Narrative Asset Pricing:

Interpretable Systematic Risk Factors from News Text*

Leland Bybee
Yale University

Bryan Kelly
Yale University, AQR Capital Management, and NBER

Yinan Su
Johns Hopkins University

April 27, 2022

Abstract

We seek fundamental risks from news text. Conceptually, news is closely related to the idea of systematic risk, in particular the "state variables" in the ICAPM. News captures investors' concerns about future investment opportunities, and hence drives the current pricing kernel. This paper introduces a method to extract a parsimonious set of risk factors and eventually a univariate pricing kernel from news text. The state variables are reduced and selected from the variations in attention allocated to different news narratives. As a result, the risk factors attain clear text-based interpretability as well as top-of-the-line asset pricing performance. The empirical method integrates topic modeling (LDA), latent factor analysis (IPCA), and variable selection (group lasso).

JEL Classification: C38, C52, G11, G12

Keywords: news, narratives, textual analysis, cross section of returns, ICAPM, factor model, IPCA, variable selection

*Kelly (corresponding author): bryan.kelly@yale.edu; Yale School of Management, 165 Whitney Ave., New Haven, CT 06511; (p) 203-432-2221; (f) 203-436-9604. Bybee: leland.bybee@yale.edu; Yale School of Management, 165 Whitney Ave., New Haven, CT 06511. Su: ys@jhu.edu; Carey Business School, 100 International Drive, Baltimore, MD 21202.

We are grateful to the seminar participants at Yale SOM and JHU Carey for their helpful comments.

AQR Capital Management is a global investment management firm, which 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.

1 Introduction

A central premise of asset pricing is that differences in expected returns stem from differences in risk exposures—but what are the fundamental risks that investors care about in pricing risky assets? Merton's (1973) Intertemporal CAPM provides theoretical guidance: "State variables" that entail changes in the "future investment opportunity set" determine agents' current consumption. Hence, such state variables represent the fundamental risks, and an asset's covariances with these risks shall dictate the asset's risk premium.

However, the identity of the state variables has remained largely conceptual as the empirical pursuit of interpretable fundamental risk factors has had limited success. Some attempts follow the ICAPM closely and propose various *macroeconomic indicators* (or the unexpected components extracted from these indicators) as proxies for the state variables.[1] Whereas *firm characteristics-sorted portfolios*, which are detached from interpretable fundamentals, often achieve better pricing performance when it comes to explaining the risk premiums of "anomaly" portfolios (which are themselves also based on characteristic sortings).[2] There is not yet a systematic risk model with both top-of-the-line pricing performances and theory-based interpretability.

We bridge this gap between theory and empirics using *news text*, which is indeed close to the original spirit of the ICAPM. Reviewing the theory, the state variables should be *unexpected new information* that is just acquired by investors and foretells changes in future investment opportunities. In fact, the unexpected components of macroeconomic indicators, when used as state variables' proxies, are often labeled as "news" metaphorically.[3] We take this theoretical guidance literally and build fundamental risks from the archive of the *Wall Street Journal*. The information contained in economic and business news should be predictive and pertinent to pricing agents' concerns. These properties are conceptually guaranteed by media competition and the induced editorial incentives. There should be no other publicly available signals that predict more timely or accurately than news, barring frictions in news production. Moreover, news should cater to market participants' key risk concerns. Events like recessions and pandemics that are closely related to the market participants' pricing kernel should make up the majority of news reporting by the nature of business journalism.

[1] For example, Chen et al. (1986), Cochrane (1996), Bali and Engle (2010), and Rossi and Timmermann (2015) use macroeconomic indicators such as industrial production, investment, and inflation to proxy for the state variables.

[2] Examples include Fama and French (1996), Fama and French (2016), Hou et al. (2015).

[3] For example, Vassalou (2003) is titled "*News Related to Future GDP Growth as a Risk Factor in Equity Returns.*"

Applying news text to a classical asset pricing model provides unparalleled benefits in terms of understanding the nature of fundamental risks. Even if one builds ICAPM state variables from macroeconomic indicators, the interpretation of risks does not go beyond some transformation of the indicators themselves. Whereas the *narratives* contained in the news are much more diverse and concrete, while at the same time directly readable. Daily news content covers a wide range of economic issues, while the editorial process can locate the most pertinent narratives behind the risks at each point in time. With news narratives, we attempt to bring the central but abstract theoretical concept of systematic risk closer to what actually happens in the real world and how the market perceives it.

Although news has appealing conceptual advantages, the textual form of data comes with practical hurdles for empirical analysis. The crux of the problem is distilling a parsimonious set of risk factors (and eventually a univariate pricing kernel) from the vast amount of textual data. The problem is tackled in stages of dimension reduction. First, we extract topical narratives in the *Wall Street Journal*, and construct daily innovations in news attention allocated to each narrative.[4] As advocated by Shiller (2017), textual analysis is a powerful tool to quantify fluctuations in popular narratives and understand their impact on the economy. We use Latent Dirichlet Allocation (LDA) to automatically group words and phrases into interpretable topical themes based on their co-occurrences in news articles.[5] This unsupervised analysis yields two objects: the word composition of each topic, which gives concrete and interpretable meanings to the news narratives; and each article's attention allocation to the narratives. The article-level narrative attention is aggregated at the daily frequency, from which we calculate the daily innovations in narrative attention by subtracting the trailing averages. Thus, we transform the textual news data into daily time series of *innovations in narrative attention*, which are the quantitative building blocks supplied to the subsequent econometric analysis of the systematic risk model. (There are 180 series since LDA identifies 180 narratives.)

We posit a model in which the observed narrative innovations are related to the ICAPM state variables that price the cross section of risky assets. The estimation of the *narrative-to-state variable relationships*, as well as the narrative-based risk factors, contains the bulk of our methodological con-

[4]This initial step of constructing narrative innovations largely follows Bybee et al. (2021, henceforth BKMX). The procedure is summarized in Subsection 4.1.1.

[5]For example, terms like "economic downturn," "steep decline," "hardest hit," "steep drop," etc. shows up together in a narrative. Following BKMX, the label of a narrative is manually assigned by summarizing the common theme displayed in the automatically grouped topic words. This example is labeled as "Recession."

tribution. Let us first illustrate the economic mechanism with one narrative as an example. Upon reading an increase in the "Recession" narrative, investors shall expect worse future investment opportunities, perhaps as a result of worsening macro conditions or increased uncertainty. According to the ICAPM, in a state with deteriorated "future investment opportunities," the concurrent consumption is proactively adjusted downward, which accordingly increases the marginal utility and the stochastic discount factor.[6] As a result, a risky asset that (ex ante) more positively covaries with the "Recession" narrative should earn a lower risk premium for its benefit in hedging consumption risk (all else equal and vice versa).

The 180 narratives cover a wide-range of themes. Besides just "Recession," we want to identify all the aspects of investor concerns and account for how much each matters to fundamental risks. Existing statistical studies of stock returns suggest the dimensionality of the systematic risk space is in the single-digit range, far smaller than 180. Moreover, some narratives are likely completely irrelevant to investors' risk concerns, such as "Arts" and "Humor/language." We formalize these ideas with a model in which the 180 narrative innovation series are noisy observations that load on a small number of underlying state variables. It requires a dimension reduction to recover the state variables from the narrative innovations. Importantly, we induce *sparsity* in the narrative-to-state variable relationships—some narratives have zero loadings on state variables. As a result, our estimator selects the relevant narratives and filters out the irrelevant to gain a concrete understanding of the risks' composition.

The estimation method utilizes the fact that the state variables' tradable mimicking portfolios should be systematic risk factors that "fit" realized individual stock returns well. The estimation can be understood as an upgrade to the two-step regression of Fama and MacBeth (1973), which, in its original form, estimates the mimicking portfolios of *observed* state variables. An upgrade is necessary because the state variables are no longer directly observed, but need to be *reduced* and *selected* from the 180 narratives. The first step is mostly unchanged: for each stock, calculate daily time-series covariances with narrative innovations in trailing windows. These realized *narrative covariances* shall provide valuable conditioning information about the stock's exposure to the risk factors (a.k.a. loading, β), since the factors are related to the narratives according to the model. The second step

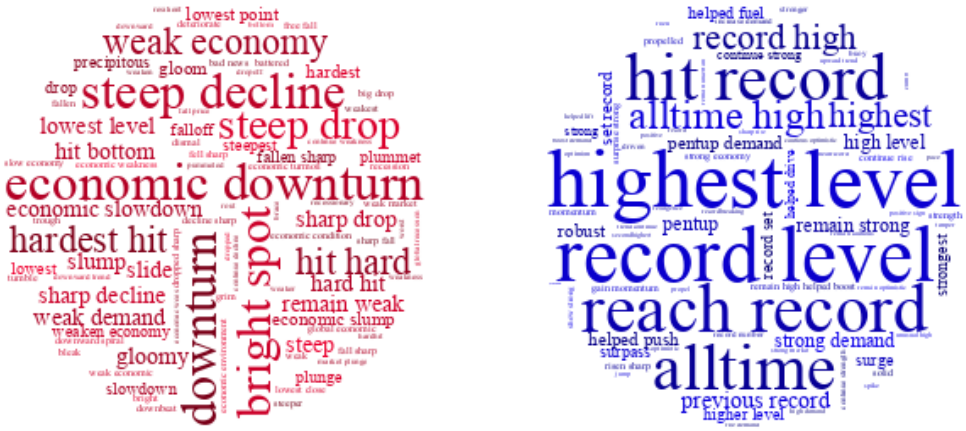
[6] In the more modern variants, intermediary capital instead of consumption is adjusted (He and Krishnamurthy, 2013; He et al., 2017). Either form of the pricing kernel shall respond to news about future concerns. Our framework view the two channels as compatible.

is changed from period-by-period cross-sectional regressions to a panel-wise Instrumented Principal Components Analysis (IPCA, Kelly et al., 2017). It uses narrative covariances as stock-month level instruments that map into each stock's β (via a dimension reduction subject to estimation). The narrative covariances become a whole new set of firm characteristics in parallel with the traditional ones such as size, book-to-market, and so on (Kelly et al., 2019). IPCA estimates both the instrumental mapping (which gives the narrative-to-state variable relationships) and the latent systematic risk factors.

To add selection, we introduce a new extension to the IPCA framework called *Sparse IPCA*. Sparse IPCA adds a regularization term to IPCA's target function that (additive-separately) penalizes each instrument's effects on factor loadings. The group lasso penalty zeroes out an individual instrument's effect. (Each instrument's effect on all factor loadings needs to be "grouped" together since the factors are not individually identified.) The strength of the penalty is controlled by a regularization constant. Starting from the unregularized IPCA benchmark, where all instruments are included, increasing regularization filters out the less relevant narratives incrementally. Meanwhile, systematic factor space estimation significantly improves, as measured by the Sharpe ratio of the factors' mean-variance efficient (MVE) combination not just in, but also *out of sample*. As regularization strength further increases, the number of selected narratives continue dropping to single digits, while the MVE's Sharpe ratio reverses back down as the selected model becomes too sparse. We pick the peak of the Sharpe ratio path as the tuned in-sample estimation, which is subsequently brought to later periods for out-of-sample construction. The tuning criteria is based on the economic condition that the true systematic factor space should span the global MVE portfolio.

The estimation result provides a literal *interpretation of the fundamental risks* with news narratives, which are comprised of human-readable words and phrases with clear economic contents. The benchmark estimation with three factors selects around a dozen narratives out of the 180 candidates. We first focus on interpreting the MVE combination of the state variables x^{MVE} by organizing the narratives according to their orientations with respect to the MVE direction. Although the factors are statistically equivalent after rotations, this direction is particularly important. Theoretically, x^{MVE} is the univariate pricing kernel, which is the single determinant of risk premiums (Hansen and Richard, 1987, the SDF is a linear transformation of x^{MVE} with a negative coefficient). In our model, x^{MVE} is a linear combination of the selected narrative innovation series.

Figure 1: MVE State Variable (x^{MVE}) Interpretation at the Word Level



Negative Positive

Note: Each words' impact on the MVE ($I_{w\to\text{MVE}}$). A word's font size corresponds to the absolute magnitude of its impact. The construction method is detailed in Subsection 5.1.

We find the most prominent narrative with a negative impact on x^{MVE} is "Recession", whereas "Record high" and "Optimism" are the leading positive components. Upon closer inspection, the articles that fall under these narratives are typically organized around a piece of quantitative information (see example articles in Table 1). The underlying quantitative events cover a wide range of issues, from consumer sentiment surveys to home-building statistics. These news events all have significant impact on investors' concerns about the future investment outlook at the time of reporting. Unlike traditional macroeconomic statistics, had one attempted to capture these with quantitative datasets, for one it would be hard to gather and aggregate datasets that cover such extensive issues. Moreover, even the same numeric reading of such specific issues, such as automobile registration or Wall Street bonuses, can have different risk implications in the ever-changing economic environment. Our method essentially relies on the news production process to locate the most "newsworthy" quantitative events at the time. It utilizes the editor's language choice, to extract the risk contents from the heterogeneous quantitative events. We can "zoom-in" the interpretation results from the level of narratives to the level of words by tracing the composition of each narrative. Figure 1 visualizes the words with negative and positive impacts on x^{MVE}, respectively, producing our most granular display of the pricing kernel.

Besides the polar narratives, we find "Trading activity," "Bear/bull market," and "International

exchanges" are also important sources of systematic risk. However, their contributions to the state variables are largely *orthogonal* to x^{MVE}. They reflect a sense of market activeness in general, and do not have a clear positive or negative connotation. As a result, covariances with these narratives, though important for explaining realized returns, have little impact on the stocks' expected returns.

