposure to the individual features.

[9] A firm-specific feature tracks a single company's news rather than an economic concept and is therefore classified as junk.

2.6.1 Interpretability Metrics

Two SAEs with the same Sharpe ratio can differ greatly in the interpretability of their features: one may have features that each correspond to a clear news subject, the other features that fire on unrelated articles, track a single company, or duplicate one another. We develop four label-based interpretability metrics to assess whether features capture economic content, retain their meaning over time, differ from one another, and faithfully describe the news that activates them. We evaluate these metrics alongside the Sharpe ratio and the sparsity of the representation.

The interpretability metrics focus on the labellable features, $\mathcal{K}$. Label stability compares each feature's two labels, and the other metrics use its in-sample label and type. Several metrics compare texts, which we do by embedding them with OpenAI's `text-embedding-3-small` sentence-embedding model (OpenAI, 2024); $\cos(a, b)$ denotes the cosine similarity between the embeddings of texts a and b. We use the following metrics.

Label stability. An economically relevant factor should capture a stable meaning over time. If the label assigned to the same feature changes radically across periods, the factor cannot be reliably used for economic analysis. We measure label stability as the average similarity between a feature's in-sample and out-of-sample labels,

$$\text{Stability} = \frac{1}{|\mathcal{K}|} \sum_{j \in \mathcal{K}} \cos\Big(\ell_j^{\text{IS}}, \ell_j^{\text{OOS}}\Big). \tag{6}$$

Distinctiveness. We want each feature to capture a distinct economic concept as far as possible. Several features may capture different nuances of the same phenomenon, just as different definitions of momentum do,[10] but the set of text-based factors should avoid unnecessary duplication. We measure how far each in-sample label is from the others, on average and from its nearest neighbor,

$$\begin{aligned} d_{\text{avg}} &= \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \Big(1 - \frac{1}{|\mathcal{K}| - 1} \sum_{j \neq k} \cos\Big(\ell_k^{\text{IS}}, \ell_j^{\text{IS}}\Big)\Big), \\ d_{\max} &= \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \Big(1 - \max_{j \neq k} \cos\Big(\ell_k^{\text{IS}}, \ell_j^{\text{IS}}\Big)\Big), \end{aligned} \tag{7}$$

[10] The JKP momentum theme includes price momentum from $t-3$, $t-6$, $t-9$ and $t-12$ to $t-1$, residual momentum from $t-6$ and $t-12$ to $t-1$, the current price relative to its high over the past year, and the average return over the previous year's nonannual lags.

where j ranges over $\mathcal{K}$. Larger values mean more distinct labels. d_{avg} captures overall redundancy, while $d_{\max}$ specifically penalizes near-duplicate pairs of features.

Label–headline alignment. Our labels are LLM-generated summaries of the top activating headlines. The LLM can supply a plausible label even when those headlines share little coherent content or when its summary fits only a few examples. We therefore assess labeling quality by checking whether the label is semantically close to all of the headlines it summarizes. For feature $j \in \mathcal{K}$, let $\mathcal{H}_j^{\text{IS}}$ be its $\underline{K}$ top in-sample headlines. We average across features the mean, the minimum and the standard deviation of the cosine similarities between each label and its headlines,

$$\text{Align}_{\text{mean}} = \frac{1}{|\mathcal{K}|} \sum_{j \in \mathcal{K}} \underset{h \in \mathcal{H}_j^{\text{IS}}}{\text{mean}} \cos\Big(\ell_j^{\text{IS}}, h\Big). \tag{8}$$

We also compute $\text{Align}_{\min}$ and $\text{Align}_{\text{std}}$ by replacing the within-feature mean with the minimum and standard deviation, respectively. A high mean indicates that the label captures its headlines; a high minimum indicates that it fits even the least typical one; a low standard deviation indicates that the headlines are homogeneous.

Clean fraction. The fraction of in-sample labels that are not junk. A low clean fraction means that few features capture the interpretable economic concepts we seek.

Junk Sharpe. The *Sharpe ratio* is that of the news portfolio re-estimated by MSRR on the in-sample-labellable features $\mathcal{K}^{\text{IS}}$ only, so that the portfolio uses no out-of-sample information. To measure how much of this performance comes from features that cannot be interpreted, we re-estimate MSRR on the clean features alone, $\{j \in \mathcal{K}^{\text{IS}} : \tau_j^{\text{IS}} = \text{economic theme}\}$, which earns SR_{clean}, and define the *junk Sharpe*

$$\text{Junk Sharpe} = 100 \times \left(1 - \frac{\text{SR}_{\text{clean}}}{\text{SR}}\right). \tag{9}$$

A junk Sharpe near zero means that junk features add little economic value and can be safely discarded without sacrificing performance. A large value means that performance relies on features that are not economic themes.

Sparsity. We report the average number of features active (nonzero) on an in-sample article. This is not a measure of interpretability itself, but a diagnostic of whether, and to

what extent, the sparsity penalty works as intended. Increasing κ penalizes active features more heavily and should therefore reduce their number, making the representation sparser.

2.6.2 Benchmarks

We evaluate these metrics for our main SAE specification with 256 features across different values of the penalty κ. Holding the dimension fixed lets us examine how sparsity changes both performance and interpretability, and select a penalty that balances the two. We compare these results with the following benchmarks, each addressing a different question.

Dense embedding. Our first benchmark is the original embedding with all 4,096 coordinates. The labeling procedure applies to these coordinates just as it does to SAE features: for each coordinate, we collect its strongest associated headlines and obtain independent in-sample and out-of-sample labels. Because dense coordinates can be positive or negative, we select headlines by the absolute value of the coordinate rather than by its signed activation. These labels allow us to compute the same interpretability metrics for the unmodified embedding. Improving on this benchmark is central to our objective: the SAE should make the information in the original representation more interpretable, rather than merely reproduce its predictive performance.

Principal components. We also compare the SAE with a conventional dimension-reduction method, retaining the first 256 principal components of the original embedding. This benchmark has exactly the same dimension as our main SAE but has no penalty coefficient enforcing sparsity.[11]

Shuffle benchmark. Because most of these metrics are new, we need intuition for the scores a random classification would achieve. To provide this benchmark, we randomly reassign an SAE's top headlines across its features and apply the same labeling and scoring procedure to the resulting groups. This preserves the pool of headlines while removing the coherent grouping learned by the SAE, giving a reference for how much the interpretability scores exceed what random grouping would produce.

[11] The figure additionally reports the 256 raw embedding coordinates with the highest in-sample absolute Sharpe ratios. This subset is distinct from PCA: it selects existing coordinates rather than constructing new components.

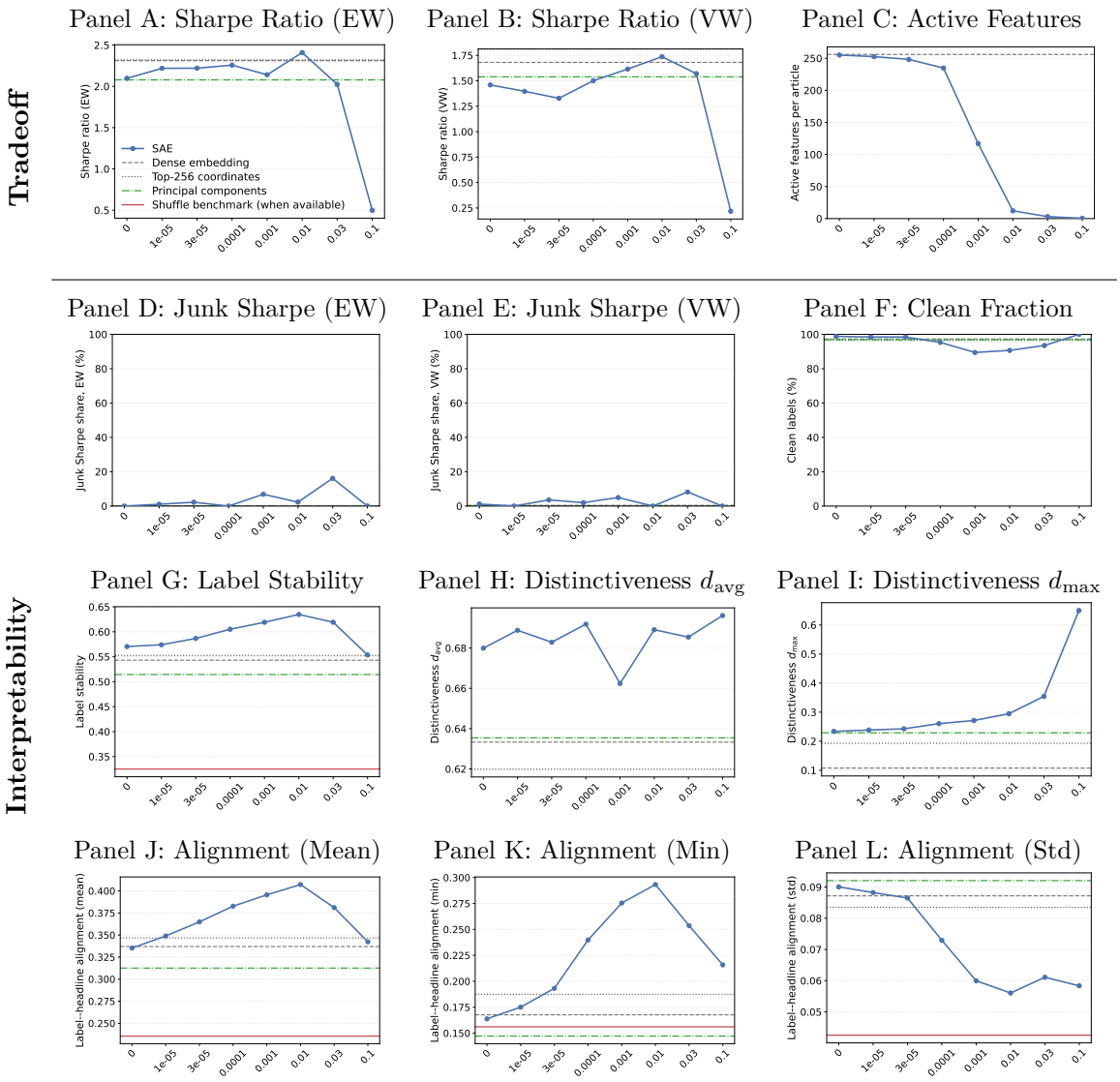


Figure 4: Interpretability Metrics Across the SAE Penalty

Note. Performance and sparsity (A–C) and interpretability (D–L) across penalties for $H = 256$. Each penalty uses the median-Sharpe run, as in Figure 3. Horizontal lines show the dense embedding, its 256 highest-Sharpe coordinates, 256 principal components, and shuffled headlines. Labels are in-sample except for stability. Junk Sharpe is floored at zero. Only three features are labellable in both periods at $\kappa = 0.1$.

2.6.3 Results

Figure 4 reports the interpretability metrics alongside the equal- and value-weighted Sharpe ratios across penalty levels for our main SAE width, $H = 256$, allowing us to assess whether improvements in interpretability come at a cost in performance.

We begin with sparsity and its tradeoff with performance in Panels A–C. With little or no penalty, sparsity barely takes effect: a typical article activates essentially all 256 features. Raising κ reduces this number to about 117 at 10^{-3} and 12 at 10^{-2}, without sacrificing performance. At $\kappa = 10^{-2}$, the equal- and value-weighted Sharpe ratios peak at 2.4 and 1.7, respectively, above those of the full embedding (2.3 and 1.7) and the principal components (2.1 and 1.5). The selected dense coordinates earn 2.3 and 1.8. Almost all SAE features (246 of 256) remain *labellable* in both periods, so the penalty makes each article's representation sparse without eliminating the range of concepts available across articles. Pushing the penalty further destroys this balance: only 77 features survive at $\kappa = 0.03$ and three at 0.1, where performance collapses. The useful region is therefore a moderate penalty that sharply reduces the number of active features per article while preserving the information needed for investment.

Turning to interpretability, the panels show that the gains in sparsity translate into clearer economic concepts. The features remain predominantly economic themes, and junk Sharpe stays low, while label alignment, stability, and distinctiveness improve with the penalty. At $\kappa = 0.01$, the semantic measures exceed the dense and PCA benchmarks without a loss in performance. Label–headline alignment rises from 0.34 without a penalty to 0.41, and its minimum across headlines rises from 0.16 to 0.29. Label stability rises from 0.57 to 0.64, and distinctiveness $d_{\max}$ from 0.23 to 0.29. Label stability of 0.64 compares with 0.55 for the selected dense coordinates, 0.54 for the full embedding and 0.51 for PCA. Alignment of 0.41 compares with 0.35, 0.34 and 0.31, respectively. Both measures also exceed the shuffle null (0.33 for stability and 0.24 for alignment). Junk Sharpe remains below 7% at every penalty up to $\kappa = 0.01$, indicating that the economic-theme features carry nearly all of the performance.

In the rest of the paper we therefore work with the SAE with $H = 256$ features and penalty $\kappa = 10^{-2}$. It activates about 12 of its 256 features per article, keeps nearly all of them labellable in both periods, and has the most stable and best-aligned labels in the grid, at no cost in Sharpe ratio.

3 News-Based Factors

Of the 256 features of our main SAE, 251 are *labellable* in sample.[12] The LLM classifies 16 as firm-specific[13] and seven as textual patterns or unclear, such as duplicate or corrected alerts. After removing these 23 features, we end up with 228 news-based factors whose in-sample labels identify economic themes.[14] We refer to this set of news-based factors as the *factor library*, a text-based counterpart to the *factor zoo* constructed from price and accounting data (Harvey et al., 2016; JKP). All subsequent results use these 228 factors. Appendix F lists their labels.

3.1 From Individual Articles to Firm Exposures

We first build intuition through article-level examples and changes in individual firms' news exposures. Figure 5 extends the Microsoft illustration in Figure 1, showing four articles and their four most strongly activated SAE features.

A strength of our methodology is that it allows a news article to load continuously on multiple topics, with different weights, rather than assigning it to just one. The examples span familiar financial subjects, sector developments and firm-specific events. Intel's credit-rating announcement activates features about *credit ratings*, *earnings*, *semiconductor manufacturing* and *financing*. Nvidia's earnings report combines *earnings and analyst reactions* with *semiconductor production* and *after-hours price movements*. General Electric's manufacturing problems activate *regulatory enforcement* and *medical-device* features, while Exxon's pipeline spill activates *oil operations*, *industrial disruptions* and *regulatory enforcement*. The labels summarize related content rather than describe every event exactly: GE's article concerns a manufacturing inspection rather than a drug approval. The SAE nevertheless captures several recognizable topics within each article.

Figure 6 illustrates how these article-level activations translate into Ford's changing position in four news factors as new articles arrive and older ones leave the six-month averaging window. Each line reports the firm's monthly percentile among all firms on the factor's portfolio-construction signal. The green band marks the top decile, and the

[12] A feature is labellable if it has at least 20 activating alert headlines in the relevant period, as defined in Section 2.6.

[13] Their most strongly activating headlines are dominated by a single company, such as IBM's or Exxon Mobil's earnings releases.

[14] Selection uses only in-sample labels and therefore no out-of-sample information. Removing these features barely impacts performance: the news portfolio's Sharpe ratio moves from 2.41 to 2.35 under equal weighting and from 1.74 to 1.83 under value weighting.

Panel A: Intel

Fitch Affirms Intel's Ratings at 'A+'; Outlook Stable
Fitch Ratings has affirmed Intel Corporation's Issuer Default Rating at 'A+' with a Stable Outlook [1]. Fitch views Intel as one of the strongest credits in the technology space driven by its position as the leading provider of microprocessors for PCs and servers worldwide [3]. [...] The company expects to incur a greater loss in this segment of its business in 2014 [2] as it invests in this development. [...] Fitch expects share repurchases and the company's dividend to be financed by free cash flow [4]. [...]

Top-4 SAE features	vs. avg.
1 Credit rating changes and debt covenants	297×
2 Corporate earnings and business updates	192×
3 Semiconductor process node and chip manufacturing advances	157×
4 Corporate financing and balance-sheet actions	56×

Panel B: Nvidia

After the Bell – Nvidia shares rise after results
NEW YORK, May 10 (Reuters) - Shares of Nvidia Corp. rose 4.2 percent to $34.20 in extended trade [3] on Thursday after the chip maker [2]'s results [4] beat estimates [1].

Top-4 SAE features	vs. avg.
1 Earnings, guidance, and analyst rating reactions	151×
2 Semiconductor process node and chip manufacturing advances	133×
3 Pre-open and after-hours stock price moves	126×
4 Technology company earnings and outlook reports	40×

Panel C: General Electric

US FDA warns GE health unit on plant problems
U.S. regulators have warned [2] General Electric Co.'s health-care unit about manufacturing problems uncovered during an inspection of an Ohio plant [3]. Food and Drug Administration inspectors [1] found shortcomings at a site where the company makes coils used in magnetic resonance imaging systems and breast biopsy plates [4]. [...] The letters can lead to product seizures, injunctions or other penalties [2].

Top-4 SAE features	vs. avg.
1 FDA drug approval and review milestones	213×
2 Regulatory investigations and enforcement settlements	205×
3 Transportation and industrial accidents/disruptions	152×
4 FDA approvals for medical devices	120×

Panel D: Exxon Mobil

Exxon fined $1.7 mln over Yellowstone River oil spill
The U.S. Department of Transportation has fined Exxon Mobil Corp $1.7 million over pipeline safety violations [3] relating to a 2011 oil spill in the Yellowstone River [1]. Exxon's Silvertip pipeline, which carries 40,000 barrels per day of crude in Montana [2], leaked about 1,500 barrels of oil into the river [1] after heavy flooding. [...] Flooding could cause the pipe in the river to lose physical support and potentially rupture [4]. [...]

Top-4 SAE features	vs. avg.
1 Transportation and industrial accidents/disruptions	190×
2 Oil and gas drilling well operations	131×
3 Regulatory investigations and enforcement settlements	123×
4 Refinery outages and maintenance disruptions	104×

Figure 5: News Articles and Their Most Active SAE Features

Note. News excerpts and their four most active topics. Values express activation relative to each topic's article-level average. Author-added shading links passages to numbered topics. Ellipses mark omissions.

annotations paraphrase the news associated with moves into that band. These are relative exposures to news topics, so a high position reflects the subject of the news rather than whether the news is favorable.

Panel A places Ford in the top decile around both Fitch's 2016 upgrade and Moody's 2018 and 2019 downgrades, as well as subsequent debt transactions and its 2023 upgrade.

The feature identifies credit events in either direction. Panel B connects litigation exposure to the 2018 court penalty, the 2022 dispute over connected-car sales, and the 2024 antitrust claims against Blue Cross Blue Shield. Panel C captures a different and recurring subject: executive compensation. Reports of CEO and CFO pay, the new CEO's salary in 2020, and reductions in managers' bonuses in 2024 accompany renewed high exposures to that feature. Ford therefore enters different factors as the content of its news changes, and can enter the same factor again when that subject returns.

Panel D provides another example, linking Ford's exposure to the *inventory-and-sales* feature to news about its automotive operations and sales. Appendix G provides a second walkthrough using Intel.

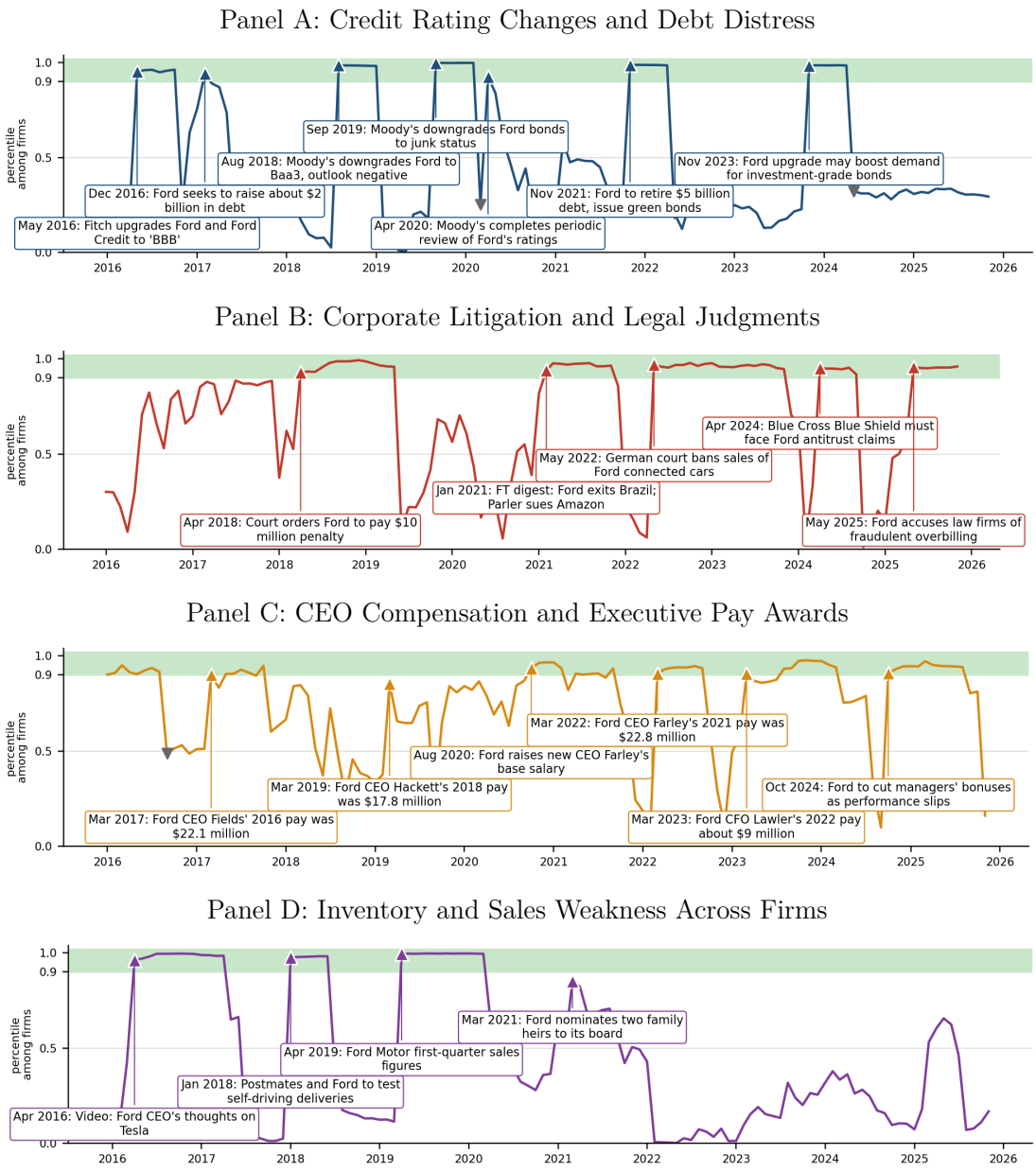


Figure 6: News-Driven Exposures of Ford to Individual News Factors

Note. Ford's monthly cross-sectional percentile on each news factor's equal-weighted signal, constructed as in (3). The green band marks the top decile. Upward triangles mark entry and identify the strongest activating article from that month or the preceding two months, with paraphrased headlines. Downward triangles mark falls from above the 90th to below the 60th percentile. January 2016–November 2025.

3.2 News Themes

To describe this set of news-based factors, we group them into thirteen *themes* by reading their in-sample labels. Six themes cover the routine flow of corporate information: earnings results (31 news factors), guidance and outlook (22), analyst actions and estimates (21), credit ratings and debt financing (19), equity issuance and payouts (9), and management and governance (8). Four cover corporate events and their consequences: mergers, acquisitions and restructuring (17), labor and workforce (5), legal, regulatory and misconduct news (11), and healthcare and biotech approvals and trials (9). The remaining three are sector- and market-wide: energy, utilities and commodities (18); markets, the financial sector and the macroeconomy (18); and industry operations and products (40), which collects the sector-specific operating news of retailers, airlines, technology, media, telecommunications, manufacturing and defense firms. Table 6 in Appendix F lists every news factor under its theme.