Besides x^{MVE}, the same framework also supports interpreting the market portfolio (as well as other pre-specified factors) by projecting it onto the narrative-based systematic factors space. In an event retrieval exercise, we trace the specific news articles that are associated with realized market return spikes. To do so, we calculate each article's model implied impact to the market portfolio based on its textual content. Then, on the same day of the realized market return spike, we find the articles that most align with the spike. The retrieved articles display the concrete events behind realized market variations over three decades, which range from worries about interest rates and government debts to uncertainties about a double-dip recession and the European crisis's ripple effects. The method could be potentially deployed as a live translation from textual news to quantitative risk factors. In summary, examining the narrative-to-state variable relationships yields insightful new observations about the nature of fundamental risks.

Evaluating from the quantitative perspective, the narrative asset pricing model excels in tests of its theoretical properties. First, we evaluate the cross-sectional pricing properties of the narrative systematic risk factors. To avoid the usual critique of in-sample overfitting, we *only* evaluate the out-of-sample (OOS) estimates. To do this, notice the narrative factors (and their MVE combination) can be seen as dynamically formed portfolios that treat each stock's narrative covariances as trading signals. We guarantee the portfolio weights are ex ante available with parameters estimated in prior training samples. The MVE attains an annualized Sharpe ratio well above 1, which is significantly higher than characteristics-sorted portfolios, such as the Fama-French/Carhart six factors, and their (OOS) MVE combinations. Moreover, the narrative factors price the cross section of a comprehensive list of anomaly portfolios with comparable (and sometimes smaller) pricing errors compared to the characteristics-sorted factors (as measured by $|\alpha|$, GRS test, etc.). The result is remarkable in the sense that it only relies on the new set of conditioning information of narrative covariances, without any firm fundamental information. Yet the pricing capability applies to the anomaly portfolios formed on firm characteristics, which the literature has been iterating on for decades.

Second, we evaluate the forecasting properties of the narrative-based state variables implied by the

ICAPM. The estimated state variables are linear combinations of the selected narrative innovations. We show that the state variables, in particular the univariate x^{MVE}, predict future market return, consumption growth, and a list of other macroeconomic indicators. Moreover, the signs of the predictive relationships are consistent with the ICAPM—the state variable that earns a positive risk premium is indeed "good news" in the sense that it predicts positive changes future investment opportunities (and negatively predicts changes in counter-cyclical indicators like credit spread and unemployment). In summary, the quantitative tests indicate the identified narrative asset pricing model is consistent with the classical asset pricing theory.

Literature: Using news text for asset pricing is a new and promising area of research (see the surveys of Loughran and Mcdonald, 2016; Gentzkow et al., 2019). It provides a unique way to observe central theoretical concepts like public information and investor attention, which otherwise could only be inferred indirectly. Engle et al. (2020) and Liu and Matthies (2021) are similar to this paper in terms of extracting risk factors from the fluctuations in news text. Engle et al. (2020) provide a dynamic trading strategy to hedge climate change risk, which is measured by tracking the fluctuations in climate change news attention. They focus on a particular narrative, which corresponds to a particular priced risk, whereas we select from all the automatically generated narratives in order to map out the whole systematic risk space. Liu and Matthies (2021) construct a risk index by tracking the frequencies of a set of pre-specified words about economic growth in news text.[7] They show the index predicts long-run trends in consumption growth, and therefore constitutes the pricing kernel. Their motivation with the long-run risks model is similar to, and to some extent compatible with, the ICAPM framework. Our method looks for a broader set of "future investment opportunities" rather than targeting just climate change or long-run consumption growth.

More generally, the literature exploring asset pricing applications of news text has proven fruitful in recent years. Some of these, such as Baker et al. (2016) rely on manual labeling to identify relevant text aspects—in their case manually labeling the types of news associated with large market returns to better understand the drivers of these large movements. Alternatively, a growing body of studies employ machine learning methods to process the underlying text. Ke et al. (2020) and Manela and Moreira (2017) develop textual machine learning methods to predict individual stock returns, and aggregate volatility (VIX), respectively. Jeon et al. (2021) attribute news as sources of jumps in

[7] Liu and Matthies (2021) pick "Consumption", "GDP", "Gross Domestic Product", "GNP", "Gross National Product", and "Pig Iron" as a set of key words in their baseline construction.

stock returns. In contrast, this paper has a closer connection with classical multi-factor asset pricing models. We approach risk premiums from narrative-based risk exposures, rather than directly using news to predict returns, which embeds an assumption of informational inefficiency.

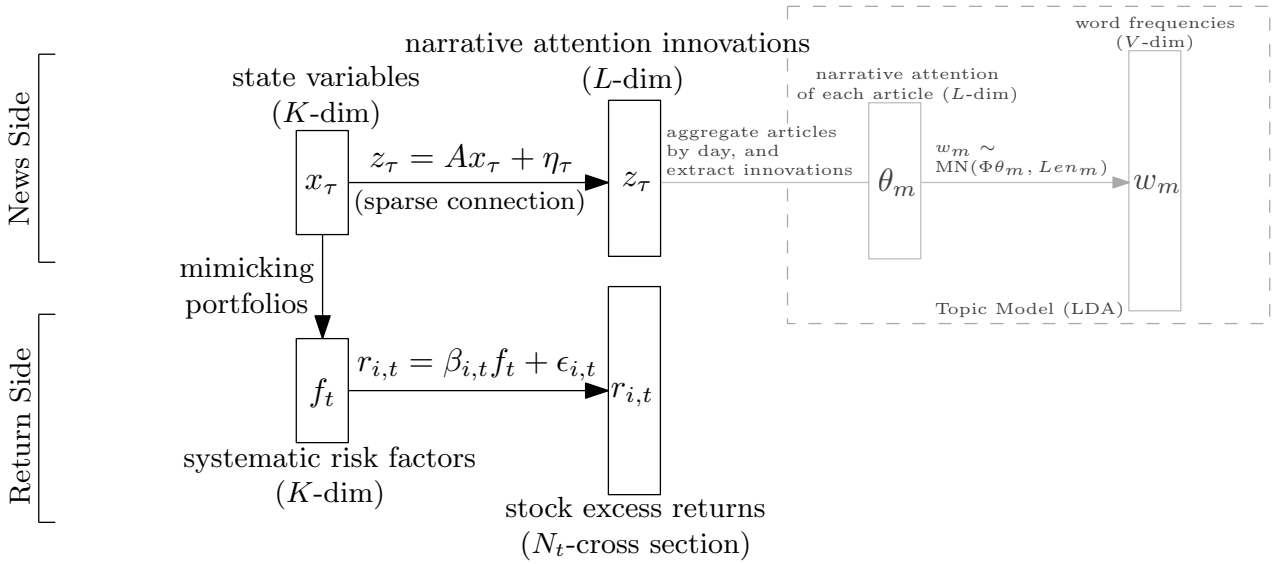
In terms of statistical methods, Sparse IPCA is a generalization of the selection and shrinkage functions of (group) lasso (Tibshirani, 1996; Yuan and Lin, 2006) from regression to latent factor analysis. Sparse IPCA is similar to Sparse Principal Components Analysis (SPCA) of Zou et al. (2012), which imposes lasso-type regularization on factor loadings. Pelger and Xiong (2020) imposes hard-thresholding on factor loadings. They also emphasize the improved interpretability from sparse estimates. The difference is that Sparse IPCA selects instruments according to their effects on factor loadings, in contrast to the two above that select factor loadings themselves. As an extension to Fama-MacBeth with regularization, our work is also related to Bryzgalova (2015).

The rest of the paper is organized as the following. Section 2 builds the data generating process and Section 3 introduces the estimation method. Section 4 reports the data and estimation results. Section 5 reports the textual interpretation of the fundamental risks. In terms of quantitative evaluations, Section 6 reports the asset pricing performances, and Section 7 reports the forecasting properties of the narrative-based state variables. Section 8 concludes.

2 Model

Figure 2 summarizes the data generating process for innovations in narrative attention (z), stock excess returns (r), as well as the word frequencies of individual news articles (w). The figure has three parts: The main novelty is connecting state variables to news narratives (x to z). The empirical goal is to estimate this relationship in order to understand the fundamental risks from the perspective of news narratives. The return generating process (f to r) is the canonical latent factor model with time-varying factor loadings. We estimate the narrative-based f for factor pricing tests. The text generating process (θ to w) follows that of LDA, which is taken as a stand-alone data pre-processing step to prepare narrative innovations (z).

Figure 2: Data Generating Process Illustration



Note: τ indexes days, t for months, and m for articles. L is the number of narratives, V is the size of the vocabulary.

2.1 The News Side of the Model

Let x_t ($K \times 1$ vector) be the ICAPM state variables. It contains both the growth of the wealth portfolio, which is the single state variable of the CAPM, and the additional state variables with intertemporal predictability that the ICAPM adds to the CAPM.[8] We do not distinguish the wealth portfolio from the rest of the state variables. Both are pooled in x_t and identified with news narratives in a unified fashion. In theory, all entries in x_t together constitute the fundamental risks of asset pricing, in the sense that the covariances with x_t explain the cross section of expected returns: $\mathbb{E}_{t-1} r_{i,t} = \gamma \mathbb{C}\text{ov}_{t-1}(x_t, r_{i,t})$ for any excess return $r_{i,t}$.[9] The unified estimation based on narratives also circumvents Roll's critique on proxying wealth by the market portfolio. For example, the variation in the human-capital component of wealth might be reflected in news narratives, but is not revealed by the market portfolio. (The narrative-based pricing kernel indeed forecasts changes in the labor market conditions, see Section 7.)

Let z_τ be the innovation of each narrative's attention on day τ arranged in an $L \times 1$ vector, L being the number of narratives. We assume z_τ is related to the K ICAPM state variables x_τ via an

[8] Due to the rotational unidentification problem, there is not a distinct entry in x that corresponds to the wealth portfolio. It would require a linear combination of x to recover the wealth portfolio, which is not our estimation interest.

[9] $1 \times K$ vector γ represents the risk prices of the state variables. After we defined x's mimicking portfolios f in 2.2, it is easy to see that $\gamma := \mu_f^\top \Sigma_{ff}^{-1}$.

$L \times K$ matrix A:

$$z_\tau = Ax_\tau + \eta_\tau, \tag{1}$$

where η_τ represents the part of narrative innovations that is irrelevant to the state variables and the asset market in general. (Throughout the paper, τ indexes days and t indexes months. Since z and x are innovations, they can be accumulated at both daily and monthly frequencies, written as x_τ and x_t, respectively.[10] The same subscripting convention goes for returns below.)

This equation is the key specification that relates news narratives to asset pricing state variables. One interpretation is that the pricing agents respond to news by proactively adjusting consumption (or other forms of the pricing kernel) concurrently. More realistically, investors directly respond to some other more timely (intraday) public signals, which might be conveyed in any format (word of mouth, social media, quantitative reports, live feeds, and so on). The newspaper shall quickly (by the next morning) report the signals, given their importance to risks, and thereby record them in the textual format, allowing econometricians to harvest for asset pricing studies.[11]

The number of narratives is an order of magnitude more than the commonly perceived dimensionality of the underlying risks ($L \approx 180 > K \approx 3 \sim 6$).[12] This means many narratives are connected to similar risks, to different degrees and potentially with opposite signs. More importantly, we posit some other narratives are *entirely irrelevant* to the asset return dynamics, such that the corresponding rows of matrix A are all zeros. The estimation will support such *row-wise sparsity* in A with a selection method. Estimating A eventually tells us which narratives matter for which risks and by how much, thereby providing a textual interpretation of fundamental risks (detailed in Section 5). The selected topics are the narratives relevant for risk, while the remaining ones contain news unrelated with investment concerns.

Narrative attention innovations (z_τ) are constructed from LDA topic analysis, which is taken as a standalone procedure to build numeric time series from raw textual data. For interpretation

[10] To be exact, the discrete time ICAPM is a linear approximation of the original continuous time version. If the instantaneous state variable is X_s, the discrete time accumulated state variables are: $x_\tau := \int_\tau^{\tau+1\text{day}} \mathrm{d}X_s$, and $x_t := \int_t^{t+1\text{month}} \mathrm{d}X_s$.

[11] To account for the asynchrony, we match the news printed in the morning with the returns realized on the previous day in calculating their covariances. See 4.1.1 and footnote 28.

[12] It is argued the dimensionality of the factor space is parsimonious with firm characteristics, e.g. Kelly et al. (2017), Kozak et al. (2020). Our empirical results suggest that dimensionality of the factor space is still low with the narrative-based conditioning information (see Section 6).

purposes, LDA also gives the word composition of narratives, providing a way to display the risks at the granular level of words and phrases. In a nutshell, LDA's textual data generating model says each topic is a (weighted) set of words and phrases that concentrate on a common theme ($\phi_l \in \Delta^V$, a V-dimensional simplex; V: size of the vocabulary).[13] The words in each article are generated as a mixture of these topics,

$$w_m \sim \text{Multinomial}(\Phi\theta_m, Len_m), \tag{2}$$

where the mixing weights $\theta_m \in \Delta^L$ are article m's attention allocated to each topic. (w_m: observed word frequencies of article m, $V \times 1$ vector; $\Phi := [\phi_1, ..., \phi_L]$, $V \times L$ matrix; Len_m: article m's total number of words.)

LDA is an unsupervised machine learning technique that automatically categorizes words into interpretable topics based on their associated occurrences in articles. It estimates each topic's word composition (ϕ_l) and each article's attention allocation to the topics (θ_m). Then, the news articles on each day are aggregated into the daily topic attention levels θ_τ. Lastly, the innovation in attention is calculated against the ex ante mean level. In the baseline configuration, we take the moving average of the previous five days as the mean level, $z_\tau := \theta_\tau - \frac{1}{5}\sum_{\iota=1}^{5}\theta_{\tau-\iota}$. More details about the empirical construction and its alternatives are in 4.1.1.

Model (1) is a first attempt to formally connect news to systematic risks. Admittedly, one could argue the narrative-to-state variable relationships are neither linear nor time-invariant. Moreover, LDA topic analysis is a method based on bag-of-words, which loses considerable details contained in the context of the language. Still, this paper shows encouraging successes in bringing news to risks, even with the somewhat coarse model. We are confident more sophisticated natural language processing (NLP) and econometric techniques could well deliver better pricing performances and generate more economic insights along this route.

[13] We omit "and phrases" henceforth and just call them "words" for short. By "words and phrases," we specifically mean bi-grams are included in the bag-or-word construction. For example, "economic downturn" and "highest levels" are seen in Figure 1. See other details of the bag-or-word construction in Appendix A.1.

2.2 The Return Side of the Model

The modeling of the stock return cross section is canonical. Let f_τ be the projection of x_τ onto the excess return space (zero-cost payoffs). We refer to x_τ specifically as the state variables and their mimicking portfolios f_τ as the systematic risk factors. We are interested in estimating f_t to examine the asset pricing properties of our narrative-based asset pricing model. Let $\nu_\tau := x_\tau - f_\tau$ be the projection residual, which is orthogonal to any excess return. As a result, we could rearrange model (1) such that

$$z_\tau = Af_\tau + g_\tau, \tag{3}$$

where $g_\tau := A\nu_\tau + \eta_\tau$ is the composite residual of z_τ. We have assumed both ν_t and η_τ are uncorrelated with any excess return, hence so is g_τ.

The panel of stock returns (in excess of the risk-free rate) follows a latent factor structure:

$$r_{i,t} = \beta_{i,t} f_t + \epsilon_{i,t}. \tag{4}$$

ICAPM theory implies that there is no α, and that $\epsilon_{i,t}$ has zero expectation and is orthogonal to f_t. We assume $\beta_{i,t}$ is slow-moving, allowing the stock's exposure to fundamental risks to drift as the firm evolves.

Given the observed narrative innovations are associated with underlying risks (Eq. 3), the covariances between an asset's return and narratives shall serve as instruments that provide guidance about the asset's risk exposures (β). Let us formally write out the instrumental mapping, which is the foundation of the IPCA-based estimation method introduced in the next section. The covariance between stock return and narrative innovations is $cov_{i,t} := \mathbb{C}\text{ov}\left(r_{i,\tau}, z_\tau \mid \beta_{i,t}\right) = \beta_{i,t}\Sigma_{\text{ff}}A^\top$ ($1 \times L$ vector, calculated from Eq. 3 and 4). This expression is a conditional covariance given a particular $\beta_{i,t}$. If $\beta_{i,t}$ were constant over time, we could equivalently use the unconditional covariance. Since $\beta_{i,t}$ is slow-moving and Eq. 4 also holds at the daily frequency, $cov_{i,t}$ can be estimated with realized daily time-series covariance between $r_{i,\tau}$ and z_τ in the trailing window up to time t (detail in Eq. 6 below). Invert the above expression to express risk exposure ($\beta_{i,t}$) in terms of the estimable narrative

covariances:

$$\beta_{i,t} = cov_{i,t} A \left(A^\top A\right)^{-1} \Sigma_{\text{ff}}^{-1} := cov_{i,t} \widetilde{\Gamma}, \tag{5}$$

where $L \times K$ matrix $\widetilde{\Gamma} := A \left(A^\top A\right)^{-1} \Sigma_{\text{ff}}^{-1}$ parameterizes the instrumental mapping from $cov_{i,t}$ to $\beta_{i,t}$.

Equations (4) and (5) matches the IPCA framework, which simultaneous estimates the instrumental mapping (parameterized by $\widetilde{\Gamma}$) and the latent factors f_t. Once $\widetilde{\Gamma}$ is estimated, A can be accordingly backed out by the reverse relationship.[14]

3 Estimation Method

3.1 Estimation Procedure Summary

Model (1)–(4) is estimated by the following procedure, where steps 1 and 2 are the upgraded counterparts to the Fama-MacBeth two-step regressions, respectively.

1. Calculate covariances between $r_{i,\tau}$ and z_τ at daily frequency in a trailing window every month (t)

$$\widehat{cov}_{i,t} := \sum_\tau \kappa(\tau; t) r_{i,\tau} z_\tau^\top - \left(\sum_\tau \kappa(\tau; t) r_{i,\tau}\right) \left(\sum_\tau \kappa(\tau; t) z_\tau^\top\right), \qquad \forall i, t \tag{6}$$

   where $\widehat{cov}_{i,t}$ is a $1 \times L$ row vector, and $\kappa(\tau; t)$ is an exponentially decaying weighting function (kernel) of the trailing window that ends before the start of month t.[15]

2. Append a constant 1 to the covariances to form a set of $1 \times (L+1)$ instruments $c_{i,t} := [1, \widehat{cov}_{i,t}]$, which is supplied to the IPCA model

$$r_{i,t} = c_{i,t} \Gamma f_t + e_{i,t}, \qquad \forall i, t, \tag{7}$$

   where Γ is $(L+1) \times K$, whose rows are indexed from 0 to L ($\Gamma := [\Gamma_0; \Gamma_1; \ldots; \Gamma_L]$), such that equivalently $c_{i,t}\Gamma = \Gamma_0 + \sum_{l=1}^{L} \widehat{cov}_{l,i,t} \Gamma_l$.

[14] The expression of A in terms of $\widetilde{\Gamma}$ is $A = \widetilde{\Gamma} \left(\widetilde{\Gamma}^\top \widetilde{\Gamma}\right)^{-1} \Sigma_{\text{ff}}^{-1}$. It is reversed from the definition $\widetilde{\Gamma} := A \left(A^\top A\right)^{-1} \Sigma_{\text{ff}}^{-1}$. To do that, first, $\widetilde{\Gamma} \Sigma_{\text{ff}} = A \left(A^\top A\right)^{-1}$. Then, $\Sigma_{\text{ff}} \widetilde{\Gamma}^\top \widetilde{\Gamma} \Sigma_{\text{ff}} = \left(A^\top A\right)^{-1}$. Plug that into the first, $A = \widetilde{\Gamma} \Sigma_{\text{ff}} \left(A^\top A\right)$.

[15] In the baseline setup, we use exponentially decaying weights in the trailing window. That is $\kappa(\tau; t) := \xi^{(t - t_\tau)} / (\sum_{\tau < \tau_t} \xi^{(t - t_\tau)})$ for $\tau < \tau_t$ and 0 otherwise, where $\xi := 0.99$, τ_t is the last day before the start of month t, and t_τ is the month of date τ.

Estimate $\{f_t\}$ and Γ with the Sparse IPCA method, which is defined as the optimization:

$$\min_{\Gamma,\{f_t\}} \frac{1}{2} \sum_{i,t \in \mathbb{S}} (r_{i,t} - c_{i,t}\Gamma f_t)^2 + \lambda N_{\mathbb{S}} \sum_{l=0}^{L} \sigma_l^c \|\Gamma_l\|_2 + \sum_{t \in \mathbb{S}} \|f_t\|_2^2 , \tag{8}$$

where λ is the regularization constant (tuning detailed in 3.5); $N_{\mathbb{S}}$ is the number of $\{i,t\}$ observations in the sample panel ($\mathbb{S}$); σ_l^c is the standard deviation of $\widehat{cov}_{l,i,t}$ across $\mathbb{S}$ (with σ_0^c assigned as 1); and $\|\cdot\|_2$ is the Euclidean norm (L^2 norm).

⟨3⟩ Wrap-up step. Given estimated $\widetilde{\Gamma} := [\Gamma_1; \ldots; \Gamma_L]$, back out the estimate of A (formula in footnote 14), and construct the estimated state variables x_τ as $\left(A^\top A\right)^{-1} A^\top z_\tau$.

Next, we explain the estimation construction step by step.

3.2 Upgrading Fama-MacBeth with an IPCA Second Step

The two-step estimation can be seen as an upgrade to the Fama-MacBeth procedure to deal with situations where the state variables need to be reduced and selected from the observed innovations. In the degenerate case where each narrative represents a standalone state variable, Fama-MacBeth finds the mimicking portfolios f_t of the observed state variables. In detail, suppose $L = K$ and $A = \mathbb{I}_K$ in model (1), then the first step calculates times-series covariances between each stock and the state variables ($\widehat{cov}_{i,t}$).[16] The second step runs cross-sectional regressions $r_{i,t} = \widehat{cov}_{i,t} f_t + e_{i,t}$ at every t to estimate the K mimicking portfolios f_t.

In our setting, the observed z_τ is an L-dimensional expansion of the underlying state variables. The first step is largely the same as Fama-MacBeth, except for using daily time series. Being able to work in higher frequencies is a big advantage of building state variables with textual data, as macroeconomic indicators are hardly observed more often than monthly.[17] The main innovation is in the second step. Each stock's covariances with narrative shocks contain instrumental information about the stock's $\beta_{i,t}$ (rather than directly reveals $\beta_{i,t}$ in the degenerate case). According to (5), matrix $\widetilde{\Gamma}$ parameterizes the mapping from the L-dimensional $\widehat{cov}_{i,t}$ to the K-dimensional $\beta_{i,t}$. IPCA estimates $\widetilde{\Gamma}$ and the latent factors f_t simultaneously.

[16] Originally, the first step of Fama-MacBeth outputs regression coefficients rather than covariances. It is equivalent to use covariances for finding the factor space. See Feng et al. (2020) for a discussion, who also use covariances.

[17] Manela and Moreira (2017) raised the similar point on textual data at higher frequency.

When implementing the IPCA estimation, we extend the linear instrumental mapping to allow for a constant term. To wrap it in the matrix form, we append a constant 1 to the L covariances and stack $1 \times K$ parameters Γ_0 on top of $\widetilde{\Gamma}$. As a result, $\Gamma = [\Gamma_0; \widetilde{\Gamma}]$ and $c_{i,t}\Gamma = \Gamma_0 + \widehat{cov}_{i,t}\widetilde{\Gamma}$. Once Sparse IPCA estimates both $\{f\}$ and Γ, the wrap-up step $\langle 3 \rangle$ cuts out $\widetilde{\Gamma}$ from the estimated Γ, from which matrix A is backed out accordingly.

3.3 Narrative Selection with Sparse IPCA

As mentioned in 2.1, some narratives are allowed to be entirely irrelevant for risks, leaving the corresponding rows of A with zero entries. Such narratives' covariances will then be irrelevant instruments for $\beta_{i,t}$. That is to say Γ inherits A's row-wise sparsity structure—if a narrative l is irrelevant, the l'th row of both A and $\widetilde{\Gamma}$ are all zero entries. Therefore, we conduct narrative selection by inducing sparsity in Γ when fitting the IPCA model.

Sparse IPCA's target function (8) is the original IPCA's target plus two new terms for regularization. The original target is the model fit term calculated as the sum of squared errors. It captures how well the latent factors fits realized stock returns. The second term is the key that facilitates variable selection. It penalizes the Euclidean norms of each Γ row $\|\Gamma_l\|_2$ and corresponds to a standard group lasso penalty. Each penalty term achieves the non-differentiable minimum at $\Gamma_l = \mathbf{0}_{1\times K}$. The optimization balances a row's benefit for improving the model fit with the penalty of being away from $\mathbf{0}_{1\times K}$. As a result, all rows are shrunk towards $\mathbf{0}_{1\times K}$, and the rows that do not contribute enough for the model fit will be *entirely* zeroed out. The corresponding instruments become irrelevant for the final estimate, effectively resulting in a more parsimonious model with only a subset of selected instruments.

We design Sparse IPCA with the row-wise group lasso regularization instead of the simpler element-wise lasso which penalizes the absolute value of each element $|\Gamma_{l,k}|$. The reason is it is not meaningful to distinguish *for which* factor an instrument matters or not, since the K individual factors are rotationally unidentified anyway. Instead, it penalizes the norm of Γ_l without distinguishing the direction in which it deviates from $\mathbf{0}_{1\times K}$.[18]

Each narrative's penalty is controlled by three weights: λ, σ_l^c, and $N_{\mathbb{S}}$. The aggregate penalty of all instruments is scaled by *regularization constant* λ, which controls the relative strength of the

[18] The penalty term is still invariant to an orthonormal rotation (which rotates Γ to ΓR with an orthonormal matrix R), preserving the IPCA property.

regularization versus model fit. A greater λ returns a more parsimonious model with more shrinkage and less selected instruments. It is the only hyperparameter subject to tuning, which is detailed in subsection 3.5. The purpose of σ_l^c is to place the strength of regularization of each narrative on the same scale. We are effectively regularizing the panel-wise standard deviation of $\|c_{l,i,t}\Gamma_l\|_2$, which is the part of $\beta_{i,t}$ variation contributed by instrument l.[19] Lastly, $N_{\mathbb{S}}$ (number of $\{i,t\}$ in $\mathbb{S}$) re-scales λ such that the magnitude of the regularization term keeps up with the model fit term as the sample size changes. The re-scaling is necessary with expanding window based OOS construction (appearing in 6.1). Otherwise, the effective strength of the same λ value would be varied across samples with different sizes.

The third term is added for a technical reason. Notice the model fit term is invariant if we shrink Γ and expand f_t by the same multiple. Therefore, without the third term's restriction on f_t, the minimization will return an infinitesimal Γ that bypasses its penalties. The third term is merely to balance the shrinkage effect on Γ, by also regularizing the sum of squared of f_t.

3.4 Numerical Minimization Algorithm of Sparse IPCA

Minimization problem (8) is solved numerically with an Alternating Regularized Least Squares (ARLS) algorithm.It alternates between minimizing over Γ while holding $\{f_t\}$ fixed, and minimizing over $\{f_t\}$ while holding Γ fixed.[20] The process is terminated when the joint target function's descent is small (or when the first order condition is satisfied) within a numerical tolerance.