Themes Coverage. Figure 7 shows how often each news theme appears in our sample over time. Each article is split across the themes in proportion to its summed SAE activation on each theme's news factors, and we report each theme's share of the month's articles.

The coverage shares track familiar economic developments. Coverage of *guidance and outlook* falls from 12.4% in 2019 to 10.1% over March–December 2020, consistent with firms withdrawing guidance, while *credit ratings and debt financing* peaks at 11.2% in May 2020 amid downgrades and emergency debt issuance. Appendix O reports the remaining seven themes and discusses further changes in coverage.

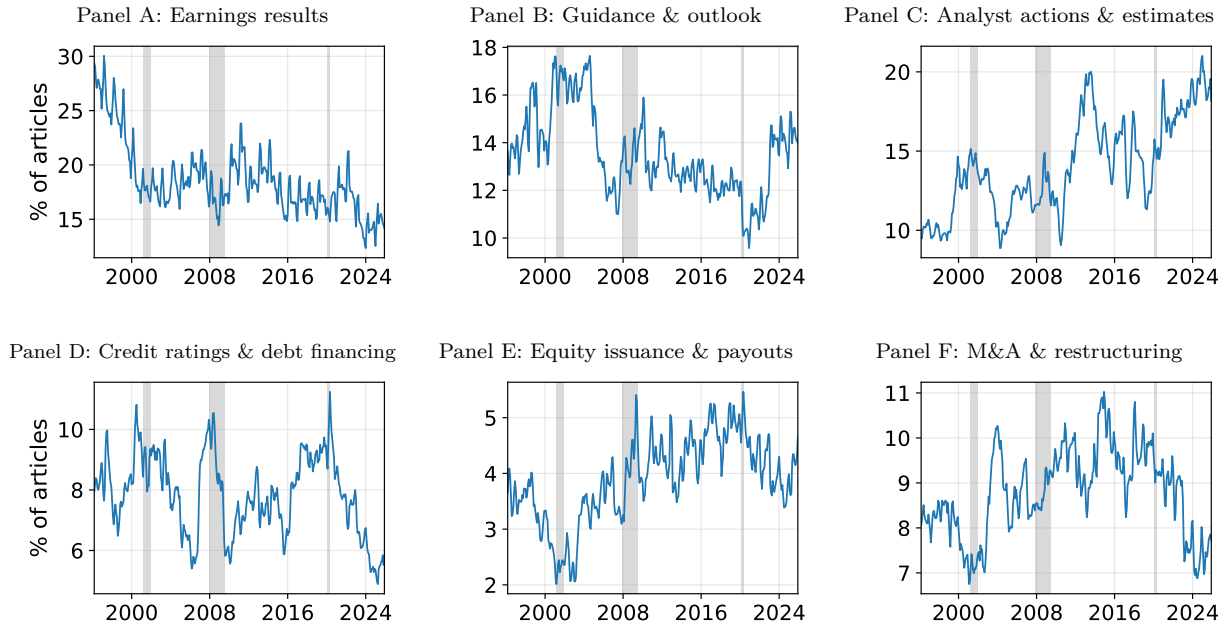


Figure 7: News Themes Over Time

Note. Monthly coverage shares for six news themes. Each article with an active factor is allocated across themes in proportion to their summed SAE activations. All thirteen shares sum to 100%. Lines are centered three-month averages with panel-specific scales. Gray bands mark NBER recessions. March 1996–December 2025.

Themes in the Portfolio. The preceding analysis shows which themes receive the most coverage in firm-specific news. More importantly, we use a Shapley decomposition to measure each theme's contribution to the news portfolio's Sharpe ratio.[15]

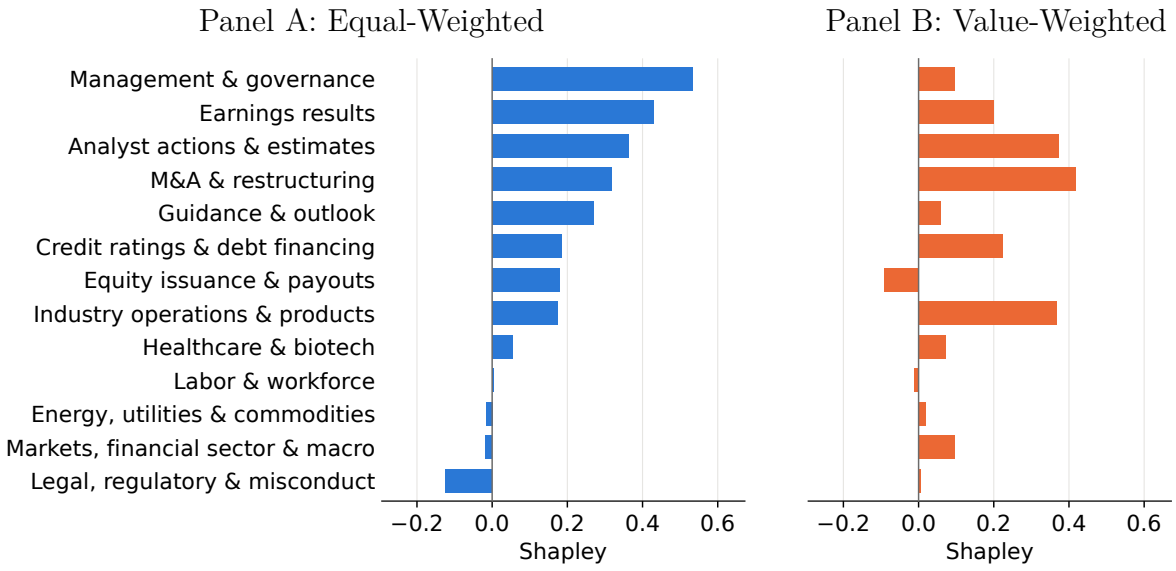


Figure 8: News Themes' Shapley Contributions to the Sharpe Ratio

Note. Themes' Shapley contributions to annualized portfolio Sharpe ratios under equal weighting (A) and value weighting (B). January 2016–November 2025.

Figure 8 shows that several economically distinct themes contribute to performance. Under equal weighting, *management and governance*, *earnings results*, and *analyst actions* make the largest contributions. Under value weighting, *M&A and restructuring*, *analyst actions*, and *industry operations and products* stand out. *Guidance* and *management* contribute more under equal weighting, while *industry operations* contributes more under value weighting. The gains thus extend beyond routine earnings and analyst news to corporate decisions and operating developments. Appendix P separates the themes' contributions to the mean and volatility.

[15]For each subset of the 13 themes, we form a portfolio by retaining only the factors belonging to those themes, with their monthly weights held at the values estimated for the full news portfolio using the expanding-window MSRR procedure. We assign each subset its annualized out-of-sample Sharpe ratio and the empty set a value of zero. Each theme's Shapley contribution is its marginal effect on the Sharpe ratio averaged over all orders in which themes can enter the portfolio. We compute these averages exactly by evaluating all $2^{13} = 8{,}192$ subsets, without re-estimating portfolio weights for any subset. Contributions sum to the full portfolio's Sharpe ratio. We perform the decomposition separately for equal- and value-weighted factors.

3.3 News Factors and the Macroeconomy

Connecting stock-return factors to macroeconomic risks is a longstanding goal in asset pricing (Chen et al., 1986), but identifying their underlying economic drivers remains challenging (Fama and French, 1995). For some of our news factors, the topic labels suggest intuitive connections: layoffs to labor-market conditions, oil production to energy prices, and debt refinancing to financial conditions. We examine whether these connections appear in correlations between factor returns and the corresponding macroeconomic quantities.[16]

Figure 9 shows six value-weighted news factors whose returns co-move with a macroeconomic series and whose labels tell a coherent story.[17] The news factor built on investment bank and capital markets news tracks the intermediary capital risk factor of He et al. (2017) closely (correlation of 0.64). The news factor built on oil and gas drilling moves with the oil price (0.29, and 0.34 against the next month's change, since the FRED-MD oil price is a monthly average), and it leads inflation by a month (0.20 in the same month, 0.42 a month ahead) as energy prices pass into consumer prices. The dividend-declarations news factor rises with the dividend yield (0.46): the yield rises when prices fall, and dividend payers fall less. The news factor built on foreign investment and state asset transactions falls when the dollar strengthens (-0.25), as commodity and emerging-market exposures do, and the news factor built on debt refinancing and credit-facility amendments falls when market volatility is high (-0.24 against the level of the VIX).

Appendix Q also shows that news attention, measured by each topic's activation-weighted share of articles and complementing the theme-level coverage in Figure 7, co-moves with macroeconomic conditions: layoffs with labor-market weakness, analyst revisions with activity and stock prices, and credit news with financial stress.

[16] Note that our factors are built exclusively from firm-specific news, unlike Bybee et al. (2024), who study the broader business news in the *Wall Street Journal*. We do not construct them from a corpus of macroeconomic news, so their links to aggregate phenomena are not obvious ex ante.

[17] For this exercise, we omit size and industry controls to retain exposures to aggregate shocks.

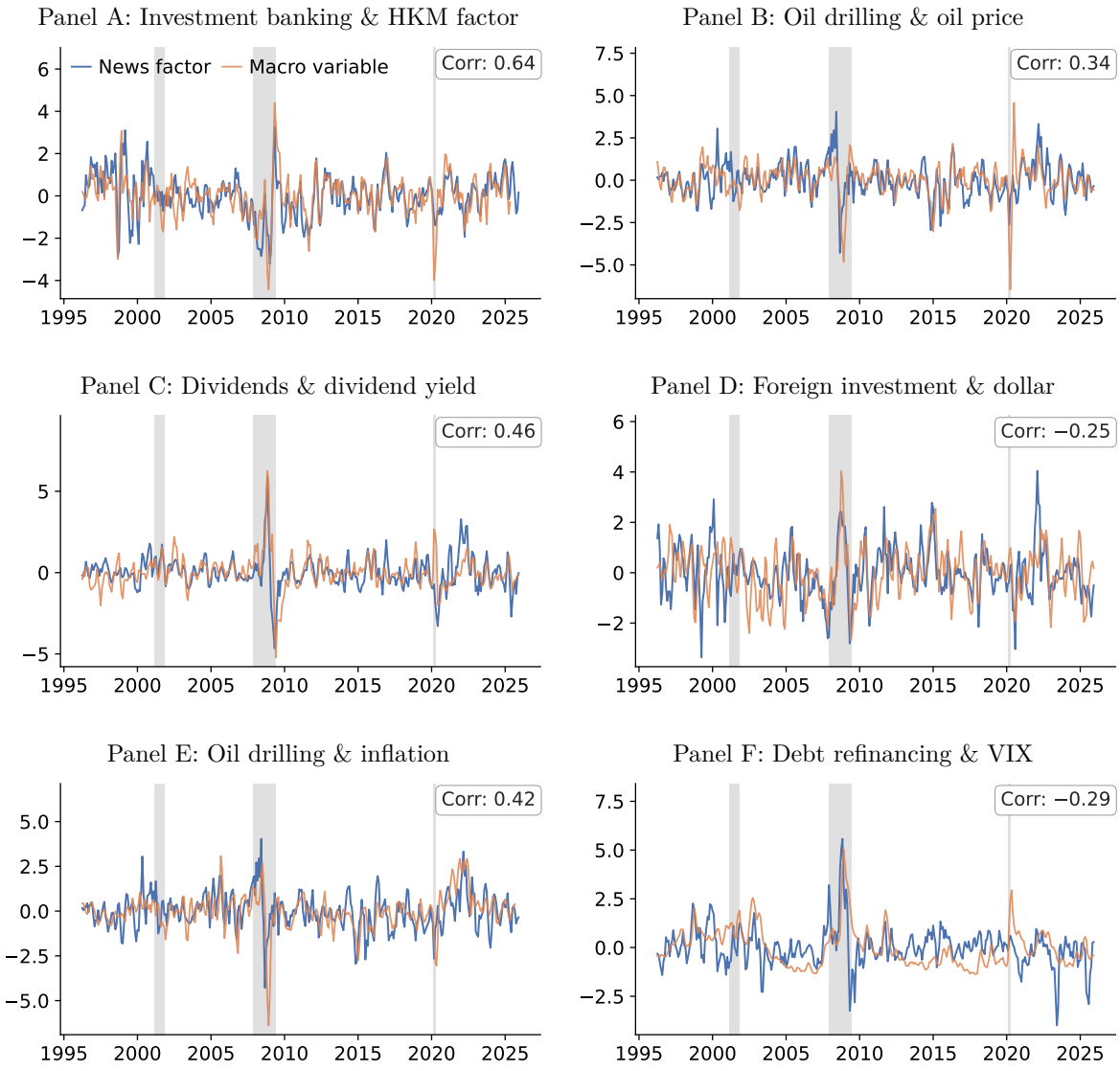


Figure 9: News Factor Returns and Macroeconomic Series

Note. Value-weighted news-factor returns (blue) and macroeconomic series (orange), matched by return month. Macro series use FRED-MD transformations, except the intermediary capital risk factor, used as published. Foreign-investment and debt-refinancing returns are sign-reversed for display, but reported correlations use original signs. Lines are standardized three-month averages. The box in each panel reports the largest correlation between the monthly series within six months of lead or lag. Gray bands mark NBER recessions. 1996–2025.

4 The Library versus the Zoo

The implicit motivation of this paper is that, had researchers always been able to process financial text at scale, news would have been a natural starting point for constructing firm-level factors. Our conjecture is that the prominence of accounting and price characteristics reflects, in part, the ease with which they could be measured using the tools available. News describes a much broader range of developments affecting firms. For this conjecture to have empirical content, however, news-based factors should be able to capture the information in traditional characteristics while adding information beyond them.

In this section, we assess how far our news factors meet that requirement. They expand the efficient frontier relative to the characteristic zoo, and adding characteristics to news yields only modest further gains. A stochastic discount factor (SDF) constructed from news alone therefore captures most of the performance attainable by combining the two portfolios. We also show that news factors explain substantial variation in characteristic-factor returns, with much of this explanatory power concentrated in three intuitive themes: *earnings*, *analyst actions*, and *guidance*. The reverse relationship is considerably weaker. Together, these results suggest that news recovers much of the information captured by traditional characteristics while providing a richer set of firm exposures.

We use monthly excess returns and characteristics of individual stocks from JKP. We apply coverage filters following Didisheim et al. (2024) to arrive at the 132 JKP characteristics with the highest coverage over our sample period. All characteristics are then rank-standardized each month to $[-\frac{1}{2}, \frac{1}{2}]$ following Gu et al. (2020). The remaining missing characteristic observations are imputed with zeros, the cross-sectional mean rank within each month. Appendix H describes the construction of the characteristic factors.

4.1 SDF Performance and Pricing

In empirical asset pricing, estimating an SDF amounts to finding a combination of factor returns that explains differences in average returns across assets. A successful SDF should capture the return opportunities available in those factors and leave little of the test assets' average returns unexplained. Following Kelly et al. (2026a), we construct an SDF from each set of factors using its maximum-Sharpe portfolio:

$$M_{t+1} = 1 - \hat{c}\, f_{t+1}, \tag{10}$$

where f_{t+1} is the excess return of the news or characteristic portfolio and $\hat{c}$ is a normalization coefficient.

We estimate each portfolio using maximum Sharpe ratio regression (MSRR), equation (5). Each month, we estimate factor weights using only past returns, select the regularization penalty within that estimation window, and apply the weights to the following month's returns.[18] This produces out-of-sample returns for both portfolios. We then normalize each SDF to price its own portfolio over the evaluation period.[19] We compare the two constructions along two dimensions: the Sharpe ratios of their associated portfolios and their pricing errors on both news and characteristic factors.

SDF performance. Figure 10 compares the news portfolio, the characteristic portfolio, and their optimal combination from January 2016 to November 2025. The news portfolio achieves annualized Sharpe ratios of 2.35 under equal weighting and 1.83 under value weighting, compared with 1.54 and 1.12 for the characteristic portfolio. Each plotted return series is scaled to 10% annualized volatility over the evaluation period, allowing comparison at a common level of risk.

Even an infeasible combination chosen with hindsight places most of its weight on news and improves little on news alone. We scale the two SDF portfolios to the same volatility, estimate their means and covariance matrix using all 119 evaluation months, and choose constant weights that maximize their combined Sharpe ratio subject to summing to one. This ex-post combination uses information from the entire evaluation period and is therefore infeasible in real time. It serves as a hindsight benchmark for what characteristics can add to news. The optimal combination assigns 70.8% to news under equal weighting and 75.4% under value weighting. Yet the Sharpe ratio rises only from 2.35 to 2.49 under equal weighting and from 1.83 to 1.90 under value weighting, gains of about 6% and 4%. Characteristics retain some diversification value, but news captures most of the attainable performance even against this ex-post combination.

[18] We use an expanding estimation window and select the ridge penalty by leave-one-out cross-validation within that window. We apply the same procedure to our 228 news factors and the 132 characteristic factors of JKP, under both equal and value weighting. Appendix H details the construction of the characteristic factors.

[19] We set $\hat{c} = \overline{f}/\overline{f^2}$, which ensures $\overline{Mf} = 0$. This scalar normalization uses the evaluation-period returns and is therefore ex post, although the underlying portfolio weights are estimated using only past data. Appendix I provides the construction details.

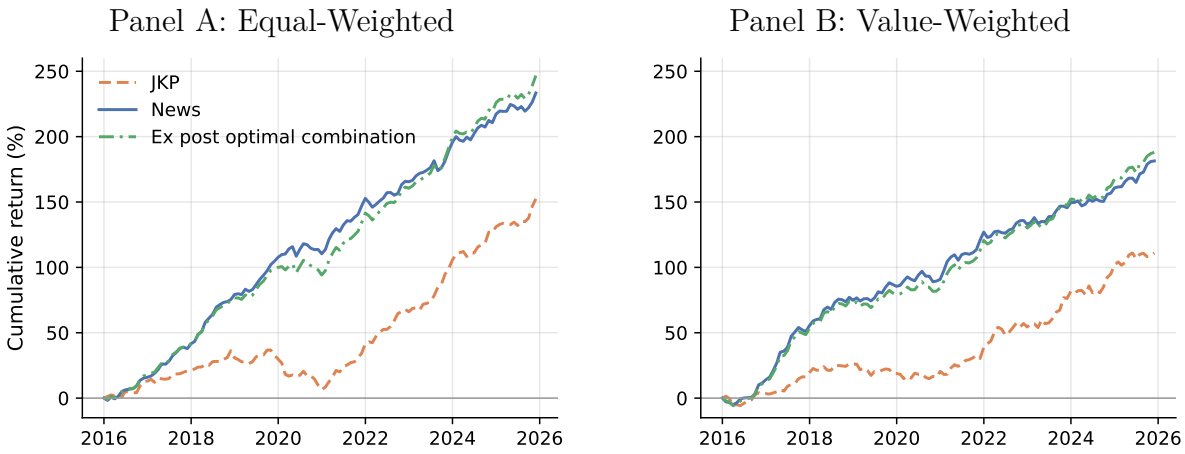


Figure 10: Cumulative Returns of the News and Characteristic Portfolios and Their Optimal Combination

Note. Cumulative excess returns of the news and JKP MSRR portfolios and their ex-post optimal combination, scaled to 10% annualized volatility. MSRR returns are out of sample. Combination weights use the full evaluation period and are infeasible in real time. Returns are summed, not compounded. January 2016–November 2025.

Pricing the factors. We next hold these two SDFs fixed and ask how well each prices both the 132 characteristic factors and the 228 news factors. We test whether the performance advantage of news also translates into smaller pricing errors on the factors themselves.

We measure pricing performance using the Hansen–Jagannathan (HJ) distance (Hansen and Jagannathan, 1997):[20]

$$D = \left(\hat{\alpha}' \hat{\Sigma}_z^{-1} \hat{\alpha}\right)^{1/2}, \tag{11}$$

where $\hat{\alpha}$ is the vector of pricing errors and $\hat{\Sigma}_z$ is the regularized second-moment matrix of factor returns. A lower distance indicates smaller pricing errors after accounting for how factor returns move together. We compare the news and characteristic SDFs on each set of factors and separately within the 13 JKP themes and the 13 news themes (defined in Section 3.2), under both equal and value weighting.[21]

[20] See also Kelly et al. (2025).

[21] For each full set, we set $\hat{\Sigma}_z = (1-z)\hat{\Sigma} + zI$ with $z = 0.5$, after scaling each factor to unit second moment. Regularization is necessary because there are more factors than evaluation months. Within themes, we use the unregularized distance ($z = 0$), holding each SDF and its calibration fixed across both factor sets and all themes. JKP themes contain 4 to 20 factors and news themes contain 5 to 40. All theme-level second-moment matrices are invertible. Appendix I provides the remaining details.

Table 1: Hansen–Jagannathan Distance of the News and Characteristic SDFs

Panel A: JKP characteristic test assets

	Equal-weighted		Value-weighted	
Test assets	News	JKP	News	JKP
Full panel (132)	**0.499**	0.534	**0.520**	0.573
Accruals	**0.125**	0.169	0.164	**0.119**
Debt issuance	**0.192**	0.263	0.231	**0.201**
Investment	**0.390**	0.402	**0.475**	0.504
Low leverage	0.272	**0.256**	**0.267**	0.278
Low risk	0.376	**0.317**	0.344	**0.340**
Momentum	0.225	**0.222**	0.214	**0.184**
Profit growth	**0.197**	0.268	0.197	**0.185**
Profitability	**0.266**	0.286	**0.276**	0.397
Quality	0.347	**0.340**	0.356	**0.351**
Seasonality	**0.203**	0.259	**0.140**	0.163
Size	**0.153**	0.188	**0.144**	0.217
Short-term reversal	**0.215**	0.228	0.169	**0.114**
Value	0.365	**0.311**	**0.265**	0.347

Panel B: News test assets

	Equal-weighted		Value-weighted	
Test assets	News	JKP	News	JKP
Full panel (228)	**0.813**	0.969	**0.861**	1.002
Earnings results	**0.414**	0.527	**0.471**	0.547
Guidance and outlook	**0.437**	0.532	**0.424**	0.481
Analyst actions and estimates	**0.380**	0.526	**0.449**	0.483
Credit ratings and debt financing	**0.438**	0.473	**0.426**	0.513
Equity issuance and payouts	0.324	**0.296**	0.319	**0.301**
M&A and restructuring	**0.388**	0.519	**0.341**	0.440
Management and governance	**0.308**	0.437	**0.335**	0.455
Labor and workforce	**0.216**	0.273	**0.108**	0.148
Legal, regulatory and misconduct	**0.272**	0.364	**0.336**	0.370
Healthcare and biotech	**0.189**	0.249	**0.239**	0.279
Energy, utilities and commodities	**0.370**	0.451	**0.345**	0.410
Industry operations and products	**0.496**	0.603	**0.581**	0.665
Markets, financial sector and macro	**0.309**	0.402	**0.381**	0.448

Note. HJ distances for the news and JKP SDFs pricing 132 JKP factors (A) and 228 news factors (B). EW and VW denote equal and value weighting. Bold marks the lower distance within each pair, not statistical significance. Calibration is ex post and fixed across test assets. Distances use $z = 0.5$ for complete sets and $z = 0$ for themes. Compare distances within rows. January 2016–November 2025.