This algorithm is similar to the un-regularized IPCA's Alternating Least Squares (ALS) method, except that the two sub-problems become *regularized* least squares. In particular, the Γ sub-problem is a group lasso regression on the $\{i,t\}$ panel:

$$\min_{\Gamma} \frac{1}{2} \sum_{\{i,t\}\in\mathbb{S}} \left(r_{i,t} - \left(c_{i,t} \otimes f_t^\top \right) \text{vect}\left(\Gamma\right) \right)^2 + \lambda N_{\mathbb{S}} \sum_{l=0}^{L} \sigma_l^c \left\| \Gamma_l \right\|_2 , \tag{9}$$

where $c_{i,t} \otimes f_t^\top$ constitutes the $(L+1)K$-variate regressor.[21] We solve the group lasso regression

[19] The adjustment with σ_l^c indeed follows the conventional procedure in lasso regressions, where regressors are standardized in order to bring the coefficients to the same level such that the coefficients are subject to the same strength of regularizations. We do not want to simply standardize the regressors (in our case $c_{l,i,t}$), as it will change the scale of the coefficients (in our case Γ). Instead, the strength of the penalty is adjusted by σ_l^c directly, such that the scale (and the interpretation) of Γ is preserved.

[20] The ARLS algorithm can be seen as special case of the Block Coordinate Descent algorithm, with the two parameters as the two "blocks."

[21] Notation: $\text{vect}\left(\Gamma\right)$ ($LK \times 1$ vector) is the vectorization of Γ that transposes and stacks up the rows of Γ.

numerically using Yang and Zou's (2015) algorithm. The $\{f_t\}$ sub-problem is done period by period by solving the cross-sectional ridge regression

$$\min_{f_t} \frac{1}{2} \|r_t - C_t \Gamma f_t\|_2^2 + \|f_t\|_2^2 , \tag{10}$$

with the analytical solution $f_t = \left(\Gamma^\top C_t^\top C_t \Gamma + 2\mathbb{I}_K\right)^{-1} \Gamma^\top C_t^\top r_t$.[22]

Both the two sub-problems easily adapt to unbalanced panels, hence so does Sparse IPCA overall. As a result, Sparse IPCA is applicable to applications with a large panel and missing entries like stock returns.

3.5 Tuning Regularization Constant λ in Sparse IPCA

The regularization constant λ controls the strength of the narrative selection. As a hyperparameter, each λ gives a different estimate of parameters $\Gamma, \{f_t\}$. The tuning of λ is based on the criteria that the true systematic risk factors' MVE portfolio shall attain the highest Sharpe ratio among all excess returns.

In detail: in sample panel $\mathbb{S}$, given a λ and the resulting Γ and $\{f_t\}$, the (annualized) Sharpe ratio of the factor MVE is $\mathrm{SR}(\lambda; \mathbb{S}) := \sqrt{12 \mu_{\mathrm{f}}^\top \Sigma_{\mathrm{ff}}^{-1} \mu_{\mathrm{f}}}$, where μ_{f} and Σ_{ff} are the mean and variance of f_t over $\mathbb{S}$. We pick the λ that maximizes the Sharpe ratio: $\lambda_{\mathbb{S}}^* := \arg\max_\lambda \mathrm{SR}(\lambda; \mathbb{S})$.[23] The corresponding Γ, A, etc. (estimated at $\lambda_{\mathbb{S}}^*$) are the tuned estimates, which we report and analyze further below. When it comes to evaluating the asset pricing performances of the factors, we focus on out-of-sample performance and form estimates in a series of training samples which we later evaluate in separate testing samples. The out-of-sample factor construction will be detailed in Section 6.1.

[22] Notation: r_t ($N_t \times 1$ vector) is the cross section $r_{i,t}$ at time t. Similarly, C_t ($N_t \times K$ matrix) is $c_{i,t}$'s stacked up.

[23] It could be a concern that $\lambda_{\mathbb{S}}^*$ might suffer from in-sample overfitting in the Markowitz optimization, which might hurt out-of-sample performances. An alternative λ tuning method based on leave-one-out mean-variance efficient portfolio formation addresses the concern. The detailed construction and empirical results are in Appendix B.2, where we show little, if any, empirical improvement compared to the simpler and more transparent tuning method reported in the main text.

4 Estimation Results

4.1 Data

4.1.1 From news text to narrative innovations via textual analysis

We construct daily narrative attention levels (θ_τ) using LDA following BKMX. The topic analysis is conducted with the complete archive of the Wall Street Journal articles from 1984:01 to 2017:06, which yields $L = 180$ topics.[24] The details about the LDA estimation method of Φ, θ_m, and subsequently aggregating to the daily levels (θ_τ) are relegated to Appendix A.2. The label of each of topic, like "Recession" and "Record high," is manually assigned by summarizing the common theme displayed in the automatically generated topic words (ϕ_l).[25] As examples, the ϕ_l and $\theta_{\tau,l}$ for the few narratives that are later found to be the most relevant for risks are visualized in Appendix A.3.[26]

Once daily attention levels θ_τ are in place, we extract the unexpected innovations by subtracting the ex ante conditional mean, which is specified as the rolling average of the previous five days, $z_\tau := \theta_\tau - \frac{1}{5}\sum_{\iota=1}^{5}\theta_{\tau-\iota}$. We use this simple specification to account for the persistence in z_τ across all the narratives, which arguably could admit different time-series dynamics. We conduct a series of robustness checks with respect to the methods of filtering the conditional mean in θ_τ, and find that our results are robust to these perturbations.[27] We leave a more complete analysis of z_τ's time-series properties, including conditional variances and trend as well as cyclicality and the frequency spectrum for future works.

A caveat about date indexing is that θ_τ is from the WSJ published *the next morning* of calendar day τ, such that it reflects the market information accrued on day τ. In this way, z_τ is synchronized with asset returns $r_{i,\tau}$, which are available by market closing in the afternoon on day τ, then their covariances can be calculated accordingly.[28]

[24] The number of topics is determined by maximizing the Bayes factor of LDA.

[25] We follow the labels used in BKMX and do not tweak the labels to fit the subjective analysis of this paper.

[26] The visualizations of all the topics can be found in structureofnews.com.

[27] Results for these robustness checks are presented in Appendix B.3. In addition to alternative daily moving-average specifications, we also use monthly innovations and macroeconomic series as candidate state variables (z_t).

[28] To avoid the potential leak of ex post information after the start of month t into the ex ante portfolio weights of f_t, $\widehat{cov}_{i,t}$ is calculated in the window up to the *second last day* before the start of month t. See footnote 15.

4.1.2 Stock return data

We use the CRSP dataset for individual stock returns (in excess of the risk-free rate) of firms listed on the NYSE, AMEX and NASDAQ.[29] To match the span of our news data, daily stock returns from 1984:01 to 2016:12 are used to construct monthly covariances $\widehat{cov}_{i,t}$ (following Eq. 6).[30] After a one-year burn-in period to prepare the earliest $\widehat{cov}_{i,t}$, the *full sample* spans 32 years from 1985:01 to 2016:12 and contains 1,850,401 stock-month observations of $\{r_{i,t}, \widehat{cov}_{i,t}\}$, 15,831 unique stocks, or on average around 4,800 stocks per month.

4.2 Estimates Under Different Regularization Constants (λ)

We report how the estimates of $\Gamma, \{f_t\}$ evolve as the regularization constant varies in order to show Sparse IPCA is effective in selecting the relevant narratives and the tuning method is robust. Figure 3 reports four statistics of full-sample estimates along the path of varying λ values. The four statistics are:

1. **Total** R^2, defined as $1 - \sum_{i,t} (r_{i,t} - c_{i,t}\Gamma f_t)^2 \,/\, \sum_{i,t} r_{i,t}^2$, reports the model fit in terms of the proportion of realized return variation explained by the factors. It simply transforms the first term in the Sparse IPCA target (8) (the sum of squared errors) to the R^2 format for ease of interpretation.

2. **Factor MVE's Sharpe ratio**, defined as $\mathrm{SR}(\lambda; \mathbb{S}) := \sqrt{12 \mu_{\mathrm{f}}^\top \Sigma_{\mathrm{ff}}^{-1} \mu_{\mathrm{f}}}$, is the maximization target for tuning the regularization constant λ.

3. **Number of selected instruments** is the number of rows of Γ that are not all zeroes.

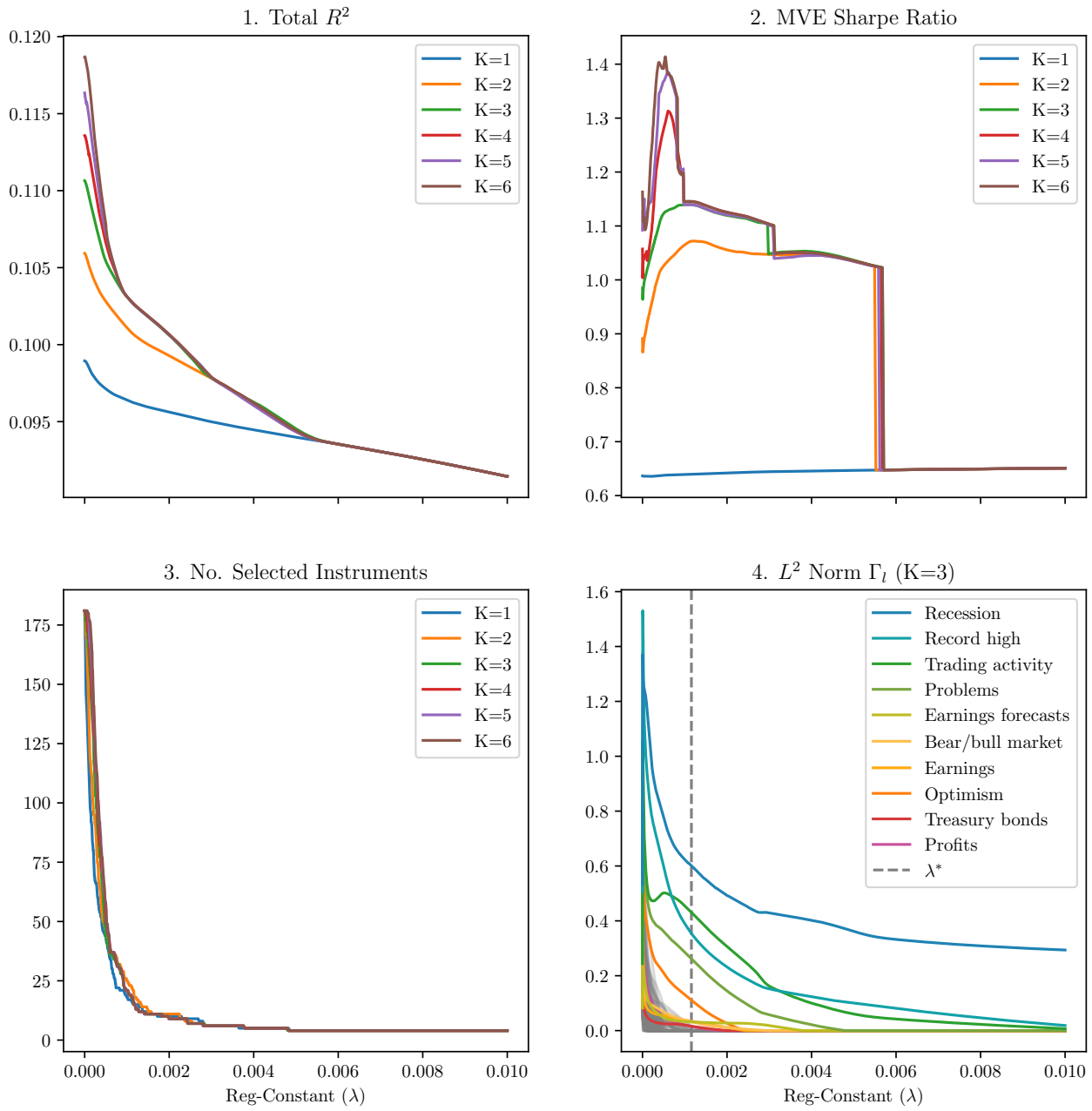
4. L^2 **Norm of each row of** Γ, defined as $\|\Gamma_l\|_2$, measures the marginal "contribution" of each instrument to the K factor loadings ($\beta_{i,t}$). Sparse IPCA penalizes these L^2 norms.

In each of the four panels, the left end starting point is the unregularized IPCA ($\lambda = 0$). As λ increases, the optimization (Eq. 8) puts more weight on the Γ penalty term relative to the model fit term. As a result, model fit (in terms of the Total R^2) deteriorates (Panel 1). Meanwhile, the quantities being penalized decrease. The rows of Γ are shrunk towards $\mathbf{0}$ (Panel 4) and are zeroed

[29] The sample is restricted to firms with observable Fama-French characteristics (Fama and French, 2016) to align with stocks used to form the benchmark factors. Using the full set of 24,162 firms which does not change the results.

[30] Effectively, only the z_τ on days when the market was open during 1984:01–2016:12 are used in calculating $\widehat{cov}_{i,t}$.

Figure 3: Statistics of full-sample estimates at each λ value



out one-by-one, resulting in more and more parsimonious models with fewer and fewer selected instruments (Panel 3).

In contrast, the Sharpe ratio curves are not monotonic but hump shaped (Panel 2). For small λ values (left end), over-fitting in unregularized IPCA hurts the estimation of the systematic risks, resulting in lower MVE Sharpe ratios. With a moderate increase in λ comes a drastic drop in the

number of selected instruments and a significant increase in the MVE Sharpe ratio. This is the first piece of evidence that the regularization is effectively improving systematic risk estimation. On the other end, excessively high λ's filter out too many instruments, which are also not conducive for factor estimation and yield lower Sharpe ratios.[31] The hump shapes have broad and smooth tops, at least for the smaller K's. This means the selected model will be relatively stable to perturbations in the tuned $\lambda^*_{\mathbb{S}}$. These observations suggest the tuning method based on maximizing MVE's Sharpe ratio is robust.

Comparing across the dimensionality of the factor space, there are significant improvements for $K = 2, 3$ on top of $K = 1$ in terms of both Total R^2 and MVE Sharpe ratio. $K = 1$ is mundane for being close to the market portfolio. Further increasing K above 3 still pushes up full-sample MVE Sharpe ratio, although the improvement becomes sensitive to the regularization constant. The "pointy" peaks indicate the high Sharpe ratios would not be sustained with moderate perturbations in $\lambda^*_{\mathbb{S}}$, which could dampen some of the OOS performances (later shown in Table 2). Hence we will use $K = 3$ as the benchmark configuration. The narrative-based factor model has roughly the same dimensionality as characteristics-based models of around 3$\sim$6. It suggests that the factor space of stock returns is still parsimonious after switching from characteristics as conditioning information to narratives covariances. Yet, we do not want to rule out the possibility that a finer textual analysis might identify a richer set of systematic risks in future works.

Panel 4 reports the evolution of Γ estimates along the λ path in the benchmark case of $K = 3$. It shows that group-lasso based Sparse IPCA has both shrinkage and selection effects. The pattern is the same as that of the canonical lasso, implying that Sparse IPCA has effectively generalized the lasso to the context of latent factor analysis. In detail, as λ increases, each instrument's contribution to $\beta_{i,t}$ (measured by $\|\Gamma_l\|_2$) first shrinks toward zero, and eventually drops to and stays at zero after a cutoff λ. We denote the cutoff as $\max_l^\lambda$, which is the maximum λ at which an instrument is still selected in the model. An instrument with a higher $\max_l^\lambda$ is more prone to be included, hence $\max_l^\lambda$ is a rough measure of an instrument's relevance. We color and label the ten narratives with the highest $\max_l^\lambda$. The rest are in gray in the background. One can already tell from the top narrative names in the legend that they are closely related to business and economics in general. The next section discusses the interpretation of fundamental risks in terms of the selected narratives in detail.

[31] We also report that with excessively high λ's the rank of Γ drops gradually from K to $K-1, K-2, \ldots$. This is why the curve "collapses" to the one below it with a smaller K as λ increases.

The tuned hyperparameter is marked by the vertical dashed line, which is at the peak of the Sharpe ratio curve (the green "$K = 3$" curve in Panel 2).

To further demonstrate that the selection method is effective in distinguishing relevant and irrelevant instruments, we conduct an experiment with simulated placebo narratives. In addition to the 180 observed narratives (z_τ), we randomly generate an equal amount of additional narratives as placebos to "jam" the estimation. The results in Appendix B.1 show Sparse IPCA can effectively filter out the placebo narratives that we know for sure are irrelevant. Moreover, the estimates of the selected narratives are largely unaffected by the interference of the irrelevant ones.

5 Textual Interpretation of Fundamental Risks

Model Eq. 1 provides a link between narrative innovations and state variables, which enables the interpretation of the fundamental risks by directly reading the associated text. Given any piece of news, we can infer how it impacts the state variables. This relationship reveals how news narratives are taken into consideration by investors to adjust their pricing kernel, and thereby sheds light on the concrete meanings of the fundamental risks.

5.1 Interpretation Method

The interpretation method can be seen as an impulse response analysis of the impact of news on state variables. Imagine a hypothetical state of the world s, where the source of the impulse is a new piece of text that brings narrative innovation $z(s)$. The impulse s can be the aggregate news during a day ($z(s) = z_\tau$), a single news article ($z(s) = z_m$), or even a particular word. Given Eq. 1, narrative attention innovations impact state variables according to

$$x(s) = (A^\top A)^{-1} A^\top z(s) := I_{z\to x}^\top z(s), \tag{11}$$

where $L \times K$ matrix $I_{z\to x} := A(A^\top A)^{-1}$ summarizes any z's (L-dimensional) impact on x (K-dimensional).[32]

Model (1)–(4) only identifies the factor space, in the sense that if A, x, β, f are rotated simul-

[32] As an impulse response analysis, the data generating errors are assumed to be zero, e.g. $\eta(s) = \mathbf{0}$.

taneously, the model is invariant.[33] Some of the research goals, for example the asset pricing tests in Section 6, can be achieved by identifying just the factor space but not a particular basis. When it comes to interpreting the risks though, we need to be specific about which "direction" of risk is of interest, and focus on one meaningful linear combination of x (or f) at a time. Therefore, we examine the impact on particular linear combinations expressed as $x^q := b^q x$, where b^q is the $1 \times K$ linear combination weights such that x^q has a concrete meaning. Our primary interest is in interpreting $x^{\text{MVE}} := b^{\text{MVE}} x$, with $b^{\text{MVE}} := \mu_{\text{f}}^\top \Sigma_{\text{ff}}^{-1}$, such that x^{MVE}'s mimicking portfolio is the MVE combination of the systematic factors ($f^{\text{MVE}} := b^{\text{MVE}} f$). The MVE state variable is special for being the univariate pricing kernel, which in theory is the single source of cross-sectional risk premium differences: $\mathbb{E}_{t-1} r_{i,t} = \gamma^{\text{MVE}} \mathbb{C}\text{ov}_{t-1}\left(r_{i,t}, x_t^{\text{MVE}}\right)$, for any excess return $r_{i,t}$.[34] Equivalently, the SDF is simply x_t^{MVE} after a linear transformation.[35] The method also supports interpreting pre-specified tradable factors such as $q_t = \text{Mkt}_t$. In these cases, we look for a b^q such that $f_t^q := b^q f_t$ is the projection of the pre-specified series q_t onto the factor space.[36] Thereby, $x_t^q := b^q x_t$ is the state variable whose mimicking portfolio is the closest to q_t and retains the concrete meaning of q_t. In either case, the impulse response of x^q is

$$x^q(s) = \underbrace{b^q (A^\top A)^{-1} A^\top}_{:= I_{z\to q}^\top} z(s), \tag{12}$$

where $L \times 1$ vector $I_{z\to q}$ summarizes the *impact* of each narrative on the state variable x^q.

We can trace the sources of the impulse one step further from the narrative level to the word level. Suppose in the hypothetical state s, a new piece of text brings changes in word frequencies $\Delta w(s)$ (a $V \times 1$ vector, V is the size of the vocabulary). Given the LDA model (2), the word frequency impulse (V-dimensional) translates into narrative attention innovations (L-dimensional) as $z(s) = \left(\Phi^\top \Phi\right)^{-1} \Phi^\top \Delta w(s)$. The induced narrative-level innovations then impact state variables

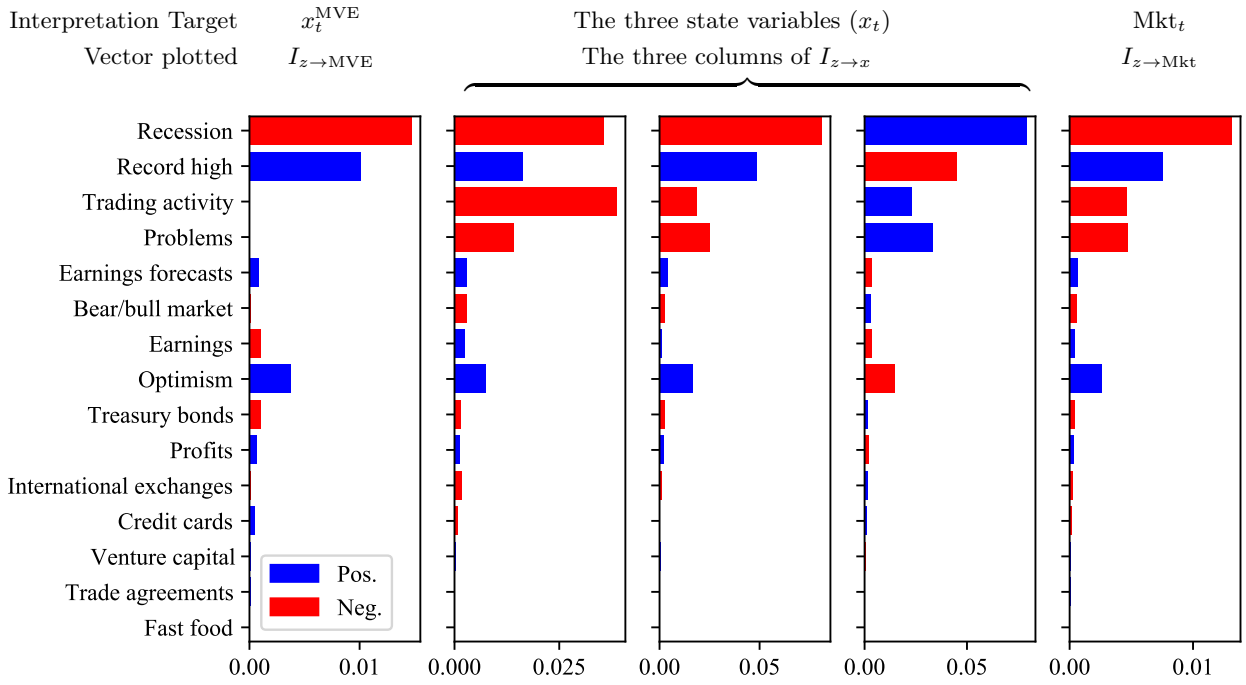
[33] Specifically, we mean the model is invariant if A, x, β, f changes to $AR^{-1}, Rx, \beta R^{-1}, Rf$, respectively, for some $K \times K$ invertible matrix R. The unidentification problem is the same as in IPCA and virtually all latent factor models, see Kelly et al. (2017) for a general discussion.

[34] $\gamma^{\text{MVE}} = \mathbb{E} f_t^{\text{MVE}} / \mathbb{V}\text{ar} f_t^{\text{MVE}}$. This theoretical result is per Hansen and Richard (1987) decomposition.

[35] The narrative-based SDF is $m_t = \rho + \zeta x_t^{\text{MVE}}$, with $\rho = 1/R^{\text{f}} + \left(\mathbb{E} x^{\text{MVE}}\right)^2 / \left(R^{\text{f}} \mathbb{V}\text{ar} x^{\text{MVE}}\right)$ and $\zeta = -\mathbb{E} x^{\text{MVE}} / \left(R^{\text{f}} \mathbb{V}\text{ar} x^{\text{MVE}}\right)$. Throughout the paper, we work with the factor form rather than the SDF form, and refer to x^{MVE} as the univariate "pricing kernel."

[36] In detail, $b^q := \mathbb{E}\left[q_t f_t^\top\right] \mathbb{E}\left[f_t f_t^\top\right]^{-1}$. If the projection R^2 is not high, meaning q_t is not spanned by f_t, then the method would not be reliable. In this general perspective, f^{MVE} is indeed the "constant" mimicking portfolio weights with the assignment $q_t = 1$.

Figure 4: Fundamental Risk Interpretation at the Narrative Level



Note: This figure reports the narrative-level risk factor interpretation results by plotting the factors' impact vectors/matrices. The lengths of the bars represent the absolute values of corresponding entries, with red color for negative impact and blue for positive.

according to (12). Combined, the word-to-state variable impulse response is

$$x^q(s) = \underbrace{b^q(A^\top A)^{-1}A^\top\left(\Phi^\top\Phi^\top\right)^{-1}\Phi^\top}_{:=I_{w\to q}^\top}\Delta w(s). \tag{13}$$

Notice $I_{z\to q}$ is a linear combination of the L columns of Φ. Hence the word-level impacts ($I_{z\to q}$) can be seen as a *composite narrative* that combines the basis narratives (ϕ_l's) according to their impacts to x^q. Notice unlike the bases, which consist of only non-negative entries, $I_{z\to q}$ has both positive and negative entries because a state variable (x^q) can be impacted in both directions.

5.2 Interpretation Results

Next, we report these impact vectors to provide a human-readable interpretation of the fundamental state variables. They are calculated with the benchmark estimation with the full sample and three factors ($K = 3$).

Table 1: Example News Articles of the Three Most Prominent Narratives

	Recession
1993-05-07	Auto Registrations Continued to Slump In Europe Last Month
2001-04-25	Consumer Confidence Slides on Fears of Layoffs
2009-02-19	U.S. News: Housing Starts Hit Lowest Level In Half-Century
2011-08-02	World News: Manufacturing Slowdown Adds to Gloom on Economy
2016-07-08	World News: U.K. Consumer Sentiment Takes Dive
	Record high
1989-07-05	Japan Vehicle Sales Rise
1994-07-01	Purchasing Managers In U.K. Survey Report Rise for June Orders
1995-02-27	Hiring Outlook For Second Quarter Appears Vigorous
2006-01-12	Wall Street Bonuses Hit a Record in 2005
2016-07-20	U.S. News: Home Building Continues Recovery as Demand Rises
	Trading activity
1993-12-30	Industrials Rise A Bit to Record; Bonds Decline
1994-10-20	Profit News Helps Boost Stock Prices — Indexes Gain Ground Despite Weakness Of Bonds and Dollar
1996-06-21	Nasdaq Sinks Amid Sell-Off Of Tech Stocks
1997-12-09	Blue Chips Fall As Dollar's Rise Causes Concern
1998-04-21	Drug Stocks Resume Gains; Blue Chips Fall

Note: Example articles with the highest attention to the three narratives, respectively (titles only, see text bodies in Appendix C.1). The shades of the yellow highlighter on each term reflects $\phi_{v,l}$, the term's loading on the corresponding narrative.

Figure 4 Panel 1 reports each narrative's impact on the MVE state variable ($I_{z\to\text{MVE}}$), the next three panels report the impacts on the three entries of x_t separately (the three columns of $I_{z\to x}$). The last panel uses the market portfolio as the interpretation target. The rows only include the selected narratives, since the excluded ones have zero impact.

The narratives fall into three categories according to their impacts on the MVE state variable: negative, positive, and orthogonal. The most prominent representatives of each category are "Recession," "Record high," and "Trading activity," respectively. To dissect the concrete contents of these narratives, Table 1 reports example articles with the highest attention to these three narratives. We highlight the words that LDA "recognizes" as related to each of three narratives, that is the words with a high $\phi_{v,l}$. Aggregating these highlighted words within an article gives rise to the article-level attention to the corresponding narrative.

Concerns about economic downturns and recessions exert a negative impact on the pricing kernel to the greatest extent. The "Recession" narrative contains words that convey negative news updates

like "downturn," "gloom," "slump" etc. When there is heightened attention to the "Recession" narrative, x^{MVE} is negatively adjusted. A reduction in x^{MVE} corresponds to an increase in the marginal utility of consumption (or equivalently the SDF). It implies risky assets that positively correlate with "Recession" reporting provide hedging benefits, hence should earn negative risk premiums.