Panel A of Table 1 shows that the news SDF prices the characteristic zoo better in aggregate, despite being constructed from a different source of information. Its regularized HJ distance is 0.499 rather than 0.534 under equal weighting and 0.520 rather than 0.573 under value weighting, reductions of 6.5% and 9.1%. By JKP theme, news achieves a lower distance in 8 of the 13 JKP themes under equal weighting and 6 under value weighting. Its largest value-weighted advantages are in *size*, *profitability*, and *value*. When the JKP SDF prices a theme better than ours, its advantage is often small, although the gaps are larger for value-weighted *accruals* and *short-term reversal*.

The reverse pricing comparison is much more one-sided. Panel B shows that the news SDF prices its own factors better in 12 of the 13 news themes under both weightings. Across all 228 news factors, its HJ distance is 0.813 against 0.969 for the characteristic SDF under equal weighting and 0.861 against 1.002 under value weighting, improvements of 16.1% and 14.1%. The characteristic SDF wins only for *equity issuance and payouts*, where its distances are 0.296 versus 0.324 under equal weighting and 0.301 versus 0.319 under value weighting. Its advantage in this sole exception is modest compared with the news SDF's improvements for *management and governance*, *labor and workforce*, and *M&A and restructuring*.

Together, the performance and pricing results favor the news-based SDF. It delivers substantially higher Sharpe ratios, gains little from adding the characteristic portfolio even ex post, and produces lower aggregate pricing errors on both sets of test assets.

4.2 Individual Factor Performance and Turnover

We next examine the individual factors behind the portfolio comparison. Figures 11 and 13 compare their Sharpe ratios and turnover, respectively.

Panels A and B of Figure 11 compare the absolute out-of-sample Sharpe ratios of individual news and characteristic factors.[22] The median absolute Sharpe ratios are similar across the two sets: 0.37 for news and 0.34 for characteristics under equal weighting, and 0.27 and 0.23 under value weighting. The news library nevertheless contains more factors with high Sharpe ratios. Among equal-weighted news factors, 9% have an absolute Sharpe ratio above one, compared with none of the characteristic factors. The value-weighted comparison shows a similar pattern.

[22]For news factors, the sign follows the SAE-based exposure signal rather than an economically chosen direction: stronger activation of a topic need not indicate favorable news or a positive expected return. Reversing a factor's long and short positions reverses its Sharpe ratio without changing its investment opportunity. We therefore compare absolute Sharpe ratios for both sets of factors.

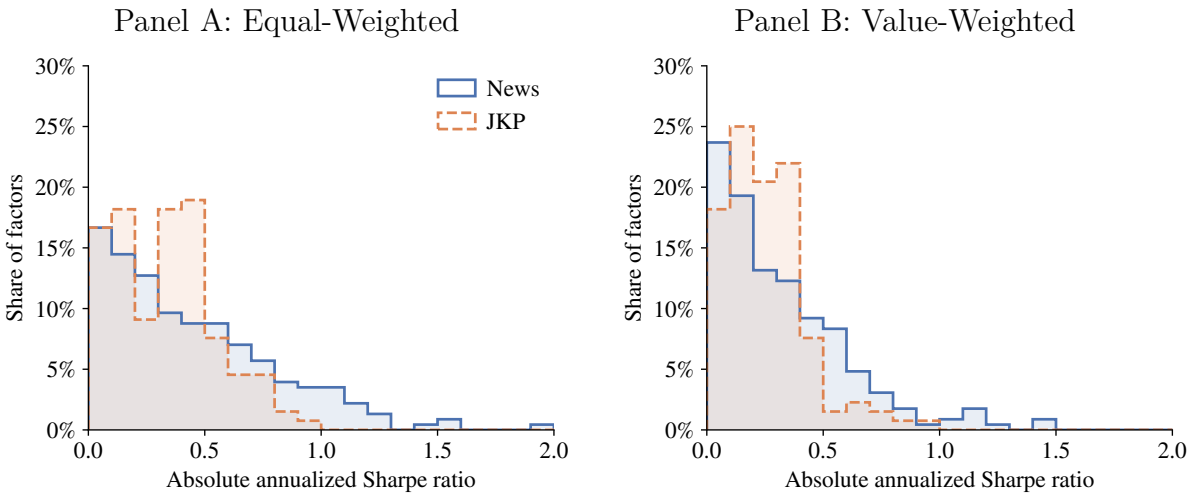


Figure 11: Sharpe Ratios of Individual News and JKP Factors

Note. Distributions of absolute annualized Sharpe ratios for 228 news (blue, solid) and 132 JKP (orange, dashed) factors under equal (A) and value (B) weighting. January 2016–November 2025. Bin heights report percentages within each library. Both panels use the same bins and axis scales.

Figure 12 helps explain why the news factors perform well when combined (see Figure 10). Characteristic factors tend to move together: five principal components explain 90% of their equal-weighted return variance, compared with 71 components for news factors. Under value weighting, the corresponding counts are 9 and 69. News factors thus capture a broader range of return patterns. This greater diversity provides more scope for diversification when news factors are combined.

A striking result is that the news factors deliver higher portfolio Sharpe ratios with substantially less turnover. Figure 13 shows that mean monthly turnover across individual news factors is 16.3%, compared with 18.2% for JKP factors under equal weighting, and 13.3% versus 20.9% under value weighting. The lower means reflect the high-turnover tail of the JKP distribution. Median turnover is slightly higher for news: 16.7% versus 12.5% under equal weighting and 13.8% versus 12.6% under value weighting. The difference is especially pronounced among the highest-turnover factors: the maximum is 34.8% for news versus 67.6% for JKP under equal weighting, and 34.0% versus 84.7% under value weighting. The latter corresponds to *short-term reversal*, whose positions respond to recent price movements. At the portfolio level, the news MSRR strategy also requires less rebalancing. Mean monthly turnover of the overall portfolio is 20.8% for news versus 29.6% for JKP under equal weighting, and 15.9% versus 26.7% under value weighting. Our news factors thus

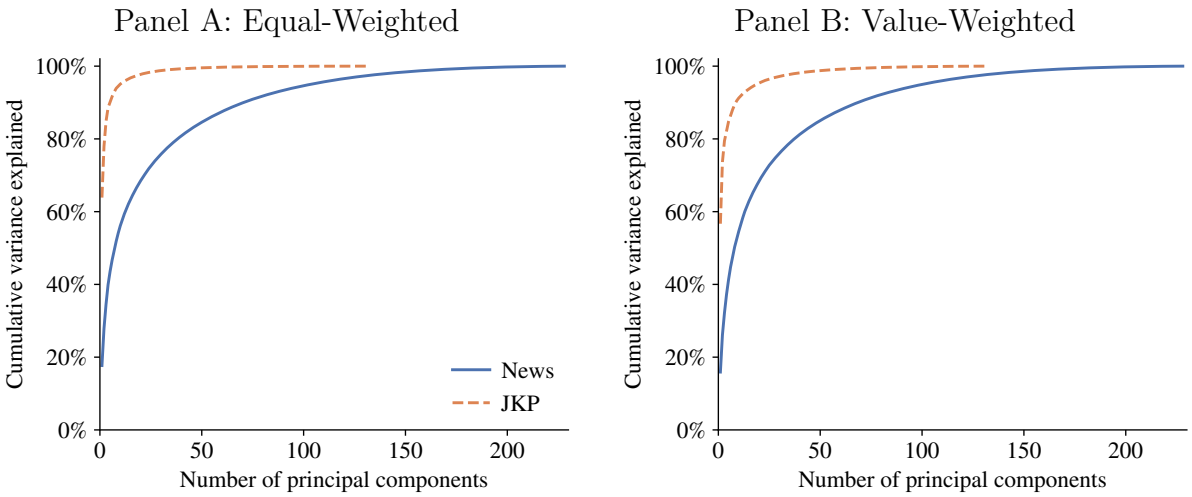


Figure 12: Principal-Component Structure of the Library and the Zoo

Note. Cumulative variance explained by principal components of news-factor (blue, solid) and JKP-factor (orange, dashed) returns. Each decomposition uses its own covariance matrix, estimated over January 1996–November 2025 (359 months), with 228 news and 132 JKP factors. Both panels use the same axis scales.

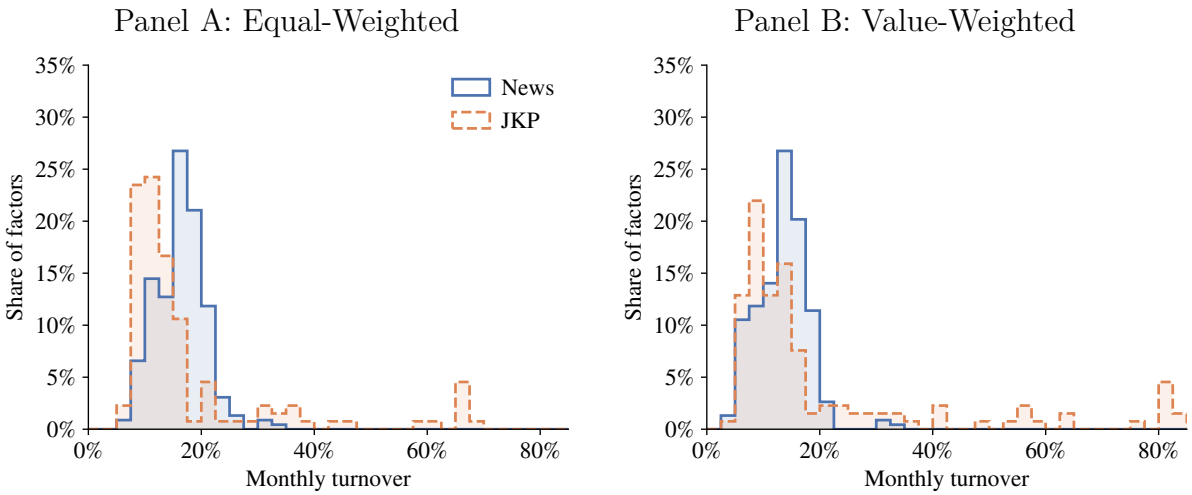


Figure 13: Turnover of Individual News and JKP Factors

Note. Distributions of monthly turnover for 228 news (blue, solid) and 132 JKP (orange, dashed) factors, normalized to one dollar of gross exposure, under equal (A) and value (B) weighting. January 2016–November 2025. Bin heights report percentages within each library. Both panels use the same bins and axis scales.

address a major implementation problem in characteristic-based asset pricing: important factors can require frequent and substantial changes in stock positions. When prices enter a

characteristic, price movements mechanically change the signal and can trigger rebalancing even without new information about the firm. Our textual approach instead measures exposure to firm narratives that can persist across successive news reports, providing a more stable basis for factor positions.

4.3 Do the Library and the Zoo Span Each Other?

To build further intuition, we ask whether a few news factors explain an individual JKP factor, and vice versa, using their labels to interpret the relationship.

We use a LASSO procedure to select the five most important news factors for explaining each JKP factor's returns using data prior to 2016, then evaluate their explanatory power out of sample.[23] The selection period coincides with the training period of the point-in-time language model and the SAE, ensuring that no information from the evaluation period enters the factor selection. The procedure consists of three steps.

1. *Scale.* We scale every factor to 10% annualized volatility, so that differences in return scales do not drive the selection.

2. *Select.* Using only months before T_0, January 2016, we choose the five news factors that best explain JKP factor k. We pool its equal- and value-weighted returns, so one set of factors is chosen for both weightings:

$$\min_{a,\beta} \sum_{w\in\{\text{EW,VW}\}} \sum_{t<T_0} \left(G^w_{k,t+1} - a - \beta' F^w_{t+1}\right)^2 + \gamma\|\beta\|_1, \tag{12}$$

   where $G^w_{k,t+1}$ is the JKP factor's return under weighting w and F^w_{t+1} is the vector of news factor returns. The penalty $\gamma\|\beta\|_1$ sets most coefficients to zero. We lower γ, tracing the LASSO path by least-angle regression, until five factors have nonzero coefficients. We denote this set by $\mathcal{J}_k$ and keep it fixed for the rest of the exercise.

3. *Refit.* Over the evaluation period, we regress the JKP factor's return on its five selected news factors:

$$G_{k,t+1} = \alpha_k + \sum_{j\in\mathcal{J}_k} \beta_{kj} F_{j,t+1} + u_{k,t+1}. \tag{13}$$

[23] See Bybee et al. (2024) for a similar approach.

The regression R^2 measures the share of return variation explained by the selected factors, while α_k measures the alpha left unexplained.[24]

We repeat the procedure in reverse, selecting five JKP factors to explain each news factor.[25] Appendix K reports a complementary exercise in which the regression coefficients are estimated on a rolling basis and 1-step ahead forecasts.

In line with the HJ results, our news factors explain substantially more of JKP return variation in aggregate than JKP factors explain of news return variation (Appendix J).

Interpreting individual factors. Table 2 shows how familiar factors connect to specific pieces of financial news. We use the published Fama–French momentum (UMD), value (HML), profitability (RMW), and CMA factors, selecting five news factors for each with the same pre-2016 procedure.[26] Their loadings reveal economically recognizable relationships rather than opaque coordinates.

Momentum provides the clearest example. Its five news factors explain 68% of its return variation, with the strongest loadings on *analyst price target downgrades* and *negative earnings and outlook revisions*. Both load negatively, as do *corporate credit rating actions* and *corporate layoffs and workforce reductions*. Momentum thus connects to a coherent set of reports about deteriorating firm prospects. Value tells a different earnings story: it loads positively on *corporate earnings and quarterly results* and negatively on *technology company earnings and outlook reports*, with an R^2 of 53%. These exposures give the value–growth distinction an interpretation in terms of the earnings news reaching firms.

Profitability has an equally direct connection to firms' operating results. It loads negatively on *quarterly earnings reports showing net losses* and positively on *quarterly revenue growth reports*, with an R^2 of 46%. CMA is less well explained ($R^2 = 32\%$), with its strongest loading on *internet platform user growth and engagement*. The examples show how individual news factors can turn familiar return patterns into specific economic narratives.

[24] We calculate

$$R^2_k = 1 - \frac{\sum_{t \geq T_0} \left(G_{k,t+1} - \hat{G}_{k,t+1}\right)^2}{\sum_{t \geq T_0} \left(G_{k,t+1} - \bar{G}_k\right)^2}, \tag{14}$$

where $\hat{G}_{k,t+1}$ is the fitted return and $\bar{G}_k$ is the evaluation-period mean.

[25] The evaluation covers 119 months from January 2016 to November 2025, with returns realized the following month.

[26] The French factors use 2×3 sorts on size and the characteristic with NYSE breakpoints, so they differ from our JKP factors and require a separate selection. We use value-weighted news factors to match the weighting of the published factors.

Table 2: Replicating Fama–French Factors with News Factors

OLS refits of momentum (UMD), value (HML), profitability (RMW), and investment (CMA) on five value-weighted news factors selected before 2016. Refits use January 2016–November 2025. Factors are scaled to 10% annualized volatility without demeaning. The first row reports annualized alpha in percent. News labels are in-sample. Parentheses report t-statistics. *, **, and *** denote significance at 10%, 5%, and 1%. Rolling estimates appear in Appendix K.

Panel A: Momentum (UMD)	
α (% p.a.)	-0.67
	(-0.45)
Negative earnings and outlook revisions	-0.287^{***}
	(-4.53)
Corporate share buyback announcements	0.040
	(0.43)
Analyst price target downgrades	-0.355^{***}
	(-8.27)
Corporate credit rating actions	-0.177^{***}
	(-2.81)
Corporate layoffs and workforce reductions	-0.100^{**}
	(-2.19)
R^2	0.677

Panel B: Value (HML)	
α (% p.a.)	0.48
	(0.18)
Technology company earnings and outlook reports	-0.316^{***}
	(-2.58)
Corporate earnings and quarterly results	0.553^{***}
	(5.14)
Corporate credit rating actions	0.101
	(0.89)
Credit rating changes and debt covenants	0.063
	(0.53)
Biotech drug development and clinical trials	-0.206^{*}
	(-1.81)
R^2	0.530

Panel C: Profitability (RMW)	
α (% p.a.)	3.24^{*}
	(1.66)
Equity secondary offerings and share issuance	0.004
	(0.05)
Earnings per share results versus estimates	-0.238^{**}
	(-2.53)
Quarterly revenue growth reports	0.426^{***}
	(4.30)
Quarterly earnings reports showing net losses	-0.467^{***}
	(-7.11)
Biotech drug development and clinical trials	-0.207^{**}
	(-2.47)
R^2	0.464

Panel D: Investment (CMA)	
α (% p.a.)	1.96
	(0.64)
Technology company earnings and outlook reports	-0.119
	(-0.78)
Analyst stock rating changes	0.246^{*}
	(1.77)
Internet platform user growth and engagement	-0.763^{***}
	(-4.39)
Earnings forecast and report corrections	-0.047
	(-0.50)
Utility power generation and regulated energy assets	0.028
	(0.20)
R^2	0.324

Which themes explain the factors? The individual regressions offer an additional advantage over the aggregate HJ comparison: we can identify which news themes explain

which JKP themes. We group their contributions into the 13 news themes and use a Shapley decomposition to allocate each regression's explained variation across them, accounting for overlap (Huettner and Sunder, 2012). We then average across factors within each of the 13 JKP themes. Figure 14 reports the resulting contributions, with each row summing to the theme's mean individual-factor R^2.

Most of the explanatory power is concentrated in a small, intuitive part of the news: *earnings*, *analyst actions*, and *guidance*. These themes describe operating results, revisions to expected performance, and the analyst assessments that connect naturally to accounting measures and momentum. For momentum, *analyst actions* and *guidance* alone account for 46 of its 54 explained percentage points under equal weighting and 51 of 56 under value weighting. *Healthcare* news also matters for some JKP themes, particularly value. A few economically recognizable news themes therefore account for much of the variation that our news factors explain in the traditional factors.

The reverse decomposition in Appendix L is both weaker and less concentrated. JKP factors draw on *quality*, *value*, *low risk*, *momentum*, and other themes to explain news-factor returns. With the exception of *equity issuance and payouts*, whose mean R^2 is 37%, mean R^2 across news themes ranges from only 4% to 25% under equal weighting, compared with 21% to 59% for JKP themes explained by news. The asymmetry is in both the amount and the concentration of explanatory power: a few news themes recover much of the familiar characteristic-based variation, while a wider range of JKP themes explains a smaller share of news-factor variation. *Equity issuance and payouts* is the best-explained news theme under both weightings, matching its status as the sole news theme that the JKP SDF prices better in Table 1.

The news themes that explain JKP factors are only part of what makes the news SDF perform well. The bottom row of Figure 14 reports each theme's Shapley contribution to the news portfolio's Sharpe ratio, repeating Figure 8. *Earnings*, *analyst actions*, and *guidance* together account for 45% of the Sharpe ratio under equal weighting and 35% under value weighting. The remaining themes account for 55% and 65%, despite playing a smaller role in explaining JKP factors. For example, *M&A* contributes 14% and 23%, and *industry operations and products* contributes 7% and 20%, respectively. The news portfolio thus draws much of its performance from economic developments beyond those that explain familiar characteristic-based factors.

Panel A: Equal-Weighted

	Earnings	Analysts	Healthcare	Guidance	Payouts & issuance	M&A	Markets & macro	Credit & debt	Operations	Energy	Labor	Management	Legal & regulation	Total
Low risk	28	17	7	2	2					2				59
Value	29		15	1	5			3		2	1			57
Profitability	29	12	11	1	3	1		1						57
Momentum	3	31		15	1			4						54
Low leverage	21	1	10		2	1	2	2		1				41
Quality	16	8		3	2	5	3	1		1	1			39
Accruals	11	6	6	2		3	5	2						34
Size	5			6	12	1			3		3			31
Short-term reversal	4	14	4	2		1	1							25
Debt issuance	4	3	7			3		3	3					23
Investment	16	1		1	1				1	2				21
Profit growth	3	4		5	1		1	3			3			21
Seasonality	2	3	1	1		6	1	2	3			1		21
Sharpe contribution	0.430	0.362	0.054	0.271	0.179	0.318	-0.017	0.185	0.173	-0.016	0.004	0.532	-0.123	2.354

Panel B: Value-Weighted

Low risk	26	15	7	1	1				2	1				53
Value	41		15		3			1		1	1			63
Profitability	31	16	7	1	5	1								61
Momentum	2	33		18	1			1						56
Low leverage	30	1	12		3	2	3	1		1				52
Quality	14	9		2	1	11	9	2		2	1			50
Accruals	11	6	2			2	5							27
Size	5	1		8	8	1			2		1			25
Short-term reversal	5	8	2	2		1	1							18
Debt issuance	6	4	8			3		1	3					24
Investment	32	3		1	1				1	3				42
Profit growth	4	4		7	2	1	2	1			1			22
Seasonality	4	4	2	1		2	1	1	2	1		1		20
Sharpe contribution	0.201	0.372	0.073	0.058	-0.092	0.420	0.096	0.223	0.369	0.021	-0.013	0.096	0.006	1.829

Figure 14: News Themes Spanning the Characteristic Zoo

Note. Cells report news themes' Shapley contributions to refit R^2, averaged within JKP themes, in percentage points. Selected factors enter jointly by theme. Values are rounded to whole percentage points, with cells rounding to zero left blank. Row totals give mean R^2. The bottom row instead reports contributions to annualized portfolio Sharpe ratios from Figure 8, holding MSRR weights fixed across all theme subsets. Panels A and B use equal and value weighting. Regressions match Figure 20.

5 Market Reactions to News Topics

In behavioral finance, a natural question is whether markets underreact or overreact to news, particularly to specific news topics. An important first step in this direction is Kwon and Tang (2025). They group news into 24 categories and document both underreaction and overreaction by markets. A fundamental limitation of their methodology is that it represents each announcement through a single news category. Our methodology suggests that this assumption may not be borne out by the data: many news articles load on several economically distinct topics.[27]

Figure 5 in Section 3.1 illustrates this overlap: the same article activates several economically distinct topics. Assigning it to only one topic would discard relevant content and attribute its entire return response to the chosen category.

These topic activations are continuous, allowing an article to load strongly on several topics and weakly or not at all on others. This gives us a richer basis for studying market reactions: we can use the same article in several topic-specific analyses while allowing its relevance to differ across topics. We use this representation to ask which kinds of news are followed by continuation or reversal, how these responses develop over time, and which topic characteristics explain the differences. While Kwon and Tang (2025) document both underreaction and overreaction across news categories, our results show predominantly underreaction.

5.1 Measuring the Response to News

We ask whether prices continue moving in the direction of the initial news-day response or subsequently reverse. Following Kwon and Tang (2025), we relate the firm's subsequent return to its news-day return.

We run a separate regression for each news topic, pooling firm-day observations across firms and dates. Let $\tilde{E}_{i,d,j}$ denote the average article-level activation $\tilde{e}_{a,j}$ of topic j across firm i's news on day d. Here we average over the day's articles, whereas portfolio construction uses the preceding six months. Each observation pairs the firm's news-day return with its cumulative return over the following h trading days. We estimate the response slope by

[27] Kwon and Tang (2025) also examine overlapping announcements. Our distinction concerns overlapping economic content within an individual article.

weighted least squares:

$$\left(\widehat{\boldsymbol{\alpha}}_j(h), \widehat{\beta}_j(h)\right) = \underset{\{\alpha_m\},\beta}{\arg\min} \sum_d \sum_i \underbrace{\frac{\tilde{E}_{i,d,j}}{\sum_\ell \tilde{E}_{\ell,d,j}}}_{\omega_{i,d,j}} \left(R^{+h}_{i,d} - \alpha_{m(d)} - \beta r^0_{i,d}\right)^2. \tag{15}$$

The sums cover eligible dates and firms. The weight $\omega_{i,d,j}$ is the firm's share of that day's topic activation. Firm-days whose news activates the topic more strongly receive greater weight, and each eligible day receives total weight one. A firm-day can enter several topic regressions with different weights.