In the opposite direction, narratives like "Record high" and "Optimism" positively impact x^{MVE}. A positive innovation in these narratives increases concurrent consumption or decreases marginal utility. As seen in Table 1, the "Record high" narrative contains good news about various aspects of the economic outlook, ranging from durable consumption to manufacturing activities. Each news article is organized around a piece of information that is quantitative in nature, such as statistic releases or survey updates. The underlying quantitative news covers a wide range of the economic events. It would be hard to collect or process comprehensive quantitative datasets to cover such extensive issues. Yet the news-based method utilizes the editorial process to locate the most pertinent quantitative events at every point in time. Moreover, it is hard to infer risk contents from numerical data for such varied issues. How much of "home building recovery" bears a comparable impact with a given "Wall Street bonus record"? Do the same "Wall Street bonus" statistics have different effects in changing economic environments? There could be complicated structural changes not to mention discontinued data when it comes to such specific issues. Once again, these difficulties are effectively left to newspaper editors. The language choice of the articles, in particular the frequency of the narrative-related words (as highlighted in Table 1), allows us to infer changes in narratives attention, from which we gain a standard way of extracting the risk content from the ever-evolving market events.

Comparing Panel 1 with the next three panels ($I_{z\to x}$) in Figure 4, we see "Trading activity" and "Problems", which have almost *no* impact on the MVE, significantly impact the individual components of x_t. The same goes with "Bear/bull market" and "International exchanges" but at smaller magnitudes. These narratives, in general, reflect a degree of activeness of the financial market, as illustrated in particular for "Trading activities" in Table 1. The articles contain news reports about market price movements in both directions. According to $I_{z\to x}$ in Figure 4, these narratives are important sources of systematic risks. Sparse IPCA selects them (and assign them large $\|\Gamma_l\|_2$) for their importance in fitting the realized stock returns. But factor premium calculations reveal that they are largely orthogonal to the MVE, which represents the univariate pricing kernel.

Therefore, an asset's covariances with these narratives has little impact on its risk premium.

By tracing the word composition of each topic, we push the interpretation from narrative level to the more granular level of words and phrases. Figure 1 (in introduction) visualizes the words' impact on the MVE ($I_{w\to\text{MVE}}$), where we separate the words with positive and negative impacts into two word clouds. This figure is a concrete display of the asset pricing kernel, which has previously been a rather abstract theoretical concept. It shows how investors update the pricing kernel given any observed news at the word level. The word clouds show two distinct extremes in the language used to describe the economic and investment outlook that one can easily recognize (yet the estimation is achieved without any human guidance on the words' semantic meanings). As explained in 5.1, $I_{w\to\text{MVE}}$ can be seen as a composite narrative made up of the basis narratives with strong impacts to x^{MVE}. The negative cloud is predominantly filled with words associated with the "Recession" narrative, whereas the positive cloud consists a blend of words in "Record high" and "Optimism."

The last panel in Figure 4 changes the interpretation target to the market return in excess of the risk-free rate ($q_t = \text{Mkt}_t$). The full-sample projection R^2 from Mkt_t to f_t is 97.6%, suggesting that the narrative-based systematic risk factors almost perfectly span the market factor.[37] We see the market factor has a similar interpretation to the MVE in terms of "Recession," "Record high," and "Optimism." This is expected as the market is often used as the proxy for the wealth portfolio which *is* the MVE factor under the (not intertemporal) CAPM. The difference lies in the additional loadings on the narratives that describe financial market activeness such as "Trading activity," "Bear/bull market," and "Problems." It is reasonable that the market factor have exposure to these aspects of the fundamental risks. But, as shown above, these narrative are orthogonal to x^{MVE}. That is to say the exposures to these narratives contribute to the systematic risk, but not to the expected return. As a result, Mkt_t is dragged away from multivariate mean-variance efficiency. We find these observations consistent with the classical ICAPM theory.

5.3 Event Retrieval

We have displayed the link from shocks in news narratives to the fundamental risk factors. Next, we apply the link to retrieve the specific news events corresponding to *realized* fluctuations in fac-

[37] The projection R^2 for the other Fama-French/Carhart factors are not as high (SMB: 60%, HML: 52%, RMW: 67%, CMA: 19%, UMD: 12%), suggesting the narrative-based factor model is different from the Fama-French/Carhart model to a large extent.

tor returns. According to the model, the textual content of each observed article has an implied quantitative impact on the risk factors. In detail, given an article m that brings narrative attention innovation $z(m)$, its impact on state variable x^q is $x^q(m) = I_{z\to q} z(m)$.[38] Among all the news on a given day, we select the article whose implied impact most lines up with the realized factor return.[39] It is particularly interesting to recover the news events behind large swings of the market factor. Hence we focus on the days with extreme $q_\tau = \text{Mkt}_\tau$ returns. The selected article on those days shall provide a concrete and human-readable account of the observed market movements. Eventually, this exercise can be applied *in real time* to "translate" from textual news to quantitative price updates as well as the other way around.

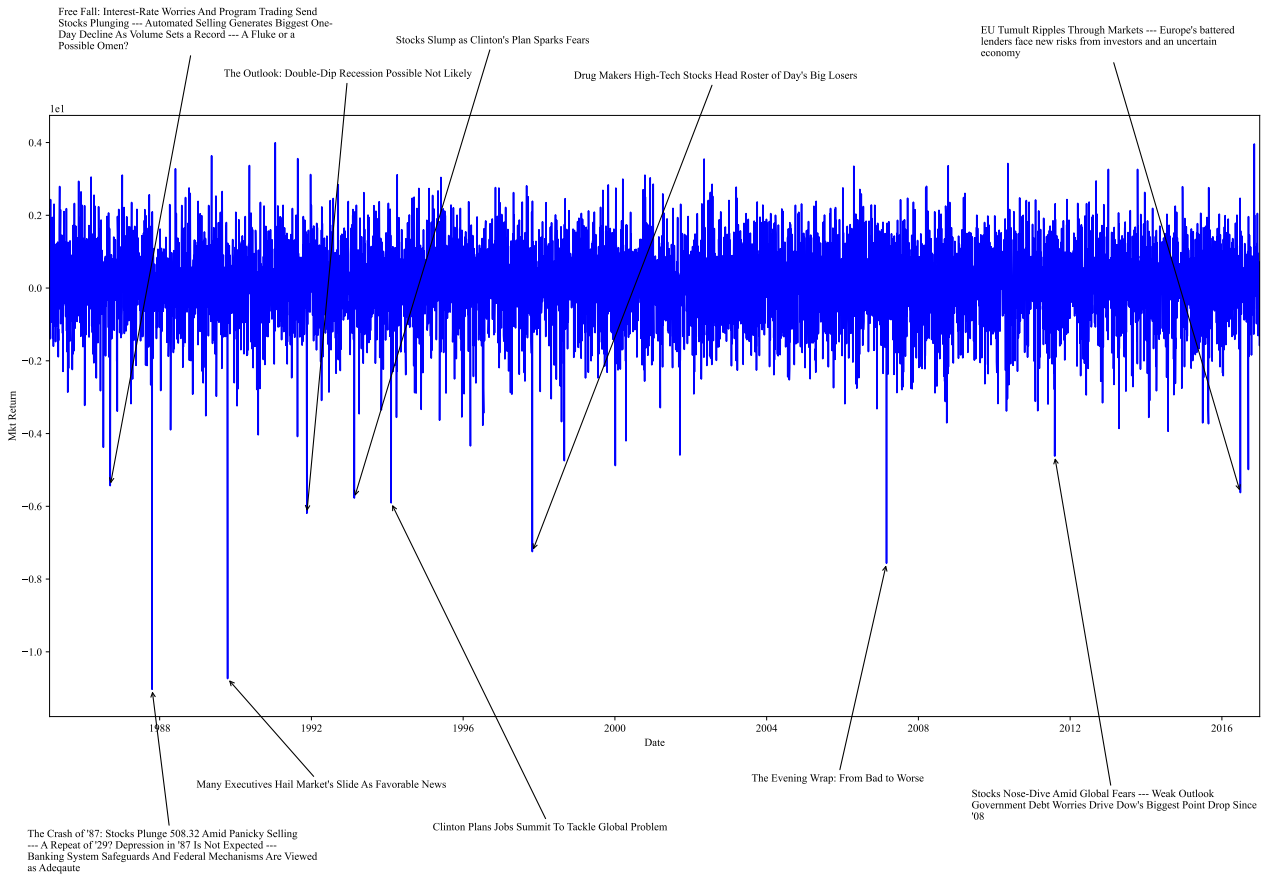
Figure 5 reports the news events behind the top ten largest daily market return spikes in the full sample period.[40] The retrieved articles shed light on specific events that triggered investor concerns and moved the market factor. For instance, we capture interest rate concerns on 1986-09-12 and Black Monday in October of 1987. Similarly, we identify policy concerns related to the Clinton administration in 1993 and 1994. Additionally, we capture the downgrading of US debt in August of 2011. The detailed news events vary considerably over the three decades. Appendix C.2 reports excerpts of the retrieved articles to display how the machine "reads" the quantitative contents from the textual articles in detail. We highlight each word according to its impact on the market return ($I_{z\to\text{Mkt}}$), with red for negative and blue for positive impacts. Aggregating these words in an article gives rise to the article's overall impact to the market return. As reported by $I_{z\to\text{Mkt}}$ (last panel in Figure 4), different narratives, such as "Recession" and "Problems," can all cause negative updates to Mkt_t. Even within the same narrative, the detailed event can be quite different over time, ranging from worries about interest rates hikes to uncertainty about a double-dipping recession. However, via the topic analysis and the narrative-based asset pricing model, we are able to retrieve the varying events that all caused large price movements throughout the sample period.

[38] We calculate $z(m)$ similar to z_τ: $z(m) := \theta_m - \frac{1}{5}\sum_{\iota=1}^{5}\theta_{\tau_m-\iota}$, where $\theta_m \in \Delta^V$ is article m's topic attention levels calculated from LDA; the summation is the average levels in the five days before the day on which article m is published.

[39] In detail, by "most lines up," we mean among all the articles on day τ, the one whose impact $x^q(m)$ has the same sign as f^q_τ and has the greatest absolute value.

[40] As representative examples, we pick 10 days with the largest absolute risk-adjusted returns within its year. The risk adjusted return is calculated by filtering out the conditional volatility: $\frac{f^{\text{Mkt}}_\tau}{\sigma^{\text{Mkt}}_\tau}$, where σ^{Mkt}_τ is the exponentially weighted conditional volatility.

Figure 5: Event Retrieval for Market Returns



Note: This figure plots the realized daily market returns and mark large return spikes with retrieved news events. Only the news article titles are reported in the picture, and the excerpts of the text bodies are in Appendix C.2.

6 Asset Pricing Performance

Now we shift gear to examine the quantitative aspects of the model. This section evaluates the asset pricing performances of the narrative-based systematic risk factors. If the factor estimates indeed capture the true systematic risks, they should pass the traditional asset pricing tests, and perform comparably to the classical characteristic-sorted portfolios. In particular, the factors' MVE portfolio should deliver a high Sharpe ratio, and the factors should price a cross section of test assets with small pricing errors (α's). We focus on comparing against a standard benchmark, the five factor model introduced by Fama and French (2016) as well as the momentum factor added by Carhart (1997). We *only evaluate the OOS estimates* of the narrative-based factors in order to avoid the standard critiques of in-sample constructions in asset pricing exercises (e.g. Lewellen et al., 2010).

6.1 Out-of-sample Factor and MVE Construction Method

The OOS factors and their MVE are constructed as tradable portfolios of individual stocks' excess returns with *ex ante available* portfolio weights, such that they can be fairly compared with characteristics-sorted portfolios in asset pricing tests.

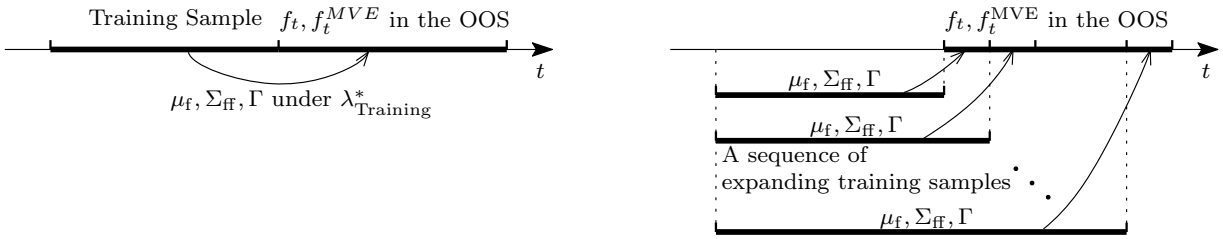
At each training sample ($\mathbb{S}$), we tune $\lambda^*_{\mathbb{S}}$ and estimate $\mu_{\rm f}, \Sigma_{\rm ff}, \Gamma$ as described in Section 4. These estimates are brought to later periods to form OOS factors and the OOS MVE, which have the same formulas as before, except using the parameters estimated in the training samples:

$$f_t = \left(\Gamma^\top C_t^\top C_t \Gamma + 2\mathbb{I}_K\right)^{-1} \Gamma^\top C_t^\top r_t, \qquad f_t^{\rm MVE} = \mu_{\rm f}^\top \Sigma_{\rm ff}^{-1} f_t, \qquad \forall t \in \text{OOS}.$$

Notice instruments (C_t) consist of narrative covariances that are available at the start of month t, parameters $\mu_{\rm f}, \Sigma_{\rm ff}, \Gamma$ are estimated in prior training samples. Hence, the portfolio weights of both f_t and $f_t^{\rm MVE}$ are indeed ex ante available.