The news-day return $r^0_{i,d}$ and subsequent cumulative return $R^{+h}_{i,d}$ are both adjusted for firm characteristics and industry. A positive $\beta_j(h)$ indicates continuation, consistent with underreaction, while a negative coefficient indicates reversal, consistent with overreaction. For example, a slope of 0.05 means that a one-percentage-point initial response is followed by another five basis points in the same direction.

Each regression includes calendar-month intercepts, and standard errors are clustered by firm and month.[28] Appendix M provides the weighting and sample-construction details.

5.2 Return Dynamics across News Themes

Figure 15 traces the return response for each of the thirteen news themes over horizons from one to 120 trading days. Shading shows 95% confidence bands that hold jointly across the nine horizons of each theme.[29]

Even after the news-day reaction, an economically and statistically significant component of the response remains delayed. For *earnings news*, this continuation adds another 9.3% to the initial response over the following 120 trading days, accounting for 8.5% of the combined move. The delayed response also extends the initial reaction by 7.7% for *guidance*, 7.8% for *analyst news*, 6.6% for *credit and debt financing*, and 7.5% for *equity issuance and payouts*.

Legal, regulatory and misconduct news shows early reversal, statistically significant from day 2 through day 20. The estimates remain near the same level at longer horizons,

[28] We do not include firm fixed effects. Subsequent returns can contain the initial returns of a firm's later news observations, so demeaning by firm can introduce a negative association between initial and subsequent returns, analogous to the finite-sample problem in Nickell (1981). This concern is relevant when a firm's news observations are few and clustered in time.

[29] We sum the raw activations of each theme's eligible topics and normalize across firms within the day. We estimate a separate weighted least squares regression for each theme and horizon, with calendar-month intercepts. The sample is held fixed across horizons, covering news days from January 1996 to July 2025 with a complete 120-day window.

reaching 4.4% of the initial response at 120 days, but are no longer statistically significant. *Markets, financial sector and macro news* also reverses early, but its 120-day response is not statistically significant. *Industry operations and products* shows estimated continuation of 2.7% over 120 days, although the estimates are insignificant at every horizon. For the remaining themes, the 120-day estimates do not distinguish continuation or reversal from zero at the 5% level.

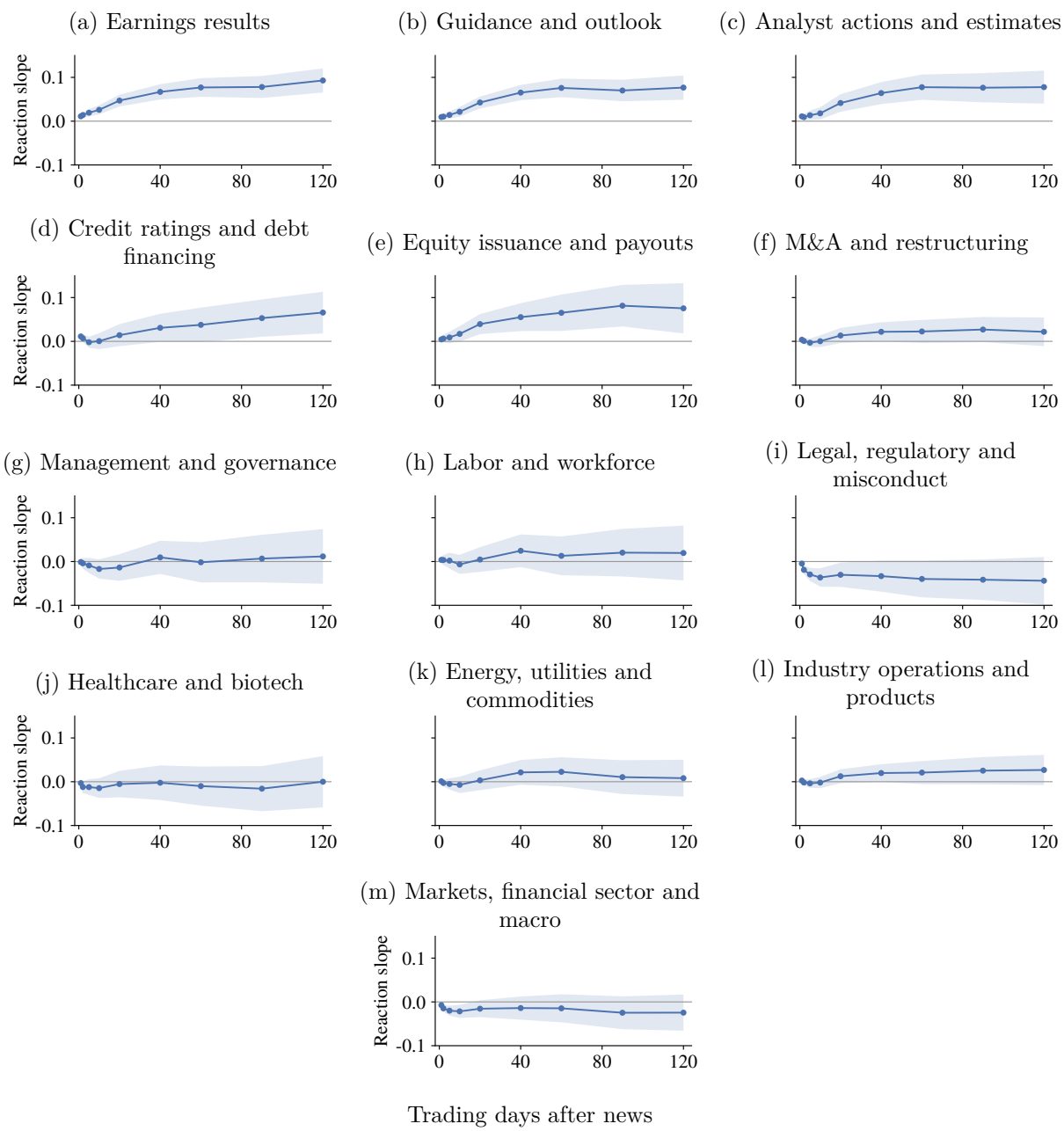


Figure 15: Return Dynamics across News Themes

Note. Slopes of subsequent cumulative residual returns on news-day returns at nine horizons from 1 to 120 trading days. Regressions include month effects and theme-activation weights normalized within day. Shading shows 95% simultaneous confidence bands across horizons within each theme, using firm- and month-clustered errors. Missing returns enter as zero and cumulative returns are winsorized by horizon. Positive slopes indicate continuation: 0.05 means five basis points per one-percentage-point initial return. January 1996–July 2025.

5.3 Return Continuation across Individual Factors

We then repeat the procedure for individual news factors, estimating a separate response regression for each of the 228 factors. We use a common horizon of $h = 60$ trading days and write $\beta_j = \beta_j(60)$. The sample runs from January 1996 to October 2025 and requires a complete 60-day window. Figure 16 reports the slopes' t-statistics, ordered from lowest to highest, with dotted lines at the two-sided 5% significance thresholds.

Underreaction predominates across individual factors. Of the 228 slopes, 164 (71.9%) are positive, and 90 (39.5%) are significantly positive. Only 13 (5.7%) are significantly negative. For the median factor, 3.7% of the combined response unfolds over the 60 trading days following the news-day reaction.

In line with the theme-level results in Figure 15, earnings, guidance and analyst news account for 55 of the 90 significant continuation slopes, although they contain only about a third of the factors. Continuation is significant for 24 of the 31 earnings factors, 19 of the 22 guidance factors and 12 of the 21 analyst factors. Legal and healthcare news account for ten of the 13 significant reversals. The broad theme patterns therefore recur among the individual factors.

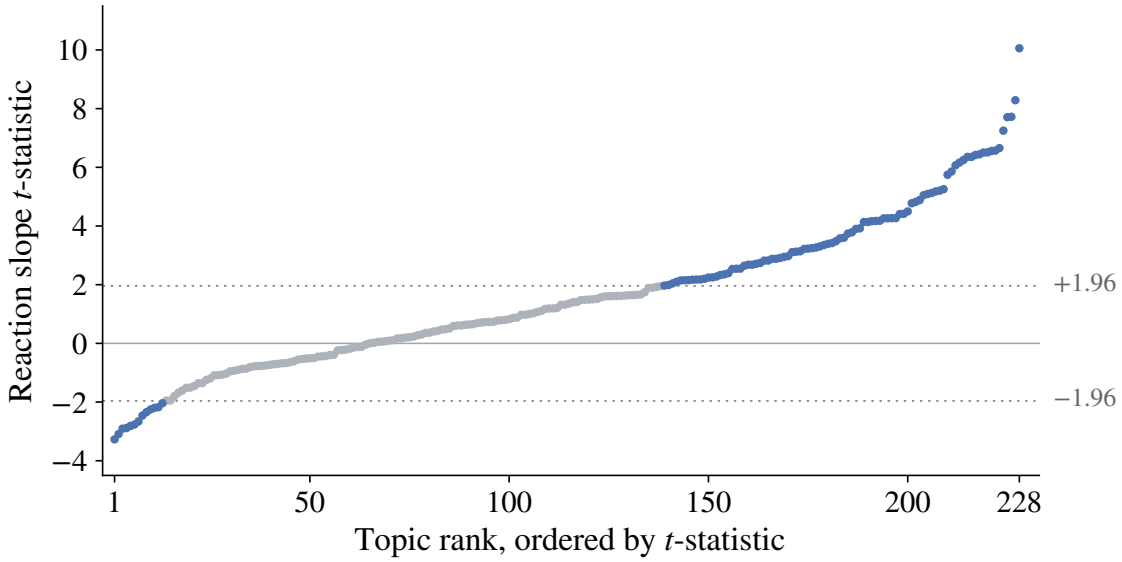


Figure 16: Return Continuation across Individual News Factors

Note. Ranked t-statistics of 228 factor-specific 60-day response slopes in (15). Positive values indicate continuation, negative values reversal. Dotted lines mark two-sided 5% thresholds. Blue points are significant and gray points are not. Errors are clustered by firm and month. January 1996–October 2025.

5.4 Behavioral Channel

Having established significant underreaction to news, we now examine whether different behavioral channels drive these response patterns across news topics. We estimate

$$\hat{\beta}_j = c + \gamma_1 \,\text{NegTone}_j + \gamma_2 \,\text{Uncertainty}_j + \gamma_3 \,\text{QuantShare}_j + \gamma_4 \,\text{Fog}_j + \gamma_5 \,\text{Attention}_j + \gamma_6 \,\text{Extremeness}_j + u_j. \tag{16}$$

The dependent variable is the slope estimated separately for each individual topic using equation (15). We standardize all six explanatory variables across topics, so each coefficient measures the change in the response slope associated with a one-standard-deviation difference in that characteristic. The text-based measures use each topic's 100 most strongly activating full articles:

- *Negative tone* (NegTone_j) is the FinBERT probability of negative tone (Araci, 2019).
- *Uncertainty* (Uncertainty_j) is the share of words in the Loughran–McDonald uncertainty dictionary (Loughran and McDonald, 2011).
- *Quantitative share* (QuantShare_j) is the fraction of numeric tokens.
- *Textual complexity* (Fog_j) uses the fog index, which combines sentence length and the share of words with three or more syllables to measure how difficult a text is to read.
- *Attention* (Attention_j) measures prior news coverage of the firms associated with the topic. For each firm-day, we take the log of one plus the firm's average monthly article count over the preceding twelve months, then average across the unique firm-days associated with the topic's top 10% of activating articles.
- *Extremeness* (Extremeness_j) is the tail parameter of absolute news-day residual returns, estimated using the same firm-days as attention. A larger value indicates a fatter tail.

Appendix M provides the construction details.

Table 3 reports the cross-topic regressions. In line with Kwon and Tang (2025), we find that more extreme news exhibits less continuation. Extremeness is the strongest predictor, explaining 33% of the variation in response slopes. The effect is economically large: a one-standard-deviation increase in extremeness reduces continuation by slightly more than its average level, effectively eliminating the typical delayed response. The estimated reduction

is 3.38 basis points per percentage point of initial return ($t = -10.46$), with a similar effect in the joint regression (-3.14, $t = -8.17$).

Topics about heavily covered firms also continue less. Greater coverage attracts investor attention (Fang and Peress, 2009), which can amplify the initial response and lead to reversal (Hou et al., 2025). In line with this channel, the attention coefficient is -2.26 on its own ($t = -5.54$) and remains negative and significant in the joint regression (-1.01, $t = -2.48$).

Quantitative content has the opposite relationship. Experiments find that statistics are remembered less durably than stories, while evidence from corporate news and analyst reports links quantitative information to delayed market reactions (Graeber et al., 2024; Hong, 2026; Ke, 2025). A one-standard-deviation increase in quantitative share raises the continuation slope by 1.81 basis points in the univariate regression ($t = 5.31$). The effect falls to 0.64 in the joint regression ($t = 1.94$), where it is significant at the 10% level.

Textual complexity also matters in the joint regression. Difficult text can take longer to process and generate greater underreaction (You and Zhang, 2009). The fog index is insignificant on its own but positive when the other characteristics are included (1.31, $t = 4.60$), linking harder-to-read news to greater continuation.

Negative tone has been shown to predict underreaction at short horizons (Tetlock et al., 2008). Appendix N shows that this holds in our setting at short horizons: over the next day, topics with more negative tone show significantly more continuation, and the effect disappears within a few days. At our 60-day horizon, however, negative tone does not explain differences in continuation. Uncertainty, which tests whether investors overinterpret weak, ambiguous signals (Augenblick et al., 2025), is also insignificant separately and jointly. Together, the six characteristics explain 39% of the variation in response slopes, compared with 33% for extremeness alone.

Illustrative topics. Table 4 lists examples for each characteristic that is significant on its own or jointly. We select topics that rank highly on both the characteristic and the response in the direction of its estimated coefficient, skipping near-duplicate labels.

The quantitative examples concern earnings, guidance and analyst estimates. *Corporate earnings and guidance reports* ranks at the 96th percentile of quantitative share and has a slope of 0.132 ($t = 6.25$): a one-percentage-point initial response is followed by another 13.2 basis points of residual return. *Analyst earnings per share estimates* has a slope of 0.111 ($t = 8.29$). The high-fog examples include *company outlooks worsened by weak demand* and *corporate share buyback announcements*, with slopes of 0.103 and 0.101. These topics

Table 3: News Characteristics and Return Continuation across Topics

Cross-topic regressions of 60-day response slopes on news characteristics, separately (1–6) and jointly (7). Text measures average each topic's 100 most activating full articles. Attention measures prior firm coverage. Extremeness measures news-day return tails. Both use the top 10% activating articles. Predictors are standardized. Coefficients measure basis points of subsequent return per one-percentage-point initial return for a one-standard-deviation increase. HC1 t-statistics appear in parentheses. *, **, and *** denote significance at 10%, 5%, and 1%. The sample contains 224 topics, January 1996–October 2025.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Negative tone	+0.30						+0.17
	(+0.76)						(+0.50)
Uncertainty		−0.16					+0.04
		(−0.48)					(+0.12)
Quantitative share			+1.81***				+0.64*
			(+5.31)				(+1.94)
Fog index				−0.35			+1.31***
				(−1.01)			(+4.60)
Attention					−2.26***		−1.01**
					(−5.54)		(−2.48)
Extremeness						−3.38***	−3.14***
						(−10.46)	(−8.17)
R^2	0.00	0.00	0.09	0.00	0.15	0.33	0.39
Topics	224	224	224	224	224	224	224

connect continuation to reports about firm performance, revisions to expectations and payout decisions.

The attention examples include government oversight of financial firms, television programming and privacy disputes. The extremeness examples include litigation outcomes, corporate fraud and executive misconduct, Middle East oil and gas deals, and FDA approvals. *Corporate fraud and executive misconduct* has a slope of −0.169 ($t = -2.46$), while *FDA drug approval and regulatory milestones* has a slope of −0.070 ($t = -2.34$). These are categories with occasional large news-day responses. Their negative slopes illustrate the association between extremeness and reversal.[30]

[30]The examples are selected to illustrate the estimated relationships, not to provide independent tests. Some individual slopes are insignificant at 5%, as the table's t-statistics indicate.

Table 4: Illustrative Topics by News Characteristic

Four illustrative topics for each characteristic significant at 5% in Table 3. Selection maximizes the lower of the characteristic percentile and the response percentile in the predicted direction. Direction follows the significant univariate coefficient, or the joint coefficient otherwise. Near-duplicate labels are omitted. β_j and $t(\beta_j)$ report the 60-day response slope and its t-statistic. Labels are in-sample.

Topic	Value	p	β_j	$t(\beta_j)$	Response p
Quantitative share (continuation)					
Corporate earnings and guidance reports *(f14)*	0.208	96	+0.132	+6.25	100
earnings results and sales reports *(f55)*	0.277	99	+0.113	+3.92	96
Analyst earnings per share estimates *(f51)*	0.191	96	+0.111	+8.29	95
non-GAAP earnings per share reports *(f3)*	0.158	93	+0.109	+6.44	94
Fog index (continuation)					
company outlooks worsened by weak demand *(f105)*	20.3	93	+0.103	+6.42	91
corporate share buyback announcements *(f123)*	18.4	87	+0.101	+4.17	89
Quarterly earnings results announcements *(f61)*	18.2	85	+0.099	+5.86	88
earnings results and sales reports *(f55)*	17.9	84	+0.113	+3.92	96
Attention (reversal)					
government oversight of financial firms *(f99)*	3.27	97	−0.107	−1.95	98
television programming and network lineup changes *(f251)*	3.04	96	−0.096	−1.51	96
legal and regulatory disputes over privacy/data *(f185)*	3.22	97	−0.092	−2.66	96
Middle East oil and gas project deals *(f173)*	3.02	96	−0.076	−1.68	94
Extremeness (reversal)					
Corporate litigation and lawsuit outcomes *(f81)*	0.62	98	−0.090	−2.81	96
Corporate fraud and executive misconduct *(f231)*	0.60	95	−0.169	−2.46	100
Middle East oil and gas project deals *(f173)*	0.60	97	−0.076	−1.68	94
FDA drug approval and regulatory milestones *(f131)*	0.62	97	−0.070	−2.34	93

Robustness and interpretation. The full-sample pattern is similar without month fixed effects, which yields 94 significantly positive topic slopes and 13 significantly negative slopes at 60 days. With month effects retained, calendar-time HAC inference with 60 lags yields 92 and 15, respectively. At a 20-day horizon, the firm/month-clustered counts are 93 and 13. Continuation is weaker after 2016: 34 of the 227 estimable topics have significantly positive 60-day slopes and 10 have significantly negative slopes. In separate 60-day theme regressions, seven slopes are significantly positive and one, legal, regulatory and misconduct news, is significantly negative over the full sample, and we reject equality of the 13 theme slopes ($\chi^2(12) = 156.1$, $p < 0.001$). After 2016, four theme slopes are significantly positive and one is significantly negative, using pointwise tests.

The estimates describe return adjustment among firms exposed to a topic. A firm-day can contain several kinds of news, and activation does not isolate the part of its initial return attributable to any one topic. The results therefore connect news characteristics to continuation and reversal without identifying a particular behavioral mechanism or each topic's contribution to portfolio profits. Topic tests share observations, and their significance levels are unadjusted for multiple comparisons.

6 Conclusion

We propose a method for constructing interpretable firm characteristics from financial news. By training a compact sparse autoencoder on point-in-time LLM embeddings, we turn dense representations into economically meaningful exposures while preserving their predictive information. The resulting factors capture recognizable developments at the article, firm, and macroeconomic levels.

Our asset-pricing results support financial news as a richer source of firm characteristics. News-based factors expand the efficient frontier, produce lower pricing errors, and require less turnover than traditional factors. Earnings, analyst, and guidance news account for much of the traditional factors' return variation, while other news themes contribute substantial additional performance. The reverse spanning relationship is considerably weaker.

The same representation also allows us to revisit market reactions to news. Accounting for each article's exposure to multiple topics reveals predominantly underreaction, with delayed responses related to extremeness, investor attention, quantitative content, and textual complexity. Together, these results show how interpretable representations of financial text can broaden empirical asset pricing beyond readily measurable accounting and price characteristics, bringing the information researchers study closer to the information investors read.

References

Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Augenblick, N., Lazarus, E., and Thaler, M. (2025). Overinference from weak signals and underinference from strong signals. *The Quarterly Journal of Economics*, 140(1):335–401.

Bills, S., Cammarata, N., Mossing, D., Tillman, H., Gao, L., Goh, G., Sutskever, I., Leike, J., Wu, J., and Saunders, W. (2023). Language models can explain neurons in language models. OpenAI. https://openaipublic.blob.core.windows.net/neuron-explainer/paper/index.html.

Bricken, T., Templeton, A., Batson, J., Chen, B., Jermyn, A., Conerly, T., Turner, N., Anil, C., Denison, C., Askell, A., et al. (2023). Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*, 2(5):6.

Bybee, L., Kelly, B., Manela, A., and Xiu, D. (2024). Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147.

Bybee, L., Kelly, B., and Su, Y. (2023). Narrative asset pricing: Interpretable systematic risk factors from news text. *The Review of Financial Studies*, 36(12):4759–4787.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *The Critical Finance Review*, 11(2):207–264.

Chen, H., Didisheim, A., Pourmohammadi, M., Somoza, L., and Tian, H. (2026a). A financial brain scan of the llm. *Working Paper*.

Chen, H., Didisheim, A., and Somoza, L. (2026b). Out of the black box: Uncertainty quantification for llms via conditional probabilities. Technical Report 34965, National Bureau of Economic Research.

Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. *The Journal of Business*, 59(3):383–403.

Chen, Y., Kelly, B. T., and Xiu, D. (2024). Expected returns and large language models. Working Paper 4416687, Social Science Research Network. https://ssrn.com/abstract=4416687.

Choi, D., Huang, V., Meng, K., Johnson, D. D., Steinhardt, J., and Schwettmann, S. (2024). Scaling automatic neuron description. Technical report, Transluce. https://transluce.org/neuron-descriptions.

Cunningham, H. et al. (2023). Sparse autoencoders find highly interpretable features in language models. *arXiv preprint arXiv:2309.08600*.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance*, 53(6):1839–1885.

Didisheim, A., Fraschini, M., and Somoza, L. (2025). Ai's predictable memory in financial analysis. *Economics Letters*, 256:112602.

Didisheim, A., Ke, S. B., Kelly, B. T., and Malamud, S. (2024). Apt or "aipt"? the surprising dominance of large factor models. Technical report, National Bureau of Economic Research.

Didisheim, A., Kelly, B., Pourmohammadi, M., and Tian, H. (2026). The inefficient pricing of news.

Elhage, N., Hume, T., Olsson, C., Schiefer, N., Henighan, T., Kravec, S., Hatfield-Dodds, Z., Lasenby, R., Drain, D., Chen, C., Grosse, R., McCandlish, S., Kaplan, J., Amodei, D., Wattenberg, M., and Olah, C. (2022). Toy models of superposition. *Transformer Circuits Thread*.

Fama, E. F. and French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance*, 50(1):131–155.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2018). Trading costs. *Available at SSRN 3229719*.

Gabaix, X. and Ibragimov, R. (2011). Rank $-$ 1/2: A simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Gao, L., la Tour, T. D., Tillman, H., Goh, G., Troll, R., Radford, A., Sutskever, I., Leike, J., and Wu, J. (2024). Scaling and evaluating sparse autoencoders. *arXiv preprint arXiv:2406.04093*.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

Graeber, T., Roth, C., and Zimmermann, F. (2024). Stories, statistics, and memory. *The Quarterly Journal of Economics*.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Gu, S., Kelly, B., and Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, 222(1):429–450.

Hansen, L. P. and Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

Hoberg, G. and Manela, A. (2025). The natural language of finance. *Foundations and Trends in Finance*, 14(4):244–365.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184.

Hong, W. (2026). Selective recall and the story-statistics gap in stock market misreaction. Working paper.

Hou, K., Loh, R., Peng, L., and Xiong, W. (2025). A tale of two anomalies: The implications of investor attention for price and earnings momentum. *Working Paper*.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5):2019–2133.

Huettner, F. and Sunder, M. (2012). Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values. *Electronic Journal of Statistics*, 6:1239–1250.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2022). Is there a replication crisis in finance? *The Journal of Finance*.

Ke, S. (2025). Analysts' belief formation in their own words. *Working Paper, Yale School of Management*. Available at SSRN: https://ssrn.com/abstract=5025830.

Ke, S. and Pourmohammadi, M. (2026). Shrinkage alignment in high-dimensional portfolios. *Working Paper*. Revised April 2026.

Kelly, B., Kuznetsov, B., Malamud, S., and Xu, T. A. (2025). Artificial intelligence asset pricing models. Technical Report 33351, National Bureau of Economic Research.

Kelly, B., Malamud, S., Pourmohammadi, M., and Trojani, F. (2026a). Universal portfolio shrinkage. *The Review of Financial Studies*. Advance Access publication, September 24, 2026.

Kelly, B., Malamud, S., Schwab, J., and Xu, T. A. (2026b). Scaling point-in-time language models: Enabling economic evaluation of embeddings. *arXiv preprint arXiv:2607.11889*.

Kelly, B., Malamud, S., and Zhou, K. (2024). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1):459–503.

Kelly, B. T. and Xiu, D. (2023). Financial machine learning. *Available at SSRN*.

Kwon, S. Y. and Tang, J. (2025). Extreme categories and overreaction to news. *Review of Economic Studies*, page rdaf037.

Lieberum, T., Rajamanoharan, S., Conmy, A., Smith, L., Sonnerat, N., Varma, V., Kramár, J., Dragan, A., Shah, R., and Nanda, N. (2024). Gemma scope: Open sparse autoencoders everywhere all at once on gemma 2. In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pages 278–300.

Lopez-Lira, A. and Tang, Y. (2024). Can chatgpt forecast stock price movements? return predictability and large language models. *Journal of Financial Economics*. Forthcoming.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Loughran, T. and McDonald, B. (2020). Textual analysis in finance. *Annual Review of Financial Economics*, 12(1):357–375.

Montiel Olea, J. L. and Plagborg-Møller, M. (2019). Simultaneous confidence bands: Theory, implementation, and an application to SVARs. *Journal of Applied Econometrics*, 34(1):1–17.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.

OpenAI (2024). New embedding models and API updates. OpenAI blog, January 25, 2024.

Sarkar, S. K. and Vafa, K. (2024). Lookahead bias in pretrained language models. SSRN Working Paper No. 4754678.

Templeton, A., Conerly, T., Marcus, J., Lindsey, J., Bricken, T., Chen, B., Pearce, A., Citro, C., et al. (2024). Scaling monosemanticity: Extracting interpretable features from claude 3 sonnet. *Transformer Circuits Thread*.

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

You, H. and Zhang, X.-j. (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting Studies*, 14(4):559–586.

A From News Embeddings to Tradable Factors

Figure 17 summarizes the complete pipeline. We first embed each news article using a point-in-time LLM checkpoint trained before the evaluation period. We train the SAE exclusively on pre-2016 news embeddings, choosing its dimension and sparsity penalty to balance predictive performance and interpretability. We then freeze the SAE and use it to encode articles in both the training and evaluation periods. Each article receives a sparse vector of topic activations.

These activations enter two parallel steps. To build portfolios, we average them across each firm's articles over the preceding six months, standardize the resulting signals, and remove their size and industry components. Each residual signal defines a long-short news factor, as described in Section 2.4. Separately, we label each feature by asking an LLM to identify the common subject of its 20 most strongly activating in-sample headlines and classify its content. Out-of-sample labels serve only to evaluate interpretability. The names used throughout the paper come from the in-sample headlines.

We retain economically interpretable topics and remove firm-specific, textual, and unclear features, leaving 228 news factors from the SAE's 256 coordinates. Finally, maximum Sharpe ratio regression combines the factors into the news portfolio. Labeling gives each factor an economic interpretation, while its returns come from the firms' news exposures and subsequent stock returns.

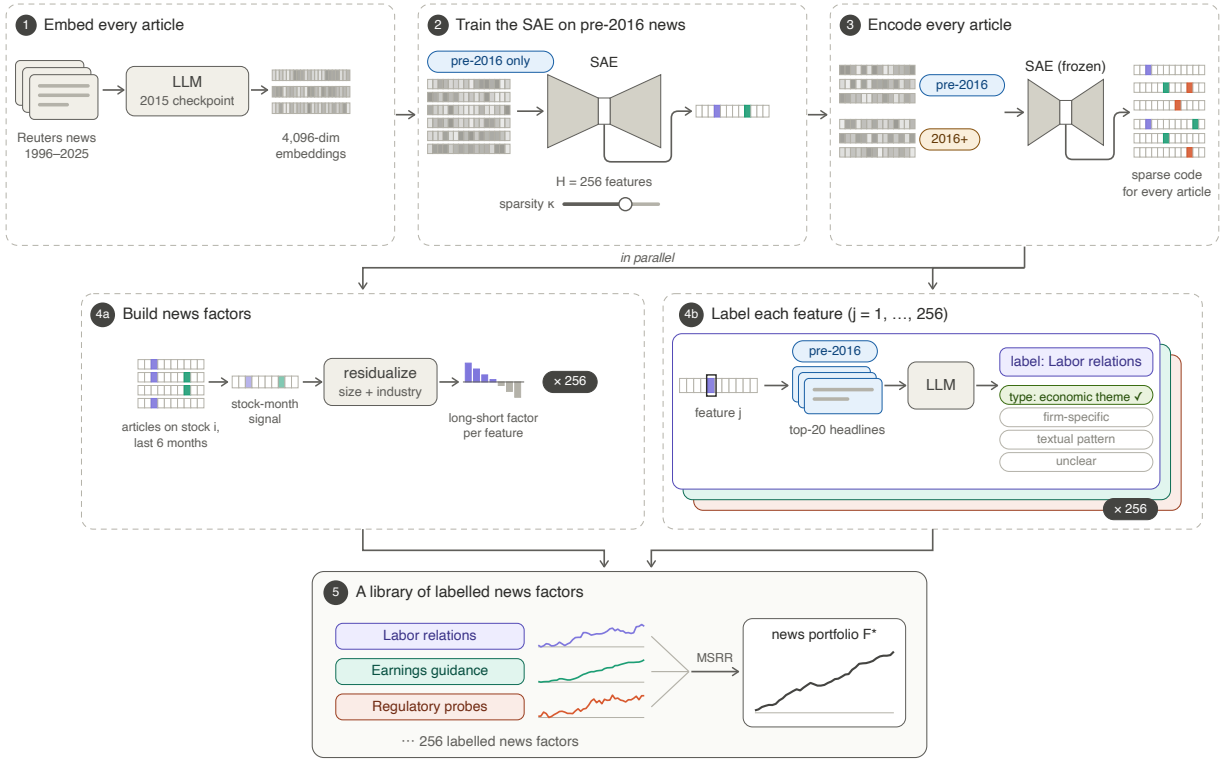


Figure 17: From News Embeddings to Tradable Factors

Note. The figure follows articles through embedding, SAE training, and encoding. The resulting activations feed two parallel steps: forming firm exposures and long-short factors, and assigning economic labels from in-sample headlines. Maximum Sharpe ratio regression (MSRR) combines the retained news factors into a portfolio. The diagram shows all 256 SAE coordinates before the removal of features without an economic interpretation.

B SAE Training Details

Architecture. The SAE in (1) subtracts the decoder bias b_{dec} before encoding, so that it encodes deviations from a learned center of the embedding distribution. Following Bricken et al. (2023) and Lieberum et al. (2024), every column of W_{dec} is rescaled to unit L_2 norm after each optimization step, with the gradient component parallel to the column projected out beforehand. Without this constraint, the SAE could evade the penalty in (2) by shrinking the activations and inflating the dictionary vectors.

PCA initialization. We initialize the decoder at the leading H principal components of the in-sample article embeddings, computed from 100,000 randomly drawn in-sample articles, and the encoder at its transpose, $W_{\text{enc}} = W'_{\text{dec}}$; the two are then trained separately. The decoder bias b_{dec} starts at the mean embedding and the encoder bias b_{enc} at zero. This initialization requires $H \leq D$, which holds for every SAE we train. It has two advantages over random initialization. First, every SAE of a given width starts from the same, data-determined point, so differences across penalty levels reflect the penalty rather than the starting point. Second, the unpenalized SAE starts at its natural solution, the principal-component reconstruction, so that the penalty's effect is measured relative to a well-understood benchmark.

Optimization. We use Adam with a learning rate of 3×10^{-4} and a batch size of 4,096 articles. Rather than training for a fixed number of epochs, we train each SAE until convergence: until the number of active features per article and the reconstruction error on a held-out 5% of the in-sample articles each change by less than 1% and 0.5% per epoch, respectively, for three consecutive epochs, with a minimum of 20 and a maximum of 200 epochs. We keep the checkpoint with the lowest held-out value of the full objective (2).

Seeds. For each combination of width and penalty we train three SAEs, which share the initialization and the data and differ only in the order in which training batches are drawn. We report the SAE whose combined equal- and value-weighted Sharpe ratio is the median of the three, so that both weightings always refer to the same model. Excluding the strongest penalty, the three SAEs of a cell typically differ by 0.08 in equal-weighted and 0.06 in value-weighted Sharpe ratio. The differences are largest near the edge of the penalty grid ($\kappa = 3 \times 10^{-2}$), where results depend on which features survive training.

C Construction of the News Portfolio

This appendix supports the two construction choices of Section 2.4: orthogonalization with respect to size and industry only, and the six-month news window. Both exhibits use the dense embedding, with all 4,096 coordinates as signals, so that the choices are made before, and independently of, any SAE design choice.

Characteristics. Figure 18 varies the set of characteristics S_t in (3), with a one-month news window. The sets are nested: none (the Z-scored embedding, demeaned cross-sectionally), industry indicators only, size and industry, the Fama–French three-factor characteristics, those of the five-factor model plus momentum, and all 132 JKP characteristics, each with industry indicators. Panel A shows that performance rises with richer orthogonalization: the equal-weighted Sharpe ratio rises from 1.8 to 2.2, and the value-weighted Sharpe ratio from 0.9 to about 1.2 once size is controlled for and 1.3 with all 132 characteristics. Panel B shows that the news portfolios earn large alphas against the 13 JKP theme factors and the market under every set: 12–19% per year equal-weighted, significant at the 1% level throughout, and 4–10% value-weighted, significant from size and industry onward. Size and industry capture most of the value-weighted gain, while the remaining gain from richer sets comes at the cost of making the news factors orthogonal to the characteristics against which we later evaluate them.

News window. Table 5 varies the length N of the news window used to construct $\tilde{E}_{i,t}$. To measure turnover, we trace the news portfolio back to the stock-level positions it holds, $w_{i,t}$, and normalize them to one dollar of gross exposure, $\tilde{w}_{i,t} = w_{i,t} / \sum_k |w_{k,t}|$. One-sided turnover is

$$\text{Turnover}_{t+1} = \frac{1}{2} \sum_i \left| \tilde{w}_{i,t+1} - g_{i,t+1} \tilde{w}_{i,t} \right|, \tag{17}$$

where $g_{i,t+1}$ is the return-based growth of the position in stock i, and we charge a trading cost of 10 basis points per dollar traded. With a one-month window, the equal-weighted portfolio turns over 68% of its positions each month and the value-weighted portfolio 57%, and trading costs remove 40% and 79% of their gross Sharpe ratios. Pooling news over six months lowers turnover to 17% and 11% while raising the gross Sharpe ratio, from 2.0 to 2.3 equal-weighted and from 1.2 to 1.7 value-weighted. The net Sharpe ratio peaks at $N = 6$ equal-weighted

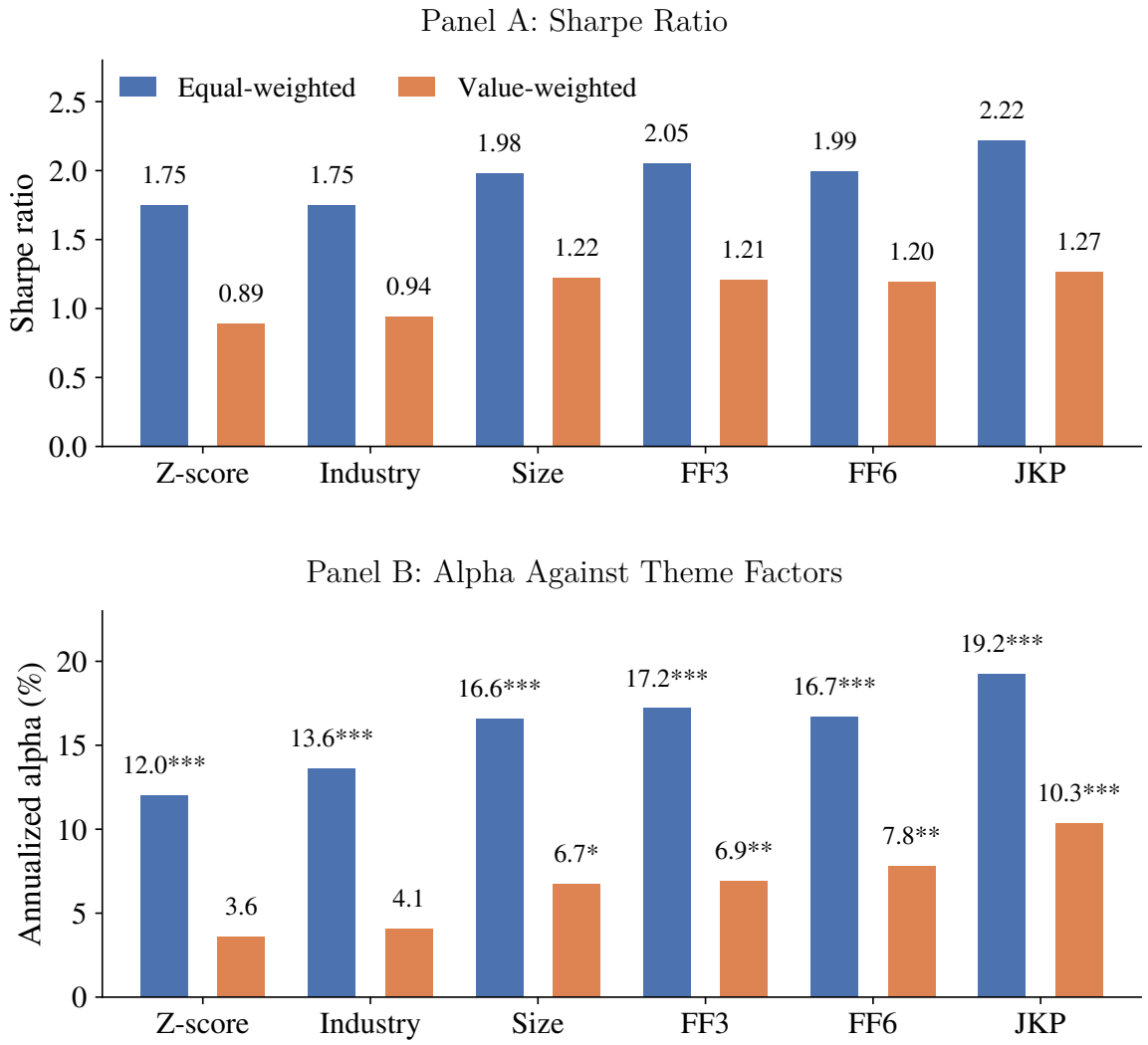


Figure 18: News Portfolio Performance by Orthogonalization Set

Note. This figure reports the out-of-sample performance of the news portfolio built from the dense embedding (all 4,096 coordinates), with a one-month news window, as the set of characteristics in (3) expands from none to all 132 JKP characteristics; every set except the first includes GICS industry indicators, and the second consists of them alone. The bars are labelled Z-score, Industry, Size, FF3, FF6 and JKP, respectively. Blue bars are equal-weighted and orange bars value-weighted. Panel A reports the annualized Sharpe ratio. Panel B reports the annualized alpha from a time-series regression on the 13 JKP theme factors and the market, with every series scaled to 10% annual volatility; *, ** and *** denote significance at the 10%, 5% and 1% levels. The out-of-sample period covers signal months January 2016 to November 2025, with returns realized the following month.

(2.1) and is nearly as high at $N = 6$ as at its peak of $N = 9$ value-weighted (1.5 against 1.6). Longer windows reduce turnover only a little further, and the signal becomes stale.

Table 5: News Portfolio Performance by News Window

This table reports the out-of-sample performance of the news portfolio built from the dense embedding (all 4,096 coordinates) as a function of the length N of the news window, in months, with news orthogonalized with respect to market capitalization and GICS industry indicators. Firms is the average number of stocks in the monthly cross-section. MSRR is the annualized Sharpe ratio of the news portfolio $F^\star$ as constructed. Gross and Net are the annualized Sharpe ratios of the same portfolio rescaled to one dollar of gross exposure each month, before and after a trading cost of 10 basis points per dollar traded following Frazzini et al. (2018), and Turn. is its average monthly one-sided turnover (17). The best net Sharpe ratio in each weighting is in bold. The out-of-sample period covers signal months January 2016 to November 2025, with returns realized the following month.

		Equal-weighted				Value-weighted			
N	Firms	MSRR	Gross	Net	Turn.	MSRR	Gross	Net	Turn.
1	2,659	1.98	1.99	1.20	0.68	1.22	1.24	0.26	0.57
3	3,507	2.10	2.04	1.73	0.26	1.39	1.38	1.01	0.20
6	3,576	2.31	2.32	**2.12**	0.17	1.68	1.67	1.45	0.11
9	3,598	2.25	2.28	2.11	0.14	1.74	1.74	**1.56**	0.09
12	3,611	2.05	2.09	1.94	0.12	1.72	1.71	1.54	0.08
24	3,636	1.57	1.58	1.46	0.11	1.62	1.66	1.52	0.06
36	3,647	1.51	1.52	1.40	0.10	1.49	1.52	1.39	0.06
48	3,655	1.42	1.43	1.31	0.10	1.41	1.44	1.31	0.06
60	3,662	1.44	1.46	1.35	0.10	1.27	1.31	1.19	0.06

D Orthogonalization of the News Embedding

This appendix details the cross-sectional regression (3) that orthogonalizes the news embedding $\tilde{E}_t$ with respect to the characteristics S_t, and the weighted version used for value-weighted portfolios.

Characteristics. The matrix S_t has a constant, market capitalization and GICS industry-group indicators. Market capitalization enters as its cross-sectional rank in month t, mapped linearly to $[-0.5, 0.5]$; a stock without a market capitalization receives zero, the midpoint. Each industry group, defined by the first four digits of the GICS code, has an indicator, and stocks without a GICS code form the omitted group. Industry indicators are not ranked.

Equal-weighted portfolios. For equal-weighted portfolios we estimate (3) by ordinary least squares,

$$\hat{\beta}_t = \left(S_t' S_t\right)^{-1} S_t' \tilde{E}_t, \qquad \tilde{\varepsilon}_t = \tilde{E}_t - S_t \hat{\beta}_t. \tag{18}$$

The normal equations give $S_t' \tilde{\varepsilon}_t = 0$. Because S_t includes a constant, every column of $\tilde{\varepsilon}_t$ has a zero cross-sectional mean and zero equal-weighted covariance with each characteristic, so the equal-weighted news factors in (4) have no equal-weighted exposure to size or industry.

Value-weighted portfolios. For value-weighted portfolios we use the portfolio weights themselves. Let $v_{i,t} = \min(\mathrm{ME}_{i,t},\, \mathrm{ME}_t^{80}) / \sum_k \min(\mathrm{ME}_{k,t},\, \mathrm{ME}_t^{80})$, where $\mathrm{ME}_{i,t}$ is the market equity of stock i and ME_t^{80} is the 80th percentile of NYSE market equity in month t, and let $V_t = \mathrm{diag}(v_{1,t}, \ldots, v_{N_t,t})$. We estimate (3) by weighted least squares,

$$\hat{\beta}_t^v = \left(S_t' V_t S_t\right)^{-1} S_t' V_t \tilde{E}_t, \qquad \tilde{\varepsilon}_t^v = \tilde{E}_t - S_t \hat{\beta}_t^v. \tag{19}$$

The normal equations now give $S_t' V_t \tilde{\varepsilon}_t^v = 0$. With the constant in S_t, the value-weighted mean of each residual column is zero, $\bar{\tilde{\varepsilon}}_{j,t}^v = \sum_i v_{i,t} \tilde{\varepsilon}_{ij,t}^v = 0$, and for every characteristic k

$$\sum_i v_{i,t} \left(\tilde{\varepsilon}_{ij,t}^v - \bar{\tilde{\varepsilon}}_{j,t}^v\right) S_{ik,t} = 0. \tag{20}$$

The value-weighted news factor in (4) holds these positions, so its value-weighted exposure to size and to each industry is exactly zero. Ordinary least squares would not deliver this: its residuals are orthogonal to S_t under equal weights, and their value-weighted exposures

to size and industry are generally different from zero. In the main text, $\tilde{\varepsilon}_t$ denotes $\tilde{\varepsilon}_t^v$ for value-weighted portfolios.

E Labeling Prompt

This appendix reproduces the prompt we use to label the news factors (Section 2.6). We label every feature with OpenAI's `gpt-5.4-mini`, submitted through the Batch API with JSON output. Each feature is labeled twice, from its in-sample headlines (published before 2016) and, separately, from its out-of-sample headlines (2016 onward). In each call, the model receives the system prompt (Prompt 1) and a user message with the feature's $\underline{K} = 20$ most strongly activating alerts in that period, ordered from most to least activating (Prompt 2). The label ℓ_j and type τ_j are the two fields of the returned JSON object.

System Prompt

```
You are analyzing a single latent feature extracted from a large corpus of financial
news headlines. You will receive headlines where this feature has high activation,
ordered from most to least activating.

Your task:
- Identify the economic theme, concept, or phenomenon the headlines share.
- Focus on the underlying economic content, NOT surface-level text patterns (shared
words, ticker symbols, company-name prefixes, or formatting).

Output ONLY a JSON object with exactly two fields:
  "label": a short description of the economic concept (5-10 words; no prefixes like
"This feature detects...").
  "label_type": one of "economic_theme", "firm_specific", "textual_pattern", or
"unclear".
```

Prompt 1: System prompt for labeling a feature.

User Message

Below are 20 news headlines where this feature has high activation, ordered from most to least.

Headlines:

1. ⟨headline 1⟩
2. ⟨headline 2⟩
   ⋮
20. ⟨headline 20⟩

Identify the economic concept this feature represents.

Prompt 2: User message for labeling a feature. The headlines are the feature's 20 most strongly activating alerts in the in-sample or the out-of-sample period.

F The News Factor Library

Table 6 lists the labels of the 228 news factors (Section 3), grouped by theme.