We construct training and evaluation samples in two standard ways, as illustrated in Figure 6. The first is *fixed training sample*, in which the 32-year full sample (1985-2016) is cut in two 16-year halves for estimation and OOS asset pricing tests, respectively. The second construction is *expanding training samples*, where the model is repeatedly estimated in a sequence of expanding windows, which are expanded by 12 months each time, starting from 1985-2000 and ending in 1985-2015. Each set of training-sample estimates $(\Gamma, \mu_{\rm f}, \Sigma_{\rm ff})$ are used for the OOS construction for the next 12 months after the end of the training-sample. The OOS portfolio returns in the 16 one-year windows are "stitched" together as the final OOS estimate. In either method, the OOS evaluation window is 2001 to 2016.

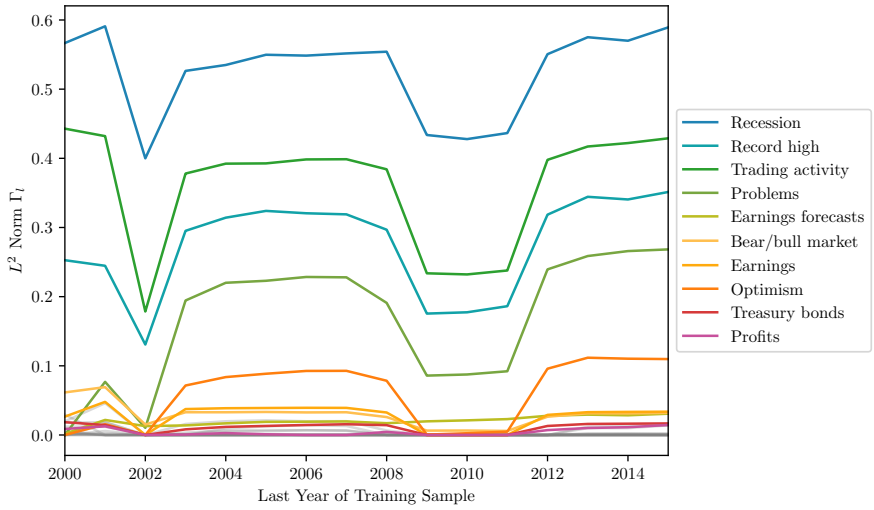
Figure 6: Two ways of OOS Construction



Fixed training sample Expanding training samples

The benefit of expanding training samples is that the parameters are estimated more recently for OOS construction, especially for the later part of the evaluation sample. However, the factor

Figure 7: L^2 Norm of Γ in Expanding Training Samples



Notes: All the expanding windows starts at 1985. The horizontal axis is the last year of the training samples.

estimates need to be "stitched" together to form a long series. We find even with minor changes in the training-sample estimates, the rotation of the K-dimensional f_t is unstable across training samples. As a result, the stitched f_t loses its coherence as a set of pricing factors. Therefore, we construct OOS f_t with the simpler fixed training sample method, such that the whole OOS time series is produced with the same set of training-sample parameters. However, the MVE combination is invariant to the rotation and hence does not suffer from this problem. Therefore, f_t^{MVE} is still constructed with the expanding method for Sharpe ratio evaluation.

Figure 7 plots $\|\Gamma_l\|_2$ across the expanding training samples. It shows the level and ordering of $\|\Gamma_l\|_2$ stay relatively stable over time. The stability of Γ estimates demonstrates the robustness of Sparse IPCA and paves the foundation for constructing OOS factors for asset pricing evaluations.

6.2 Sharpe Ratio Performance

Table 2 reports the Sharpe ratios of the OOS MVE portfolios. We present models estimated under the tuned regularization constant ($\lambda = \lambda^*_{\mathbb{S}}$) and the unregularized ($\lambda = 0$) for comparison. The "avg no. narratives" rows report the number of selected narratives averaged across the expanding training samples. Panel B is for comparisons with characteristics-sorted factors.[41]

[41] An edge for FFC6 is a longer time series. We experiment with different start year for the expanding MVE estimation windows for FFC6 and report the better Sharpe ratio results here.

Table 2: Sharpe ratios of Factor MVE

Panel A: NF MVE (f_t^{MVE})		K					
λ tuning	Statistics	1	2	3	4	5	6
$\lambda = \lambda^*_{\mathbb{S}}$	Sharpe ratio	0.48	1.00	1.10	1.26	1.32	1.31
	avg no. narratives	2.88	4.94	12.12	39.12	43.44	61.81
$\lambda = 0$	Sharpe ratio	0.44	0.66	0.73	0.73	0.78	0.9
	avg no. narratives	all 180 narratives, no selection					
Panel B: FFC6		Mkt	SMB	HML	RMW	CMA	UMD
Sharpe ratio: individual factors		0.51	0.20	0.36	0.52	0.53	0.47
Sharpe ratio: MVE cumulatively		0.31	0.41	0.46	0.81	0.97	0.65

Note: Sharpe ratios are annualized. "NF" stands for narrative-based factors. "MVE cumulatively" means we cumulatively include each of the FFC6 factors to form their MVE portfolio. (The additional results using leave-one-out MVE formation as mentioned in footnote 23 is reported in Appendix B.2.)

The table shows the MVE Sharpe ratios of the narrative-based factors (NF) dominate those of the benchmark models. The highest Sharpe ratio achieved by NF is around 1.3, while the benchmarks are less than 1. As K increases, the Sharpe ratios rise up till $K = 5$ (although the in-sample Sharpe ratio kept increasing marginally at $K = 6$, see Figure 3 Panel 2). The number of narratives selected also increases with K, meaning it takes more narratives to support a factor space with higher dimensionality. This is consistent with the in-sample observation that $\lambda^*_{\mathbb{S}}$ is decreasing in K, as seen in the Figure 3 Panel 2 where the peaks of the Sharpe ratio curves shift to the left as K increases. The lower $\lambda^*_{\mathbb{S}}$ yields more selected narratives.

Figure 8 provides additional evidences that the Sparse IPCA's regularization is effective in improving the estimation of systematic risks. The figure displays the OOS MVE Sharpe ratios under the complete path of λ values, which is similar to the in-sample plot of Figure 3 Panel 2. Compared with the in-sample plot, the OOS Sharpe ratio curves are still hump-shaped with a relatively broad top, peaking at roughly the same range of in-sample $\lambda^*_{\mathbb{S}}$, (although OOS brings noticeably more noises due to sample discrepancies). The steep increase of Sharpe ratio at the lower range of λ clearly shows the benefit of Sparse IPCA's shrinkage and selection effects.

Figure 9 displays the MVE's cumulative return time series over the evaluation sample. The first panel compares the MVE return with regularization versus without regularization, holding fixed $K = 3$. The second panel compares the effects of different K's. (We have standardized the return series such that their sample standard deviation is the same.) The figure shows that the improved

Figure 8: OOS MVE Sharpe Ratio along the Regularizing Constant Path

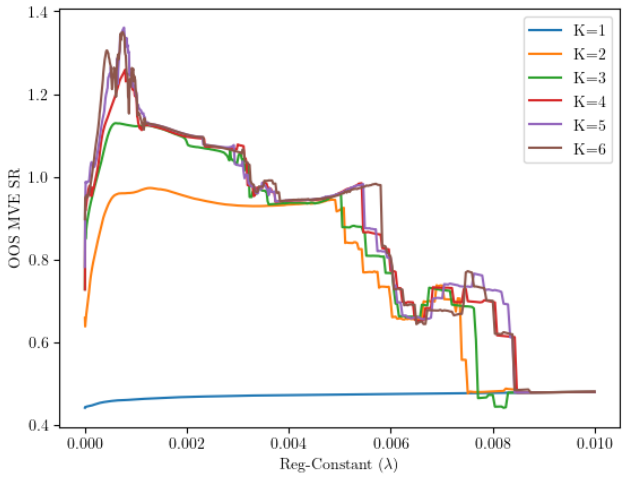
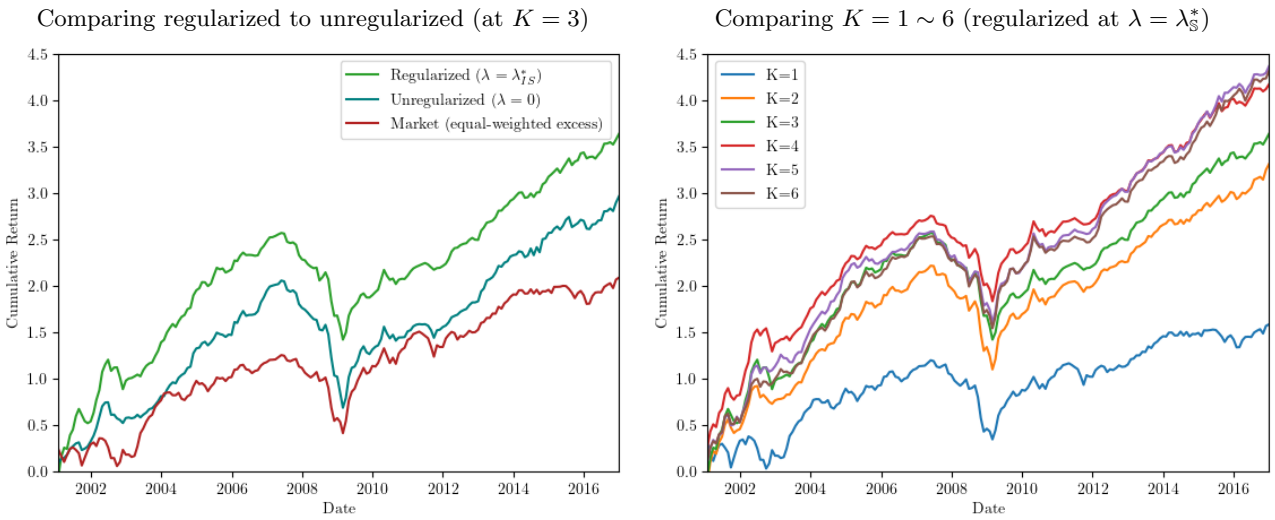


Figure 9: OOS MVE Cumulative Return



Note: We use the market return as the benchmark (shown in the first panel). All the other NF MVE returns in both panels are standardized such that their sample standard deviation is the same as that of the market return.

return performance is not concentrated in a particular period or driven by a particular event (e.g. the 2008 financial crisis).

6.3 Cross-sectional Pricing Performance

Table 3 reports the cross-sectional asset pricing results with respect to three sets of test assets. The empirical procedure is standard in the factor pricing literature, where the tests assets' return

Table 3: Cross-sectional Pricing Results

Factors	avg $\lvert\hat{\alpha}_a\rvert$	avg $\lvert t(\hat{\alpha}_a)\rvert$	$\frac{\#\lvert t(\hat{\alpha}_a)\rvert>1.96}{\#\text{test assets}}$	GRS	$p(GRS)$	R^2
Panel A: 76 anomaly portfolios as test assets						
CAPM	0.35	1.69	0.42	3.95	0.000	59.63
FF3	0.33	1.69	0.36	3.77	0.000	67.06
FF5	0.29	1.63	0.36	3.12	0.000	74.25
FFC6	0.27	1.63	0.38	3.09	0.000	80.58
NF1	0.43	2.11	0.53	3.94	0.000	69.46
NF2	0.24	1.43	0.26	3.50	0.000	77.09
NF3	0.21	1.35	0.27	3.26	0.000	79.94
NF4	0.23	1.46	0.28	3.12	0.000	81.35
NF5	0.23	1.45	0.31	3.09	0.000	81.72
NF6	0.23	1.47	0.29	3.08	0.000	82.48
Panel B: 25 size/book-to-market double sorts as test assets						
CAPM	0.28	1.50	0.44	2.60	0.000	76.82
FF3	0.17	1.40	0.12	2.34	0.001	91.57
FF5	0.11	0.92	0.12	1.76	0.019	92.51
FFC6	0.11	0.94	0.12	1.78	0.018	92.60
NF1	0.20	0.99	0.04	2.67	0.000	76.39
NF2	0.24	1.28	0.24	2.57	0.000	80.91
NF3	0.16	0.90	0.12	2.41	0.001	84.74
NF4	0.18	1.01	0.12	2.18	0.002	84.87
NF5	0.18	1.01	0.12	2.19	0.002	85.09
NF6	0.18	1.05	0.12	2.18	0.002	85.52
Panel C: 25 size/12-month momentum double sorts as test assets						
CAPM	0.35	1.81	0.44	2.56	0.000	73.80
FF3	0.23	1.52	0.28	2.37	0.001	82.75
FF5	0.16	1.14	0.16	1.88	0.011	84.65
FFC6	0.13	1.15	0.16	1.86	0.012	93.47
NF1	0.28	1.26	0.24	2.51	0.000	78.00
NF2	0.22	1.00	0.16	2.19	0.002	81.11
NF3	0.13	0.70	0.08	1.98	0.006	83.10
NF4	0.16	0.81	0.16	1.80	0.016	83.84
NF5	0.16	0.84	0.12	1.80	0.016	83.96
NF6	0.17	0.90	0.16	1.87	0.011	84.42

Note: "NF" stands for narrative-based factors, the number suffix is K. Statistics: (1) avg $\lvert\hat{\alpha}_a\rvert$, cross-sectional average absolute value of the intercepts (a indexes anomaly test assets); (2) avg $\lvert t(\hat{\alpha}_a)\rvert$, the cross-sectional average absolute value of the intercept t-stats; (3) $\frac{\#\lvert t(\hat{\alpha}_a)\rvert>1.96}{\#\text{test assets}}$, the proportion of intercept significantly different from zero; (4) GRS, the GRS statistics for the joint test of all intercepts are zero; (5) $p(GRS)$, the p-value of the said test; and (6) R^2, the R^2 of the time-series regression pooling all test assets.

series are regressed onto each set of proposed factors. The test assets of Panel A are 76 anomaly portfolios constructed as long-short portfolios of 76 characteristics used in Gu et al. (2020).[42] The test assets of Panels B and C are both 25 double-sorted portfolios based on size/book-to-market and size/momentum, respectively, from Kenneth French's data library.