Table 6: News Factor Labels by Theme

#	Label
Earnings results (31)	
0	corporate earnings results announcements
3	non-GAAP earnings per share reports
7	Corporate earnings and business updates
13	earnings per share results versus estimates
17	Quarterly earnings and financial results announcements
24	Corporate quarterly earnings announcements
41	quarterly corporate revenue announcements
47	Corporate earnings and financial results reports
53	earnings and operating results reports
55	earnings results and sales reports
61	Quarterly earnings results announcements
62	corporate earnings and operating results
66	Corporate earnings and financial results reports
67	Quarterly revenue growth reports
70	Small-cap earnings and financing news
76	Quarterly earnings per share announcements
101	technology company earnings and outlook reports
109	Earnings excluding special charges
115	Corporate earnings and sales results
116	Corporate earnings and dividend announcements
119	Quarterly earnings and revenue beats
121	earnings per share results and forecasts
127	Corporate earnings and results reports
128	Corporate earnings and quarterly results
134	company reported operating performance metrics
169	Corporate earnings and financial results
174	Quarterly earnings reports showing net losses

continued on next page

Table 6, continued

#	Label
175	Quarterly earnings results announcements
179	Corporate earnings and quarterly results
181	company operating performance updates
218	earnings reports and guidance for consumer firms
Guidance and outlook (22)	
1	Corporate earnings, forecasts, and guidance
5	Earnings guidance and forecast revisions
8	company forward guidance on growth metrics
14	Corporate earnings and guidance reports
16	Corporate outlook and earnings guidance
26	Forward-looking earnings per share guidance
30	corporate guidance and financial updates
39	Corporate earnings and outlook updates
42	Negative earnings and outlook revisions
54	Forward earnings guidance and estimates
59	Company forward guidance on capex and revenue
73	Forward earnings and quarterly guidance
75	earnings and revenue guidance revisions
83	Company financial guidance and estimates
95	earnings forecast revisions and guidance
105	company outlooks worsened by weak demand
143	earnings guidance revisions and outlook updates
161	Inventory and sales weakness across firms
168	Earnings and guidance misses
182	Corporate performance guidance and growth targets
230	earnings forecast and report corrections
247	Corporate demand and outlook commentary
Analyst actions and estimates (21)	
4	Brokerage analyst raises price targets
19	Credit ratings and analyst recommendation changes
23	analyst earnings per share estimates
27	brokerage analyst price target upgrades

continued on next page

Table 6, continued

#	Label
34	Sell-side analyst rating changes
45	credit ratings and analyst target changes
48	earnings per share analyst estimates
49	earnings guidance and analyst estimate revisions
51	Analyst earnings per share estimates
89	Investment bank analyst ratings and personnel moves
103	Brokerage rating and target revisions
106	Analyst stock rating changes
108	Analyst downgrades of stock ratings
132	analyst price target downgrades
138	Analyst initiations and rating upgrades
151	Analyst stock rating actions and targets
162	analyst price target revisions
196	earnings, guidance, and analyst rating reactions
235	Analyst rating upgrades and downgrades
246	corporate announcements and analyst actions
253	analyst price target revisions
Credit ratings and debt financing (19)	
11	Credit rating and outlook changes
32	Corporate bond issuance and pricing
35	bank capital and debt issuance
38	Credit rating changes and debt distress
58	Corporate financing and credit events
63	Large corporate cash spending and financing
65	Debt refinancing and credit facility amendments
82	Credit ratings and debt issuance
118	corporate restructuring and debt management
144	Corporate credit rating actions
146	corporate financing and balance-sheet actions
149	Debt refinancing and equity issuance
155	Credit rating changes and debt covenants
163	Credit rating downgrades and reviews

continued on next page

Table 6, continued

#	Label
190	Structured debt issuance and pricing
200	corporate debt refinancing and paydown
241	S&P credit rating actions on debt issues
248	Corporate debt issuance and bond pricing
250	Debt refinancing, tender offers, and redemptions
Equity issuance and payouts (9)	
2	Corporate dividend declarations
9	Equity secondary offerings and share issuance
52	Shareholder equity restructuring actions
72	equity issuance and IPO announcements
123	corporate share buyback announcements
125	special and increased shareholder dividends
147	Quarterly cash dividend announcements
148	SEC filing and registration announcements
236	Corporate action deadlines and expiration dates
M&A and restructuring (17)	
12	pharmaceutical M&A, licensing, and litigation
18	corporate restructuring for growth and efficiency
28	Corporate mergers, spin-offs, and restructurings
44	Corporate divestitures and spin-offs
60	utility renewable energy mergers and acquisitions
69	Corporate restructuring and strategic actions
78	Corporate acquisitions and merger deals
80	corporate restructuring and one-time charges
86	Mergers, acquisitions, and corporate restructuring
92	bank mergers and acquisitions
129	M&A, takeovers, and leveraged buyouts
139	regional bank mergers and acquisitions
186	Corporate acquisitions and mergers
198	Insurance brokerage acquisitions and consolidation
212	corporate restructuring and asset divestitures
224	companies facing financial distress or restructuring

continued on next page

Table 6, continued

#	Label
237	Corporate acquisitions and merger bids
Management and governance (8)	
21	Corporate executive appointments, especially CFO hires
113	Corporate executive appointments and departures
117	CEO retirements and succession changes
140	Corporate executive appointments and earnings reports
160	executive and board turnover events
206	shareholder board election results
215	Israeli corporate governance and market activity
245	CEO compensation and executive pay awards
Labor and workforce (5)	
114	Corporate layoffs and workforce reductions
157	union labor negotiations and contract ratification
194	labor union negotiations and disputes
201	Canadian labor disputes and restructuring
221	Corporate layoffs and workforce reductions
Legal, regulatory and misconduct (11)	
68	corporate accounting irregularities and fraud
81	Corporate litigation and lawsuit outcomes
99	government oversight of financial firms
141	tobacco industry litigation and regulation
153	Regulatory approvals for casino and development projects
172	corporate litigation and legal judgments
185	legal and regulatory disputes over privacy/data
191	Regulatory compliance and quality-control issues
199	Regulatory investigations and enforcement settlements
231	Corporate fraud and executive misconduct
255	EU antitrust and regulatory actions
Healthcare and biotech (9)	
79	FDA drug approval and review milestones
85	Biotech clinical trial results and drug development

continued on next page

Table 6, continued

#	Label
97	Biotech and pharmaceutical company news
131	FDA drug approval and regulatory milestones
142	biotech clinical trial results and regulatory milestones
176	FDA approvals for medical devices
202	Medicare/Medicaid managed care contracts and expansions
203	Biotech drug development and clinical trials
214	pharmaceutical trial results and safety
Energy, utilities and commodities (18)	
36	Regulatory approval of utility asset deals
46	nuclear reactor outages and restart updates
87	Refinery outages and maintenance disruptions
90	Industrial materials price increases and supply disruptions
96	metal and mineral mining production guidance
104	Latin American energy and financial deals
120	Regional utility and infrastructure operations
126	utility service restoration after storms
159	oil and gas sector trading activity
173	Middle East oil and gas project deals
180	California utility regulation and power disputes
183	oil and gas production volumes
205	oil and gas drilling well operations
216	commodity futures margin changes
227	Indonesian resource project disputes and stake deals
232	Utility power generation and regulated energy assets
238	natural resource extraction and energy projects
239	utility-scale solar power project development
Industry operations and products (40)	
25	Semiconductor process node and chip manufacturing advances
31	same-store sales performance in retail chains
57	Chinese regional expansion and local approvals
64	Airline capacity, traffic, and merger activity
77	property and casualty insurance losses

continued on next page

Table 6, continued

#	Label
88	commercial real estate portfolio sales and financing
93	enterprise software and IT services launches
111	media content licensing and distribution deals
122	Retail expansion and promotional pricing
124	mobile communications product launches and partnerships
130	rail freight capacity and pricing conditions
133	GMO crop approvals and seed commercialization
135	Retail same-store sales performance
150	Vehicle safety recalls and defect remediation
152	hotel expansion and portfolio growth
154	ship acquisitions, charters, and newbuild orders
158	telecommunications pricing and network expansion
167	Technology hardware and equipment firms
170	Pricing changes and fee schedules
187	Defense logistics and support contracts
189	corporate announcements and capital investments
195	restaurant expansion and same-store sales growth
197	Defense and aerospace contract awards
207	Insurers reporting underwriting and earnings growth
208	internet platform user growth and engagement
211	Airline industry distress and restructuring
213	Consumer beverage product launches and brand strategy
217	Industrial plant openings and closures
219	meat processing industry restructuring
220	apparel brand licensing and product expansion
222	Corporate contracts, licensing, and labor agreements
225	Transportation and industrial accidents/disruptions
226	Insurance company quarterly earnings and premiums
228	newspaper publishing business restructuring
233	Food recalls and product safety actions
234	Commercial aircraft order announcements
240	Pay-TV carriage agreements and subscriber growth
242	Corporate expansion and capital investment announcements

continued on next page

Table 6, continued

#	Label
244	U.S. automaker operations and performance
251	television programming and network lineup changes
Markets, financial sector and macro (18)	
10	Market order imbalance at open
15	government intervention and fiscal support
20	Single-stock intraday price moves
22	opening auction order imbalance
50	company announcements and corporate events
74	Safe-haven Treasury and market reaction moves
84	Pre-open and after-hours stock price moves
91	Government ownership and bailout restructurings
100	Economic conditions affecting hiring and sales
110	Foreign investment and state asset transactions
112	Investment bank and capital markets news
156	Asset managers and institutional investing flows
165	options and derivatives trading activity
166	exchange-traded derivatives volume and launches
193	changes in economic activity indicators
209	post-crisis bailout and restructuring of financial firms
223	Cross-border financial services and investment activity
249	Cautious recovery from recession

G Individual Firm Exposures: Intel

We illustrate the same changes in firm exposures with Intel (Figure 19). The *litigation* feature in Panel B places Intel in the top decile around its European antitrust appeal in 2016 and 2017, lawsuits over security flaws in 2018, retirement-plan litigation in 2019 and 2020, and the 2021 patent judgment requiring payment to VLSI. These events concern different parts of Intel's business but share a clear legal dimension. Panel D similarly connects the *earnings-and-guidance* feature to Intel's weak three-year outlook in 2019, the delay to its chip technology in 2020, and its sharp share-price decline in August 2024.

Other panels show why the labels should be read as summaries of the activating news rather than exhaustive definitions. The *misconduct* feature in Panel A captures reports about stolen deal information and the CEO's resignation following an internal investigation, but also activates on less direct matches. The *EU-antitrust* feature in Panel C captures Intel's antitrust appeal, while its later activations also include European chip partnerships and factory discussions. Its empirical scope is therefore broader than its label suggests. The figure also flags an article about U.S. tariff revenue that the news feed assigns to Intel even though it is not about the firm. Together, these examples show how the exposures track changing news content while making the remaining limits of topic labels and source-level firm assignments visible.

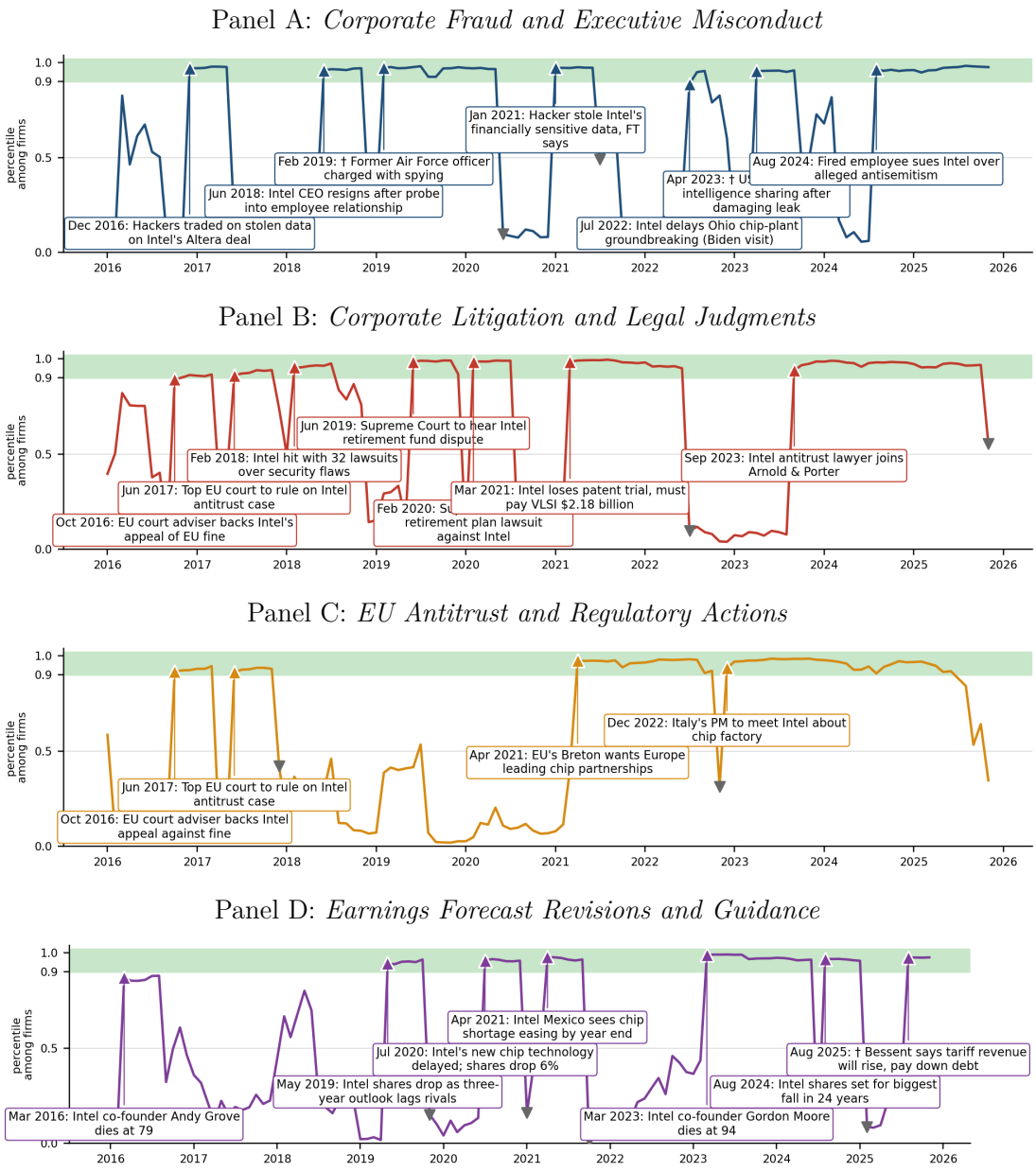


Figure 19: News-Driven Exposures of Intel to Individual News Factors

Note. Each panel follows Intel's monthly percentile among firms on one news factor's portfolio signal. The green band marks the top decile. Upward triangles mark entries into that decile, with paraphrased headlines identifying the most strongly activating article in that month or the preceding two months. Downward triangles mark falls from above the 90th to below the 60th percentile. A dagger identifies an article assigned to Intel that is not about the firm. The sample runs from January 2016 to November 2025.

H Characteristic Factors and Turnover

This appendix provides the construction and measurement details for the comparison in Section 4.

Characteristic factors. Both sets use the same stock returns and monthly rebalancing frequency. For equal-weighted characteristic factors, we rank-standardize characteristic k cross-sectionally each month to $[-\frac{1}{2}, \frac{1}{2}]$ (Gu et al., 2020), giving signals $c_{ik,t}$, and form

$$G_{k,t+1} = \sum_i c_{ik,t} R_{i,t+1}, \tag{21}$$

where $R_{i,t+1}$ is the stock's excess return. Value-weighted factors go long the top quintile and short the bottom quintile of stocks by the characteristic, value-weighted within each leg:

$$G_{k,t+1} = \sum_{i \in \mathcal{L}_{k,t}} \frac{v_{i,t}}{\sum_{l \in \mathcal{L}_{k,t}} v_{l,t}} R_{i,t+1} - \sum_{i \in \mathcal{S}_{k,t}} \frac{v_{i,t}}{\sum_{l \in \mathcal{S}_{k,t}} v_{l,t}} R_{i,t+1}, \tag{22}$$

where $\mathcal{L}_{k,t}$ and $\mathcal{S}_{k,t}$ denote the two quintiles and $v_{i,t}$ are the capped value weights of Section 2.4. We aggregate the characteristic factors by MSRR (5) into a *characteristic portfolio* $G^{\star}_{t+1}$, and form a *combined portfolio* by applying MSRR to all 360 news and characteristic factors.

Metrics. We report annualized Sharpe ratios and average monthly turnover over January 2016 to November 2025. Following Didisheim et al. (2026), we normalize positions to one dollar of gross exposure, $\tilde{w}_{i,t} = w_{i,t} / \sum_l |w_{l,t}|$, and calculate turnover as

$$\text{Turnover}_{t+1} = \frac{1}{2} \sum_i \left| \tilde{w}_{i,t+1} - g_{i,t+1} \tilde{w}_{i,t} \right|, \tag{23}$$

where $g_{i,t+1}$ is the return-based growth of the position in stock i. For the news, characteristic and combined portfolios, turnover is computed from their aggregate stock-level positions.

I Construction of the Stochastic Discount Factors

For the SDF in (10), f_{t+1} is the out-of-sample return of the news portfolio $F^\star$ or of the characteristic portfolio $G^\star$. Following Kelly et al. (2026a), we set

$$\hat{c} = \frac{\overline{f}}{\overline{f^2}}, \tag{24}$$

where bars denote averages over the T out-of-sample months. The coefficient $\hat{c}$ minimizes $\overline{(1 - cf)^2}$ for the given portfolio return. We estimate one coefficient for each portfolio and weighting. The portfolio returns are out of sample, but $\hat{c}$ is estimated on the same months, so the comparison is not a fully out-of-sample test of a calibrated SDF.

Hansen–Jagannathan distance. The test assets are the characteristic factors $G_{k,t+1}$, with equal-weighted SDFs pricing equal-weighted factors and value-weighted SDFs pricing value-weighted ones. We scale each factor to unit second moment over the out-of-sample period and stack them in G_{t+1}. The pricing errors and second-moment matrix are

$$\hat{\alpha} = \overline{M_{t+1} G_{t+1}} = \overline{G_{t+1}} - \hat{c}\, \overline{f_{t+1} G_{t+1}}, \qquad \hat{\Sigma} = \overline{G_{t+1} G_{t+1}'}, \tag{25}$$

The Hansen–Jagannathan distance in (11) uses

$$\hat{\Sigma}_z = (1 - z)\hat{\Sigma} + zI. \tag{26}$$

A lower distance means smaller pricing errors. For the full set of 132 factors we set $z = 0.5$: there are more test assets than out-of-sample months, so the sample second-moment matrix is singular. We also price the factors of each of the 13 JKP themes separately, holding the SDFs fixed. These subsets contain 4 to 20 factors, so we set $z = 0$. We compare SDFs within each set of test assets, rather than comparing distances across sets.

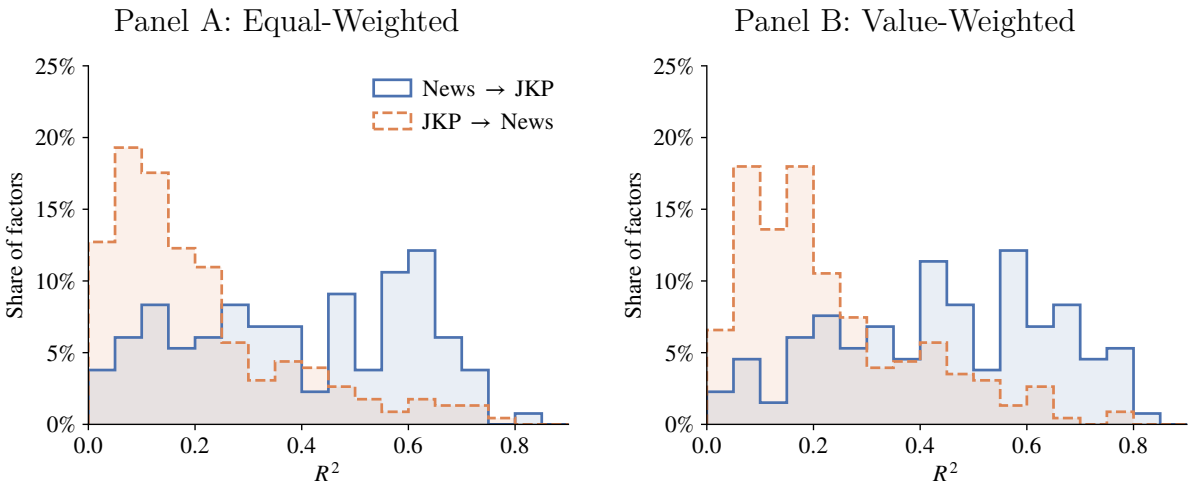


Figure 20: Refit R^2 of the Spanning Regressions

Note. The panels compare the 132 JKP factors explained by five news factors (blue) with the 228 news factors explained by five JKP factors (orange). Regressors are selected jointly across weightings by (12) and held fixed. Panel A is equal-weighted and Panel B is value-weighted. Coefficients are refitted on the evaluation sample. Bars report shares of factors within each target library. The legend identifies the direction of explanation, from the regressor library to the target library. Both panels use the same bins and axis scales. The sample contains 119 signal months, January 2016 to November 2025, with returns realized the following month.

J Aggregate Spanning Results

This appendix reports the aggregate spanning checks for the selection and refit procedure in Section 4.3. The main text focuses on the economic interpretation of individual factor loadings and news-theme contributions.

Explained return variation. Figure 20 compares the two directions. Five news factors selected before 2016 explain a median 39% of a JKP factor's return variation under equal weighting and 45% under value weighting. Five JKP factors explain only 15% and 17% of a news factor's variation. This asymmetry is consistent with the news SDF's lower aggregate pricing distance, although the exercises concern different moments: R^2 measures return variation, while the HJ distance measures pricing errors in average returns.

Unexplained average returns. Figure 21 reports absolute refit alphas. The median is 3.0% per year for news factors versus 1.9% for JKP factors under equal weighting, and 2.5% versus 2.1% under value weighting. JKP factors therefore leave more of the news factors'

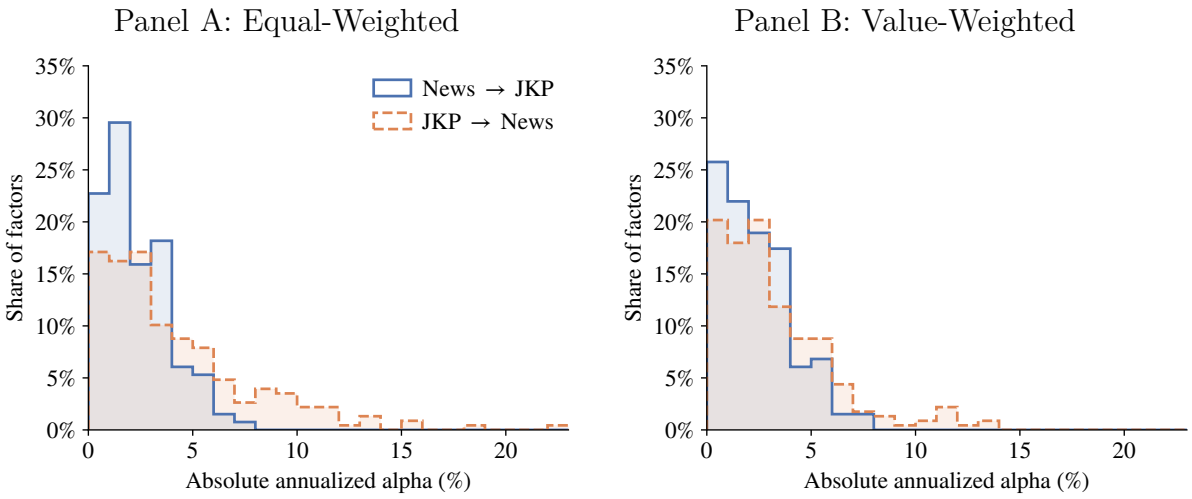


Figure 21: Absolute Refit Alphas of the Spanning Regressions

Note. The panels compare the 132 JKP factors explained by five news factors (blue) with the 228 news factors explained by five JKP factors (orange). Regressors are selected jointly across weightings by (12) and held fixed. Panel A is equal-weighted and Panel B is value-weighted. The horizontal axis reports absolute annualized intercepts of (13), in percent, with all factors scaled to 10% annualized full-sample volatility. Bars report shares of factors within each target library. The legend identifies the direction of explanation, from the regressor library to the target library. Both panels use the same bins and axis scales. Absolute estimates reflect both economic magnitude and estimation noise. The sample is the same as in Figure 20.

average returns unexplained than news factors leave of JKP average returns. Absolute estimates reflect both economic magnitude and estimation noise. A high R^2 can coexist with a large alpha because the selected regressors track month-to-month movements rather than target the mean return directly.

K Rolling Replication of the Characteristic and News Factors

This appendix complements the refit spanning results of Section 4.3. We keep the same five selected regressors for each target and re-estimate their coefficients on the preceding 60 months before each evaluation month:

$$\hat{G}_{k,t+1} = \hat{\alpha}_{k,t} + \sum_{j \in \mathcal{J}_k} \hat{\beta}_{kj,t} F_{j,t+1}. \tag{27}$$

Both sides are portfolio returns realized in month $t+1$, while the replicating weights use information available at t. This is implementable replication rather than a forecast of returns. We evaluate it using (14), which can be negative, and define the rolling alpha as the average

return of the characteristic factor hedged with the replicating portfolio:

$$\alpha_k^{\text{fc}} = \frac{1}{T} \sum_{t \geq T_0} \Big(G_{k,t+1} - \sum_{j \in \mathcal{J}_k} \hat{\beta}_{kj,t} F_{j,t+1} \Big). \tag{28}$$

Its t-statistic uses the time series of hedged returns. The reverse exercise uses the same procedure with news factors as targets. The sample matches Section 4.3.

Rolling fit. Figure 22 shows that the median R^2 falls to 0.24 equal-weighted, positive for 77% of the characteristic factors, and 0.32 value-weighted, positive for 89%. More than half of the spanning is therefore available in real time, and more so value-weighted. The selected news factors are the same in both exercises. The lower rolling fit reflects estimating loadings on a shorter, preceding sample, which introduces estimation error and can also expose changes in the relation between news and characteristic factors.

In the reverse direction, the median rolling R^2 is -0.02 equal-weighted and 0.00 value-weighted, positive for 45% and 50% of news factors. Figure 23 reports absolute rolling alphas: their medians are 3.1% for news targets against 2.0% for characteristic targets equal-weighted, and 2.6% against 2.3% value-weighted.

News-theme contributions. We repeat the Shapley allocation of Figure 14, re-estimating every subset regression on the preceding 60 months at each evaluation date. The empty subset predicts the trailing 60-month mean, so its evaluation-period R^2 need not be zero. Consequently, theme contributions sum to the full rolling R^2 minus this intercept-only baseline. Negative contributions indicate that a theme reduces replication performance on average across orders of inclusion.

Individual factor examples. Table 7 reports refit and rolling estimates for the four Fama–French examples discussed in the main text. The rolling columns report loadings from the last estimation window and fit over the whole evaluation sample.

Table 7: Refit and Rolling Replication of Fama–French Factors with News Factors

Note. Each panel regresses a Fama–French factor on five value-weighted news factors selected before 2016. Refit columns use OLS over the evaluation period. Rolling columns use preceding 60-month windows and report the last window's loadings and full-period replication R^2. All factors are scaled to 10% annualized volatility without demeaning. Alpha is annualized in percent: the refit intercept or the mean rolling hedged return. News labels and feature indices are in-sample. Parentheses report t-statistics. *, **, and *** denote significance at 10%, 5%, and 1%. The evaluation covers January 2016–November 2025, with returns realized the following month.

Panel A: Momentum (UMD)		
	Refit	Rolling
α (% p.a.)	-0.67	-0.71
	(-0.45)	(-0.46)
Negative earnings and outlook revisions *(f42)*	-0.287^{***}	-0.171^{*}
	(-4.53)	(-1.65)
corporate share buyback announcements *(f123)*	$+0.040$	$+0.233^{*}$
	$(+0.43)$	$(+1.65)$
analyst price target downgrades *(f132)*	-0.355^{***}	-0.348^{***}
	(-8.27)	(-6.18)
Corporate credit rating actions *(f144)*	-0.177^{***}	-0.208^{**}
	(-2.81)	(-2.35)
Corporate layoffs and workforce reductions *(f221)*	-0.100^{**}	-0.075
	(-2.19)	(-1.20)
R^2	0.677	0.614
N	119	119

Panel B: Value (HML)		
	Refit	Rolling
α (% p.a.)	$+0.48$	$+0.81$
	$(+0.18)$	$(+0.30)$
technology company earnings and outlook reports *(f101)*	-0.316^{***}	-0.049
	(-2.58)	(-0.28)
Corporate earnings and quarterly results *(f128)*	$+0.553^{***}$	$+0.810^{***}$
	$(+5.14)$	$(+5.71)$
Corporate credit rating actions *(f144)*	$+0.101$	$+0.255^{*}$
	$(+0.89)$	$(+1.72)$
Credit rating changes and debt covenants *(f155)*	$+0.063$	-0.179
	$(+0.53)$	(-1.00)
Biotech drug development and clinical trials *(f203)*	-0.206^{*}	-0.110
	(-1.81)	(-0.72)
R^2	0.530	0.439
N	119	119

Panel C: Profitability (RMW)		
	Refit	Rolling
α (% p.a.)	$+3.24^{*}$	$+2.68$
	$(+1.66)$	$(+1.34)$
Equity secondary offerings and share issuance *(f9)*	$+0.004$	$+0.013$
	$(+0.05)$	$(+0.12)$
earnings per share results versus estimates *(f13)*	-0.238^{**}	-0.165
	(-2.53)	(-1.18)
Quarterly revenue growth reports *(f67)*	$+0.426^{***}$	$+0.381^{***}$
	$(+4.30)$	$(+3.04)$
Quarterly earnings reports showing net losses *(f174)*	-0.467^{***}	-0.576^{***}
	(-7.11)	(-6.36)
Biotech drug development and clinical trials *(f203)*	-0.207^{**}	-0.258^{*}
	(-2.47)	(-1.93)
R^2	0.464	0.351
N	119	119

Panel D: Investment (CMA)		
	Refit	Rolling
α (% p.a.)	$+1.96$	$+4.03$
	$(+0.64)$	$(+1.30)$
technology company earnings and outlook reports *(f101)*	-0.119	-0.153
	(-0.78)	(-0.58)
Analyst stock rating changes *(f106)*	$+0.246^{*}$	$+0.207$
	$(+1.77)$	$(+0.72)$
internet platform user growth and engagement *(f208)*	-0.763^{***}	-1.089^{***}
	(-4.39)	(-3.35)
earnings forecast and report corrections *(f230)*	-0.047	-0.175
	(-0.50)	(-1.16)
Utility power generation and regulated energy assets *(f232)*	$+0.028$	-0.047
	$(+0.20)$	(-0.19)
R^2	0.324	0.151
N	119	119

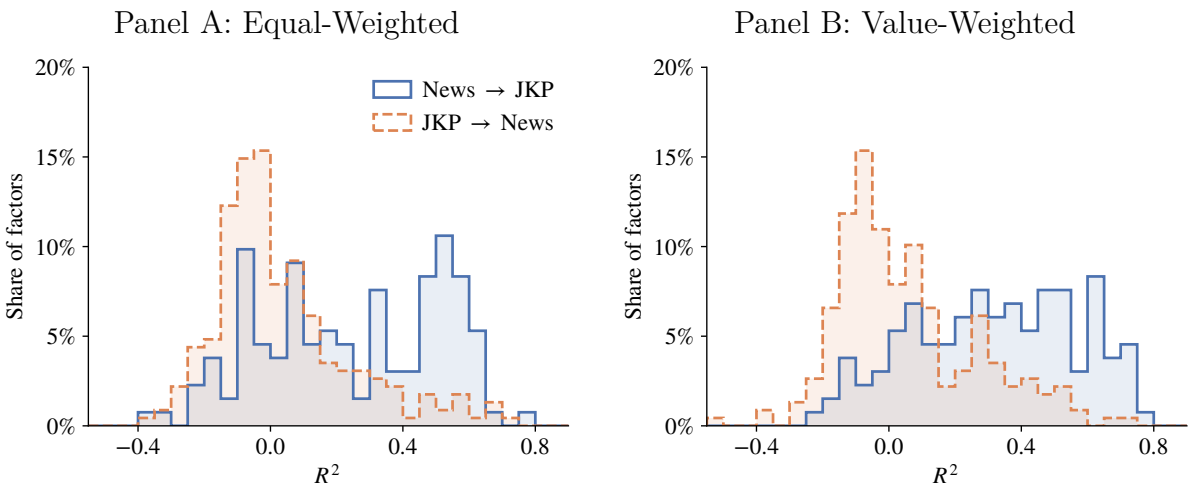


Figure 22: Rolling R^2 of the Spanning Regressions

Note. The panels compare the 132 characteristic factors explained by five news factors (blue) with the 228 news factors explained by five characteristic factors (orange). Regressors are selected jointly across weightings by (12) and held fixed. Panel A is equal-weighted and Panel B is value-weighted. Loadings use the preceding 60 months, as in (27). Bars report shares of factors within each target library. The legend identifies the direction of explanation, from the regressor library to the target library. Both panels use the same bins and axis scales. The sample matches Figure 20.

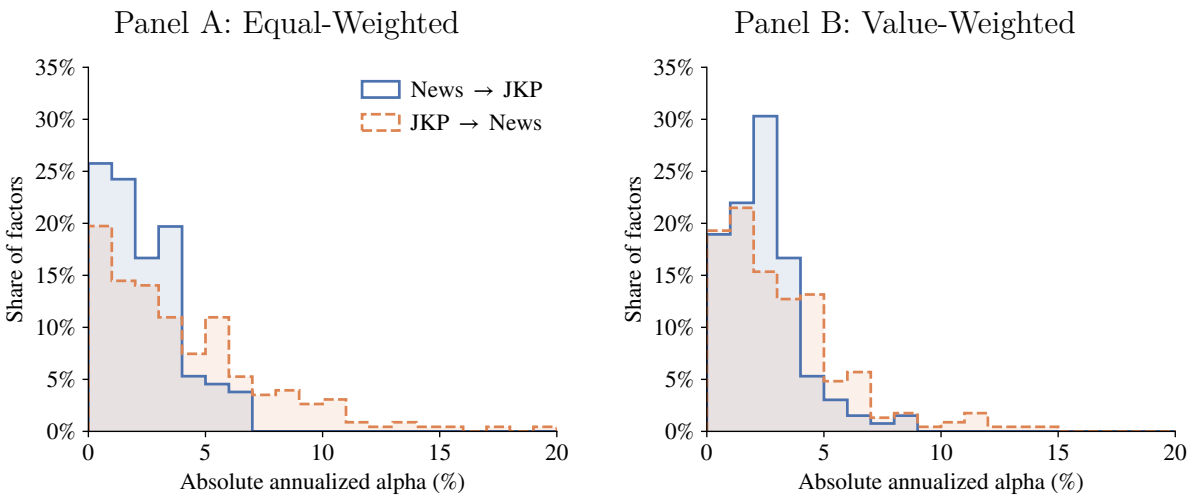


Figure 23: Absolute Rolling Alphas of the Spanning Regressions

Note. The panels compare the 132 characteristic factors explained by five news factors (blue) with the 228 news factors explained by five characteristic factors (orange). Regressors are selected jointly across weightings by (12) and held fixed. Panel A is equal-weighted and Panel B is value-weighted. The horizontal axis reports absolute annualized rolling alphas (28), in percent. Bars report shares of factors within each target library. The legend identifies the direction of explanation, from the regressor library to the target library. Both panels use the same bins and axis scales. Scaling and the sample match Figure 21.

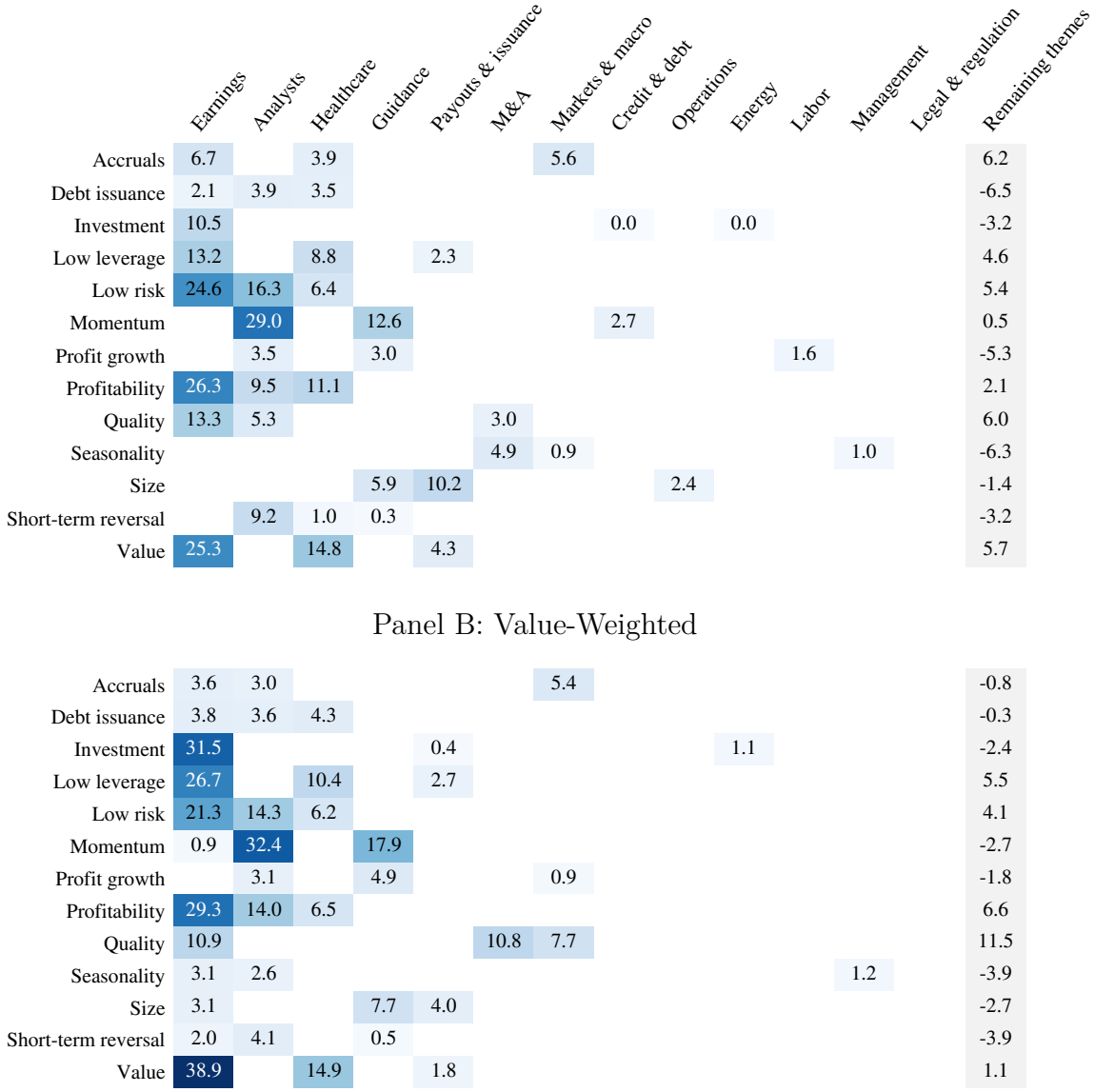


Figure 24: News-Theme Contributions to Rolling Replication

Note. The panels repeat Figure 14 using rolling subset regressions. Rows are JKP themes and columns are news themes. Each row displays its three largest Shapley contributions, in percentage points, rounded to one decimal. Blank cells identify themes included in the signed sum in the gray Remaining themes column. The shared column headings apply to both panels. Contributions are averaged equally across individual JKP factors. Their sum equals mean rolling R^2 minus the rolling intercept-only baseline, rather than total R^2 itself. The sample and selected regressors match the main spanning results.

L Characteristic Themes Spanning the News Themes

Figure 14 decomposes how news themes explain the characteristic factors. Figure 25 reverses the direction. For each news factor, we take the regression on its five selected characteristic factors and group these regressors by JKP theme. As before, a Shapley decomposition assigns each theme its average incremental R^2 over all orders of inclusion, and we average the contributions across news factors within each news theme.

Characteristic factors explain news factors less well than news factors explain characteristic factors. The mean R^2 ranges from 4 to 37 percentage points across news themes under equal weighting, compared with 21 to 59 in Figure 14. No JKP theme dominates: quality, value, low risk, and momentum each contribute to most news themes. Equity issuance and payout news is the best explained, while legal and regulatory news is almost entirely unexplained by characteristics.

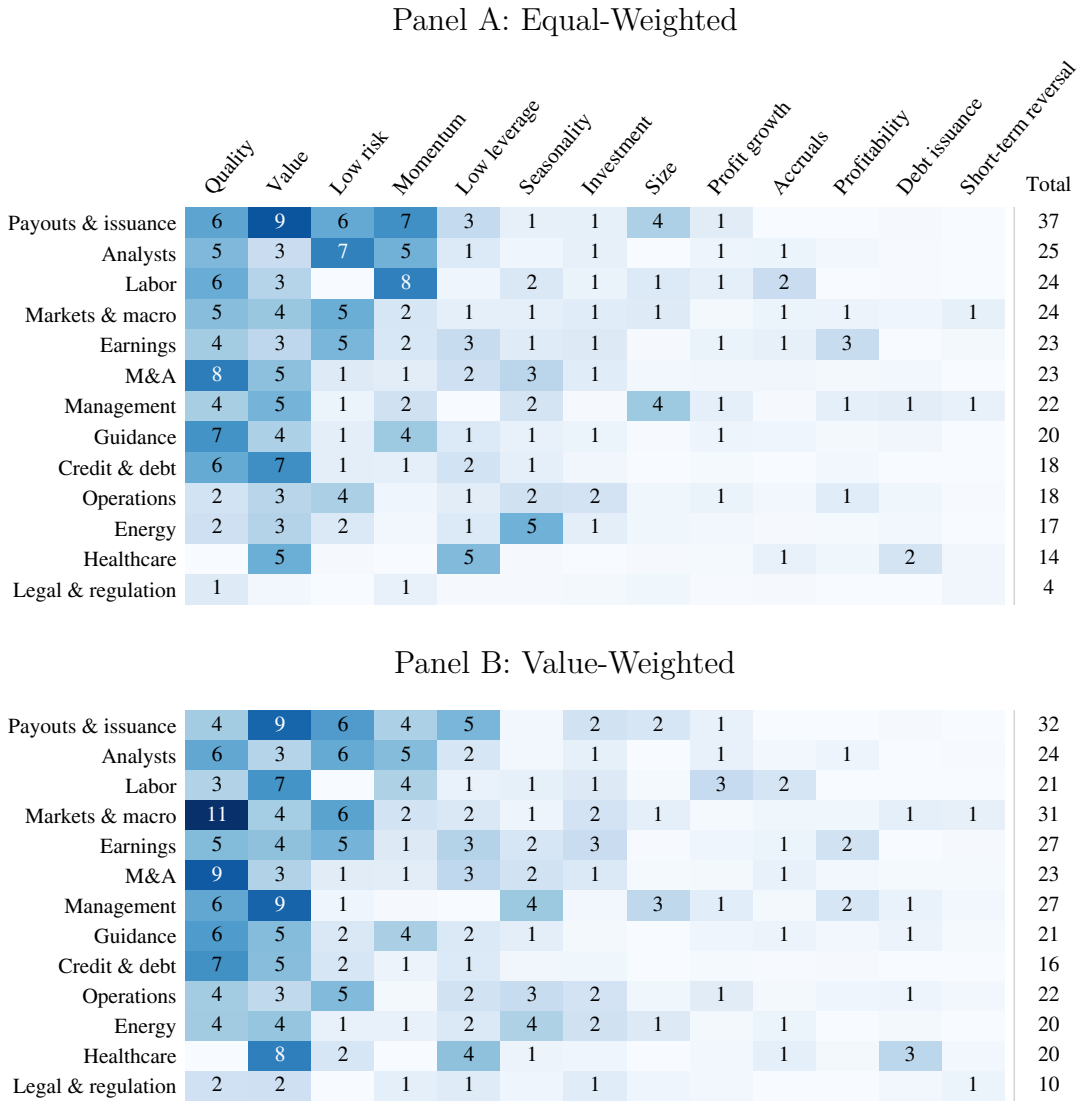


Figure 25: Characteristic Themes Spanning the News Themes

Note. The panels decompose the refit R^2 of the regressions of each news factor on its five selected characteristic factors by JKP theme. For each news factor, selected characteristic factors belonging to the same JKP theme enter together. We refit every subset regression and average each JKP theme's incremental R^2 over all orders of inclusion. We then average contributions equally across news factors within each news theme. Each cell reports the contribution of a JKP theme (column) for a news theme (row), in percentage points; each row sums to the mean individual-factor R^2, reported in the Total column. Values are rounded to whole percentage points, with cells rounding to zero left blank. Shared column headings apply to both panels. News themes are ordered by their equal-weighted R^2 in both panels, and JKP themes by their average contribution. The sample and selected factors match Figure 20.

M Construction of the Firm-Day Reaction Regressions

This appendix gives the construction details for Section 5. We date an article at its effective trading day d: its publication date if it arrives before 16:00 Eastern time on a trading day, and the next trading day otherwise. Let $\mathcal{A}_{i,d}$ denote firm i's articles effective on day d. Its daily exposure to topic j, using the article-level SAE activations from equation (1), is

$$\tilde{E}_{i,d,j} = \frac{1}{|\mathcal{A}_{i,d}|} \sum_{a \in \mathcal{A}_{i,d}} \tilde{e}_{a,j}. \tag{29}$$

The average includes articles with zero activation. We use raw nonnegative ReLU activations without smoothing, standardization or residualization. On an eligible topic-day, the regression weight is

$$\omega_{i,d,j} = \frac{\tilde{E}_{i,d,j}}{\sum_{l \in \mathcal{N}_d} \tilde{E}_{l,d,j}}, \qquad \sum_{i \in \mathcal{N}_d} \omega_{i,d,j} = 1, \tag{30}$$

where $\mathcal{N}_d$ contains firms with news and an observed initial residual return on dates with a complete 60-trading-day calendar window. The topic-day must contain at least five positive-activation firms. Zero-activation firms receive zero weight.

The daily residual return $r^{\perp}_{i,d}$ comes from a cross-sectional OLS regression of close-to-close returns on a constant, the 132 characteristics of JKP, and GICS industry indicators, measured at the preceding month-end. The regression requires a firm record in the preceding-month JKP panel. We rank-standardize each characteristic cross-sectionally to $[-0.5, 0.5]$ and replace missing characteristic values with zero, the cross-sectional median. Before winsorization, initial and subsequent returns are

$$r^{0}_{i,d} = r^{\perp}_{i,d}, \qquad R^{+60}_{i,d} = \sum_{\tau=1}^{60} r^{\perp}_{i,d+\tau}. \tag{31}$$

All baseline topic regressions, including the one-day and 20-day horizons, use firm-days with an observed initial residual return and a complete 60-trading-day calendar window. A complete calendar window means that the horizon ends within the available trading calendar, without requiring a return observation for the firm on every future day. Missing subsequent firm residual returns count as zero. We impose neither a minimum fraction of observed future returns nor a minimum number of news days per firm.

We retain every topic-day with at least five firms with positive activation. Beyond the

data needed to construct the initial residual return, we impose no additional characteristic-history or return-history filters. The sample starts with the news data in January 1996. We winsorize initial and cumulative firm residual returns separately for each horizon at the 0.5% and 99.5% quantiles across unique full-sample firm-days eligible for at least one topic. For horizons exceeding 60 days, cutoffs use only firm-days whose calendar windows extend to that horizon. These cutoffs are common across topics and themes and remain fixed in the post-2016 comparison. The return variables in the reaction regressions denote these winsorized quantities.

For theme g, let $\mathcal{J}_g$ contain its topics and let $I_{d,j}$ indicate topic-day eligibility. We sum raw activations before normalizing across firms:

$$\tilde{E}^{\text{theme}}_{i,d,g} = \sum_{j \in \mathcal{J}_g} I_{d,j} \tilde{E}_{i,d,j}, \qquad \omega_{i,d,g} = \frac{\tilde{E}^{\text{theme}}_{i,d,g}}{\sum_{l \in \mathcal{N}_d} \tilde{E}^{\text{theme}}_{l,d,g}}. \tag{32}$$

Topics are not individually rescaled before this sum. The theme regression is

$$R^{+60}_{i,d} = \alpha_{g,m(d)} + \beta_g r^0_{i,d} + \varepsilon_{i,d,g}, \qquad \text{weight } \omega_{i,d,g}. \tag{33}$$

We retain firm-day returns as separate observations. Each theme has its own month intercepts and slope, so joint estimation gives the same coefficients as separate weighted regressions. We compute the full covariance of the slopes by combining their influence scores within firms and calendar months and subtracting the firm-month intersection term. This accounts for overlapping observations across themes and permits equality tests. Neither the topic nor the theme regressions include firm fixed effects or firm-specific initial-return slopes.

Primary inference uses the uncorrected two-way firm/month clustered covariance and two-sided normal critical values. The sensitivity check aggregates influence scores by trading-calendar day and applies a Bartlett HAC estimator with 60 lags for the baseline regressions and 120 lags for theme dynamics, retaining zero scores on days with no observations. The equality test compares all 13 theme slopes using 12 restrictions and an asymptotic chi-square reference distribution. The individual-topic tests and the original 60-day theme comparisons use unadjusted significance levels. The dynamic theme figure uses simultaneous confidence bands across horizons within each theme, as described below.

Cross-topic regression. The characteristics in Table 3 describe each topic's most representative news over the full sample, January 1996 to December 2025. For the text

characteristics, we rank the topic's full articles, those with a non-empty body, by the topic's SAE activation and keep the 100 most strongly activating ones. We average over them the FinBERT probability of negative tone of the headline and body, the Loughran–McDonald uncertainty word share, the quantitative share and the fog index, defined in Section 5. Four topics have no activating full article and are omitted.

For attention and extremeness, we rank all of the topic's articles with positive activation and keep the top 10%. We map each to its firm and effective trading day, keep each firm-day once, and require an observed news-day residual return $r^{\perp}_{i,d}$. Attention is the average over these firm-days of $\log(1+n_{i,m})$, where $n_{i,m}$ is the firm's average monthly number of Reuters articles over the twelve calendar months before the month of day d. Extremeness follows equation (10) of Kwon and Tang (2025) and Gabaix and Ibragimov (2011): over the 10% largest values of $|r^{\perp}_{i,d}|$ among the topic's firm-days, we estimate

$$\log(\text{Rank}_{i,d} - 0.5) = a_j - \zeta_j^{-1} \log |r^{\perp}_{i,d}|, \tag{34}$$

where $\text{Rank}_{i,d}$ is the rank of $|r^{\perp}_{i,d}|$ in descending order. A larger ζ_j corresponds to a fatter tail. Unlike Kwon and Tang (2025), we use characteristic- and industry-adjusted residual returns rather than raw returns. The median topic has about 23,000 such firm-days, so every estimate rests on several hundred tail observations or more.

We estimate (16) by OLS with equal weight per topic and report heteroskedasticity-robust (HC1) t-statistics. These treat the topic slopes as independent, although topics share firm-days, so their first-stage estimation errors are correlated. As a robustness check, we use the covariance $(X'X)^{-1}X'\big(\widehat{\Omega}+\widehat{V}^{\text{off}}\big)X(X'X)^{-1}$, where $\widehat{\Omega}$ is diagonal with elements $\hat{u}_j^2/(1-h_{jj})^2$ (HC3) and $\widehat{V}^{\text{off}}$ contains the off-diagonal elements of the joint covariance of the first-stage slopes, which combines the topics' influence scores within firms and calendar months as for the themes above. This lowers the t-statistics by about a quarter, to -5.62 for extremeness and -2.17 for attention in the joint specification, and the quantitative share is then insignificant ($t = 1.55$). Every other coefficient keeps its significance at the 5% level.

Theme dynamics. For Figure 15, we replace $R^{+60}_{i,d}$ in equation (33) with $R^{+h}_{i,d}$, defined before winsorization as $\sum_{\tau=1}^{h} r^{\perp}_{i,d+\tau}$, and estimate a separate slope $\beta_g(h)$ at each horizon $h \in \{1, 2, 5, 10, 20, 40, 60, 90, 120\}$. The sample contains the eligible firm-days with a complete 120-day trading-calendar window and is held fixed across horizons. It spans news days from January 1996 to July 2025. Initial-return cutoffs remain unchanged. For each horizon, we compute cumulative-return cutoffs across unique eligible firm-days of the 60-day sample that

have a complete calendar window at that horizon, before restricting to the common 120-day sample. The 20-day and 60-day cutoffs therefore equal those of the topic regressions. Missing subsequent firm returns continue to enter as zero.

The plotted bands are $\widehat{\beta}_g(h) \pm c_g \widehat{s}_{g,h}$, where $\widehat{s}_{g,h}$ is the firm- and month-clustered standard error and c_g is a theme-specific critical value. Following Montiel Olea and Plagborg-Møller (2019), we simulate one million Gaussian vectors using the estimated cross-horizon correlation of the slopes and set c_g to the 95th percentile of their maximum absolute value across the nine horizons. These bands provide asymptotic 95% simultaneous coverage over the nine evaluated horizons within each theme. They do not adjust across themes or cover unobserved intermediate horizons.

N Negative Tone at Daily Frequency

Table 8 repeats the cross-topic regression (16) of Table 3 with one-day topic slopes $\hat{\beta}_j(1)$ in place of the 60-day slopes. Each topic slope is estimated as in (15), with the residual return of the trading day after the news as the dependent variable. The firm-days, weights, characteristics and inference are unchanged.

At the one-day horizon, topics with more negative tone show more continuation. A one-standard-deviation higher negative tone raises the slope by 0.14 basis points per percentage point of initial return ($t = 1.69$), significant at the 10% level. The estimate is unchanged when the other characteristics are included (0.14, $t = 1.62$). The relationship is short-lived. With two-day slopes the coefficient falls to 0.08 ($t = 0.62$), with five-day slopes to 0.04 ($t = 0.22$), and it is insignificant at every horizon up to 60 days. This pattern is consistent with Tetlock et al. (2008), who find that prices underreact to negative words only briefly. Negative tone therefore does not explain the longer-horizon differences in continuation in Table 3. Extremeness and the fog index keep the signs they have at 60 days in the joint regression.

Table 8: News Characteristics and One-Day Return Continuation across Topics

This table repeats Table 3 with the full-sample one-day topic slopes $\hat{\beta}_j(1)$ as the dependent variable: the slope of the residual return on the trading day after the news on the news-day residual return, estimated with calendar-month intercepts and the same firm-days and weights as the 60-day slopes. The characteristics are those of Table 3, standardized across topics, and coefficients are multiplied by 100, so they give the change in basis points of next-day residual return per percentage point of initial return for a one-standard-deviation higher characteristic. Heteroskedasticity-robust (HC1) t-statistics are in parentheses; *, ** and *** denote significance at the 10%, 5% and 1% levels. Four topics without an activating full article are omitted. The sample covers news days from January 1996 to October 2025.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Negative tone	$+0.14^{*}$						$+0.14$
	$(+1.69)$						$(+1.62)$
Uncertainty		$+0.08$					$+0.07$
		$(+1.07)$					$(+0.89)$
Quantitative share			$+0.18^{**}$				$+0.05$
			$(+2.57)$				$(+0.69)$
Fog index				$+0.02$			$+0.26^{***}$
				$(+0.39)$			$(+4.18)$
Attention					-0.28^{***}		-0.16
					(-3.09)		(-1.51)
Extremeness						-0.42^{***}	-0.40^{***}
						(-5.82)	(-4.10)
R^2	0.02	0.00	0.03	0.00	0.06	0.14	0.20
Topics	224	224	224	224	224	224	224

O Additional News Themes Over Time

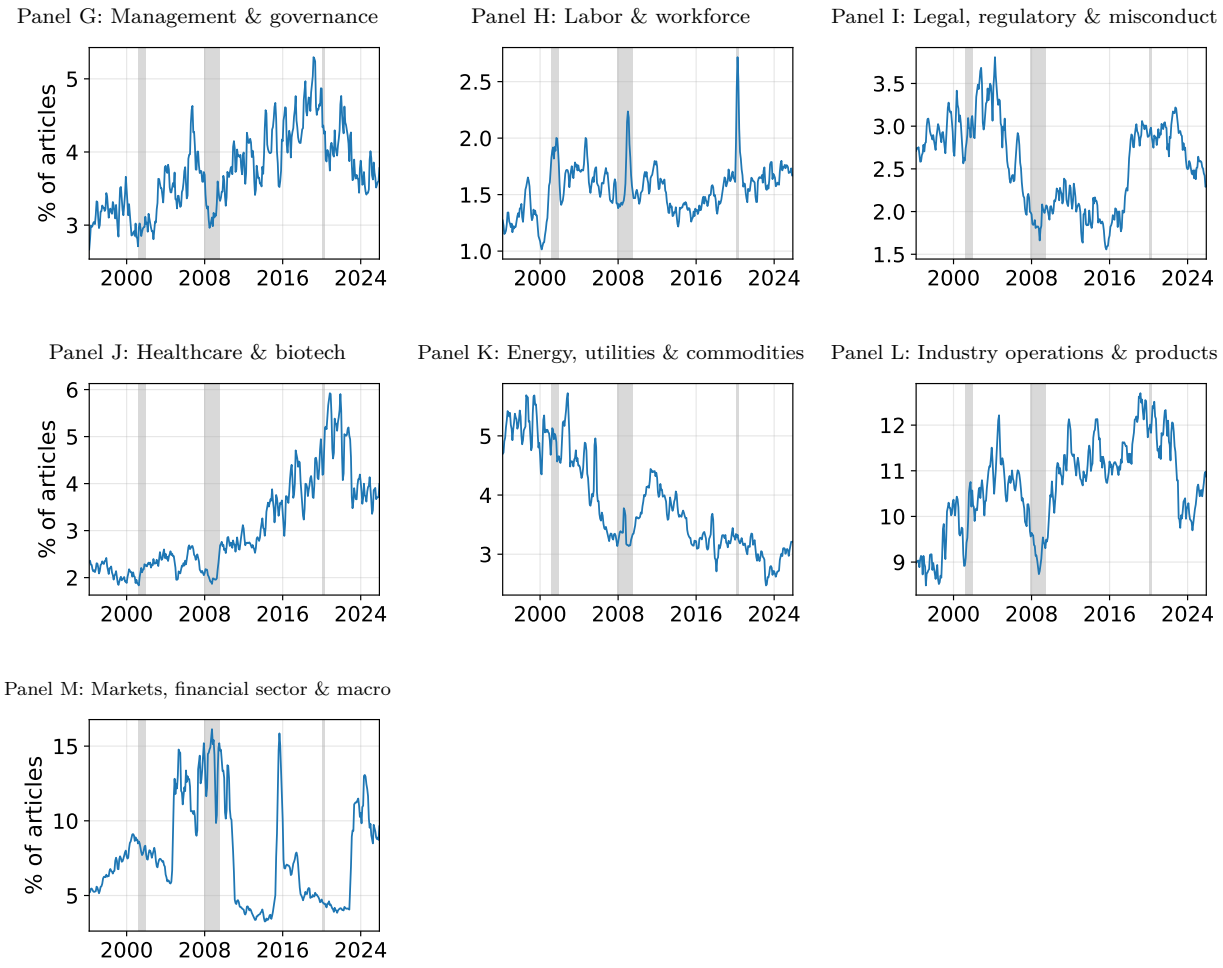


Figure 26: Additional News Themes Over Time

Note. The remaining seven themes complement Figure 7, using the same construction, sample, and smoothing. Each panel has its own scale. Grey bands indicate NBER recessions.

The pandemic provides a clear example of how news coverage shifts. The share of Reuters coverage devoted to healthcare and biotech rises from 4.3% in 2019 to 5.5% over March to December 2020, peaking at 5.9% in November. Over the same periods, the share devoted to guidance and outlook falls from 12.4% to 10.1%, consistent with firms withdrawing guidance. Recessions also leave a recognizable imprint on coverage: labor and workforce news reaches 2.0% to 2.7% of articles around the downturns of 2001, 2008–2009, and 2020, compared with 1.5% in 2010–2019. Coverage of credit ratings and debt financing peaks at 11.2% in

May 2020 amid downgrades and emergency debt issuance. These shifts in coverage reflect the economic developments firms face: the pandemic draws attention to healthcare, while downturns bring layoffs and financing pressures to the foreground.

Longer-run movements reflect changes in the composition of coverage as well. Earnings results falls from 25.0% of articles in 1996–1999 to 17.6% from 2002 onward, while analyst actions and estimates rises from 10.4% in 1996–1999 to 18.2% in 2021–2025. Some shifts also reflect changes in Reuters' reporting practices: increases in the markets, financial sector and macro theme partly capture more frequent automated alerts about stock-price movements and order imbalances, rather than greater coverage of macroeconomic developments.

Changes in which firms Reuters covers matter too. Healthcare's share of articles rises from 2.7% in 2012 to 4.3% in 2019, but holding healthcare firms' coverage share at its 2010–2012 level removes most (about 82%) of that increase. The pandemic increase survives this adjustment. Moreover, among articles about firms outside the healthcare sector, the share devoted to healthcare rises by almost 40% in 2020. This supports interpreting the pandemic shift as a change in the subjects covered, beyond a change in the firms receiving coverage.

P News Themes: Mean and Volatility Decomposition

We decompose theme contributions to the mean and volatility by keeping the MSRR weights fitted on all 228 news factors and dropping one theme at a time, without re-optimizing. With $\hat{b}_{t,j}$ the weight of news factor j, the news portfolio is $F^{\star}_{t+1} = \sum_j \hat{b}_{t,j} F_{j,t+1}$, and without theme g it is

$$F^{\star,-g}_{t+1} = F^{\star}_{t+1} - \sum_{j \in g} \hat{b}_{t,j} F_{j,t+1}. \tag{35}$$

For the out-of-sample mean and volatility, we report the contribution of theme g as the share of the full news portfolio's value lost when the theme is left out, $100 \times \Big(\theta(F^{\star}) - \theta(F^{\star,-g})\Big)/\theta(F^{\star})$; a positive contribution means the theme raises the statistic. Mean contributions add up to 100% across themes, while volatility contributions need not.

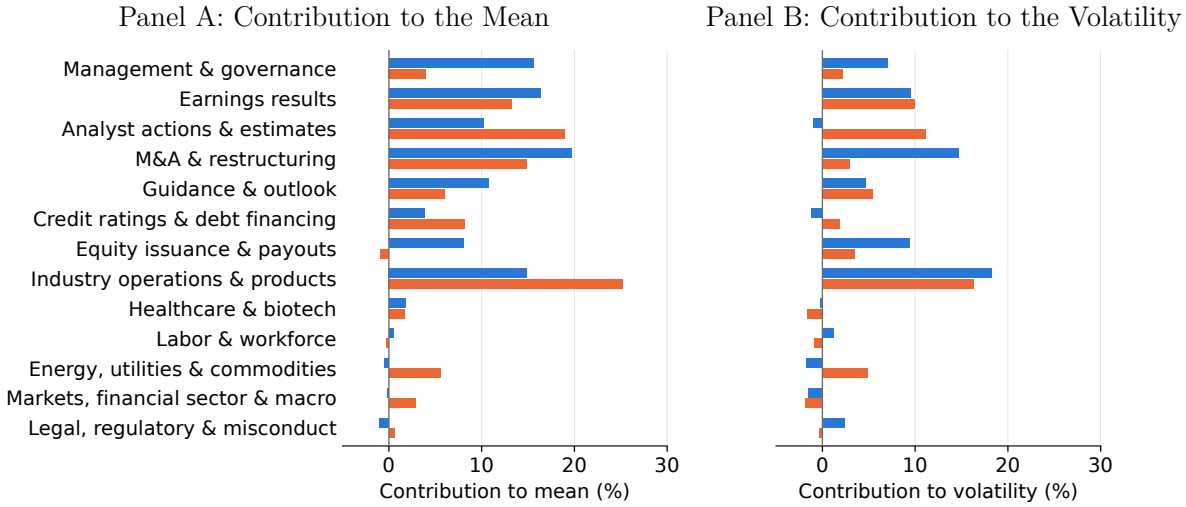


Figure 27: Mean and Volatility Contributions of News Themes

Note. Panels A and B show the percentage reduction in the mean and volatility when each theme is removed, holding the full-portfolio MSRR weights fixed. Blue bars denote equal weighting and orange bars value weighting. The theme order matches Figure 8. The sample is January 2016 to November 2025.

Figure 27 helps distinguish a theme's contribution to mean returns from its contribution to portfolio volatility. For example, *industry operations and products* contributes positively to the mean under both weighting schemes, but adds substantially more volatility under equal weighting. *M&A and restructuring* likewise contributes to the mean under both schemes while adding less volatility under value weighting.

Q News Attention and Macroeconomic Conditions

We measure the attention that news factor j receives in month t by its activation-weighted share of that month's articles, $a_{j,t} = |\mathcal{A}_t|^{-1} \sum_{a \in \mathcal{A}_t} \tilde{e}_{a,j} / \sum_k \tilde{e}_{a,k}$, where $\mathcal{A}_t$ is the set of articles published in month t with at least one active news factor and k runs over the news factors of the library. Each article is split across the news factors in proportion to its activations, as for the theme shares in Figure 7. This measure involves neither orthogonalization nor returns. Because the composition of the news corpus drifts over three decades, we work with *abnormal* attention, the log share relative to its own average over the preceding 24 months,

$$\tilde{a}_{j,t} = \log a_{j,t} - \frac{1}{24} \sum_{s=1}^{24} \log a_{j,t-s}, \tag{36}$$

which uses no future information. We correlate abnormal attention with macroeconomic series from the FRED-MD database, each in its standard stationary transformation, in the same calendar month.

Figure 28 shows six news factors whose attention co-moves with macroeconomic series. Attention to corporate layoffs moves with initial unemployment claims (correlation of 0.26), and at least as strongly with the change in unemployment, payroll growth, the VIX and industrial production growth. It behaves as a broad downturn indicator, peaking in 2001, early 2009 and April 2020. Attention to analyst price-target cuts falls with industrial production growth (-0.23) and capacity utilization (-0.27), and attention to analyst price-target raises rises with the stock market (0.34). These links are contemporaneous: the correlations peak in the same month and are near zero a few months away. Attention to credit-rating downgrades moves with the VIX (0.33), and attention to financial distress and to corporate credit events with the BAA–federal funds spread (0.24 and 0.38). The VIX and the spread enter FRED-MD in levels, so part of this co-movement is shared low-frequency variation, and attention to financial distress tends to lead the spread (its correlation rises to 0.40 six months ahead, as in 2008).

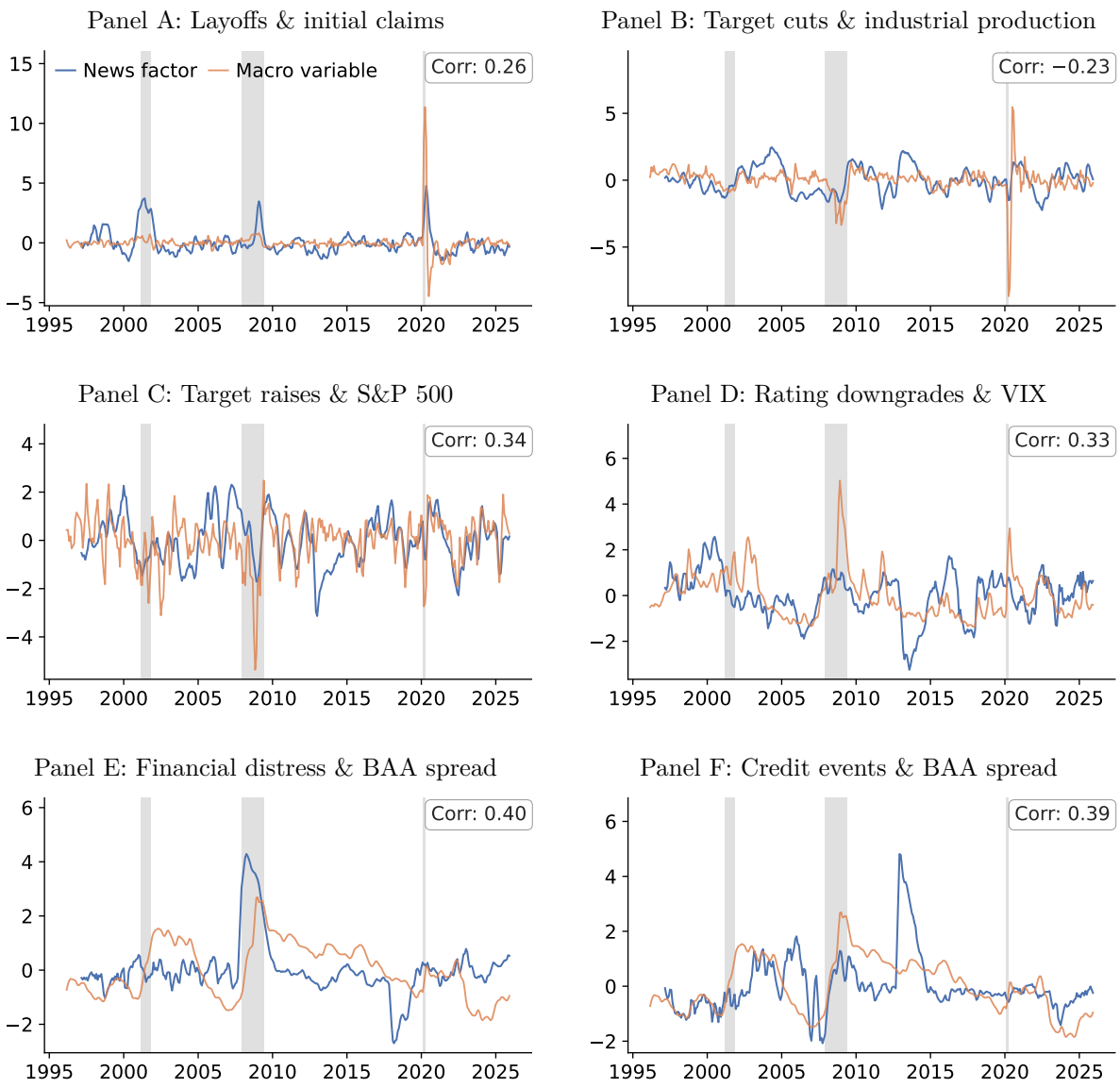


Figure 28: Attention to News Factors and Macroeconomic Series

Note. Each panel plots the abnormal attention (36) of one news factor (blue), named by its in-sample label, against a FRED-MD series (orange) in its FRED-MD transformation: log changes for initial claims, industrial production and the S&P 500, and levels for the VIX and the BAA–federal funds spread. The attention to price-target downgrades, which falls when industrial production grows, is shown with its sign reversed; the reported correlation is that of the unreversed series. Lines are three-month averages, standardized over the sample. The box in each panel reports the largest correlation between the monthly series within six months of lead or lag. Grey bands indicate NBER recessions. The sample is January 1997 to December 2025.