The narrative-based factors (NF) yield small pricing errors at a level no worse than the leading models based on characteristics-sorted portfolios. Take the benchmark model of NF3, its average absolute pricing errors are smaller than the six-factor model of Fama French plus Carhartt, when tested against 76 anomaly portfolios (Panel A) and 25 portfolios of size/momentum double sorts (Panel C). When it comes to size/book-to-market double sorts (Panel B), which Fama French has a natural advantage, NF3 still delivers a comparable performance. The average t-statistics and GRS statistics support these comparisons. Using just three factors, NFs provide a relatively parsimonious description of the systematic risk space.

We observe that increasing dimensionality beyond $K = 3$ does not further reduce the pricing errors (and sometimes makes things slightly worse). This is in contrast to the previous observations with the MVE Sharpe ratio, which keeps increasing up till $K = 5$. One explanation could be that the additional factors are still capturing new dimensions of priced risks that are related to, and revealed by, news narratives. Therefore these dimensions are not reflected in, and hence not "appreciated" by, the test assets, which after all are constructed with the traditional conditioning information of firm characteristics.

A corroborating observation is the total R^2 does not increase much beyond $K = 3$ either, while the total R^2 for characteristic-sorted models keeps increasing with more factors. This further reveals that NFs are not designed to fit the *realized* returns of anomaly portfolios. The goal we started with is to capture the priced risks, and we have approached that from a very different set of conditioning information compared to the traditional characteristics-based methods.

In addition to the main results reported above, we conduct a series of robustness checks for the asset pricing performance under alternative empirical specification and report the results in Appendix B.3. The good performance is largely unchanged when altering the benchmark narrative attention innovation specification: $z_\tau := \theta_\tau - \frac{1}{5}\sum_{\iota=1}^{5}\theta_{\tau-\iota}$. Instead of the five-day trailing moving average, the alternatives include one-day, three-day, and 20-day moving averages. The OOS MVE Sharpe ratios

[42] In detail, the test assets are managed portfolios defined as $r_{a,t} := \sum_i char_{a,i,t} r_{i,t}$, where $char_{a,i,t}$ is the rank standardized characteristics a of stock i at time t.

are still well above one, and the cross-sectional pricing errors are still at a level no higher than that yielded by the characteristics-sorted factors. The 20-day moving average in particular even slightly improves the results across the board (see Table B.2 Panel B). This evidence supports that the z_τ specification is not the driver of our results. Pushing the envelope further, we find downgrading to monthly narrative innovations indeed hurt the asset pricing performance, in which the OOS MVE Sharpe ratio drops to 0.8. When completely abandoning news-based data and using a set of 129 macroeconomic series observed at the monthly frequency as candidate sate variables (z_t), the results further deteriorates. It yields an OOS MVE Sharpe ratio of 0.7, which is still considerable, but falls short of the main results by far (see Table B.2 Panel C). This series of results illustrates the power of Sparse IPCA method in estimating the pricing kernel, and on top of that, the marginal improvements of news narratives and daily covariances, which respectively contributed to the our main results. The last robustness check appends the daily market return to the narrative innovation series (z_τ), given that the wealth portfolio is a specified state variable in ICAPM, which is commonly proxied by the market portfolio. Table B.2 Panel D shows the results are largely unchanged.

Lastly, we note it takes a grain of salt in interpreting statistical horse races between economically motivated models and characteristics-sorted portfolios (Cochrane, 2009, Ch. 7). We do not expect narrative-based factors to dominate characteristics-sorted portfolios in every aspect. In the ideal world with perfect estimation, they should at most perform equally well as there is but one theoretical pricing kernel, no matter whether approached from economic models like the ICAPM or statistical conditions like the APT. Neither do we expect narrative factors to explain all the pricing anomalies. The cross-sectional asset pricing literature has been iterating on new characteristics and signals for decades. Yet, narrative factors are estimated with a completely different data source with no fundamental information in the traditional sense.[43] We claim an empirical success so long as narrative factors' asset pricing performance is *on par* with the benchmarks.

7 Narrative State Variables Forecast "Investment Opportunities"

According to the ICAPM, a state variable enters the consumption process (and hence the pricing kernel) because it forecasts changes in "future investment opportunities." We have argued concep-

[43] We deliberately stay away from quantitative information from firm fundamentals in the estimation step in order to emphasize the method's uniqueness in its news source of conditional information. A model that combines the two sources of information is an interesting direction for future research.

tually that news text should contain such predictive information, and used the argument as the motivation to look for state variables in news data. Now with the state variables estimated, we loop back to examine whether they indeed forecast "future investment opportunities." The theory not only implies the existence of said predictive relationships, but also imposes sign restrictions. A state variable with positive risk premium must increase consumption risk due to positive association with consumption. Hence, the state variable with a positive premium must be a "good news" in the sense of predicting positive changes in the investment outlook, and vice versa.[44]

To examine these theoretical implications, we run predictive regressions of different macro-statistics that represent aspects of "future investment opportunities" onto the estimated state variable. The macro-statistics include equity market return, consumption growth, GDP growth, price and interest rate indexes, etc.[45] For each of these measures, we attempt to predict the cumulative changes in different horizons (h) with the following regressions:

$$\sum_{s=1}^{h} \psi_{t+s} = b_h \left(x_t^{\text{MVE}} / \text{std}(x_t^{\text{MVE}}) \right) + \text{error}_t^{(h)}. \tag{14}$$

On the left-hand side, ψ_t denotes the one-month change in a macro-statistic such as $\psi_t = \text{GDP growth}_t$, and the summation takes the cumulative change in the future horizon of h months. The regressor x_t^{MVE} is the univariate pricing kernel, which is formed as a linear combination of the observed narrative innovations (z_t).[46] We standardize x_t^{MVE} such that the coefficient b_h can be interpreted as the effect per one standard deviation change in the state variable.[47]

Figure 10 reports the full-sample estimation results of the predictive relationships. Each panel corresponds to a different prediction target; the horizontal axis is the prediction horizon h from 1 to 24 months; the vertical axis is the estimated coefficient b_h; the dashed lines mark the 90% confidence interval.[48]

The results show the MVE state variable predicts a wide range of macro-statistics, and the signs of the predictions are consistent with the theoretical restriction. The first panel implies a one standard

[44] This sign restriction is raised in Maio and Santa-Clara (2012) in evaluating other state variable constructions.

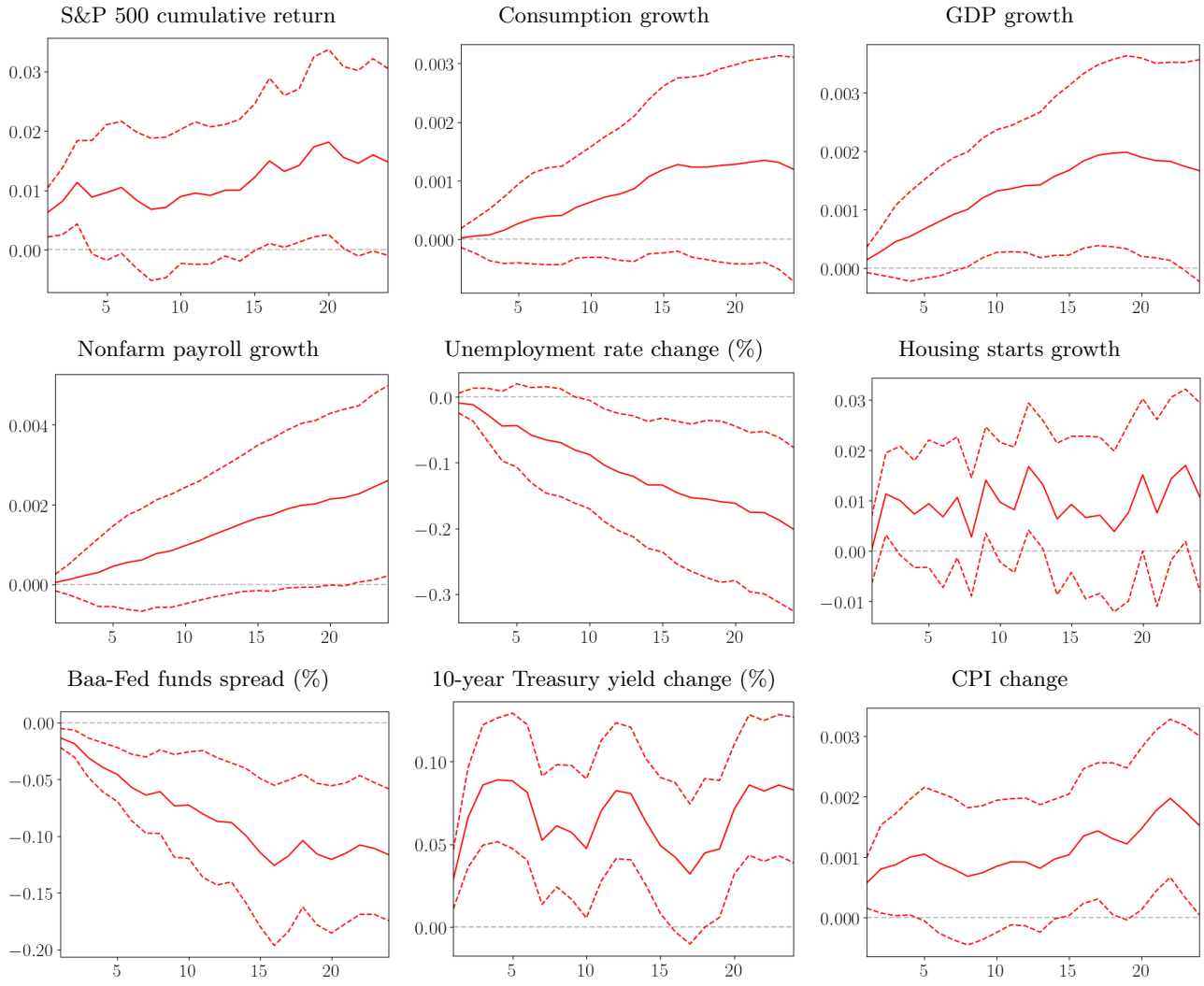
[45] The statistics are from the FRED-MD dataset (McCracken and Ng, 2015)

[46] As defined and discussed in 5.1, $x_t^{\text{MVE}} = I_{z \to \text{MVE}} z_t = \mu_{\text{f}}^\top \Sigma_{\text{ff}}^{-1} x_t^{\text{MVE}}$, is a special state variable as the univariate pricing kernel. We use the full-sample estimate of x_t^{MVE}.

[47] This empirical strategy follows the existing literature that examines the forecasting properties of state variables constructed by other methods (Maio and Santa-Clara, 2012; Liu and Matthies, 2021).

[48] We use the Newey-West standard errors computed with h lags to account for the auto-correlation.

Figure 10: State Variable x_t^{MVE} Predicts Future Investment Opportunities



deviation increase in x_t^{MVE} is associated with a 1.2 percentage points increase in the cumulative market return in the coming 24 months. This is key evidence supporting the ICAPM reasoning for how the narrative-based state variable enters the concurrent consumption process. Meanwhile, the consumption growth predictability (second panel) can be seen as capturing long-run consumption risk (Bansal and Yaron, 2004).[49] These two theoretical motivations for why news enters the pricing kernel are compatible to a large extent. News plays the same role of capturing predictive information that matters for the current consumption choice, although the prediction targets are not exactly the same in the two models. In the fourth and fifth panel, the positive prediction for nonfarm employment and

[49] See Liu and Matthies (2021) for a discussion of news-based risk measures in the context of long-run risks models.

negative prediction for unemployment rate can be attributed to state variables that predict returns in the human capital investment and consumption risks due to labor income shocks.[50] In addition, x_t^{MVE} positively predicts GDP growth, nonfarm employment, and housing construction. And as predicted by theory, it negatively predicts the counter-cyclical "bad news" statistics: unemployment rate, and a measure of credit spread.

In addition to predicting realized changes in the future, we also examine whether the narrative-based state variables are associated with the self-reported forecasts from the Survey of Professional Forecasters (SPF). The results are also consistent with the theory, as reported in Appendix B.4. In summary, this evidence shows that the narrative-based state variables indeed forecast "future investment opportunities" and explains why they should be fundamental risks that explain the cross section of risk premiums.

8 Conclusion

The concept of state variables is theoretically central to multi-factor asset pricing, but empirically elusive to measure. We argue economic and business news is a close real-world counterpart that embodies the concept and thereby provides a novel way to measure the fundamental risks perceived by investors. We demonstrate a way of processing the vast textual data and estimating a parsimonious set of state variables based on new narratives. The resulting pricing kernel is a linear combination of the variations in the attention allocated to some selected news narratives. Therefore, the narrative asset pricing model allows for concrete interpretations of the economic contents of the fundamental risks. As the most prominent examples, we found heightened attention to the "Recession" narrative negatively impacts the pricing kernel. In the opposite direction, the "Record high" and "Optimism" narratives have the largest positive impacts. The underlying news covers a variety of events that are often quantitative in nature. The news-based method effectively relies on the editorial process to locate the most pertinent events at each time, and on the language choice of the articles to reveal the risk contents from the heterogeneous events. We can display the composition of the pricing kernel with words and phrases, and trace the realized market returns to specific news articles. These granular and concrete results provide new insights about the nature of fundamental risks that are

[50] The literature that focus on labor income risks in asset pricing includes Jagannathan and Wang (1996), Julliard (2007), and Liu (2021).

not possible with quantitative datasets of macroeconomic indicators.

The narrative-based systematic risk factors admit top-notch asset pricing performances. The factors can be seen as portfolios formed from textual news data, a source of conditioning information that is completely different from the vast literature that uses fundamental characteristics. We show the narrative-based factors price the anomaly portfolios constructed from firm characteristics as well as (and sometimes better than) the leading factor models constructed from the same class of characteristic information. The factors' MVE portfolio also achieves high Sharpe ratios. The narrative-based state variables also satisfy the ICAPM implied forecasting properties of future investment opportunities.

In terms of econometric methods, the estimation is an upgrade to the Fama-MacBeth two-step procedure to deal with the situation that the state variables are not directly observed, but need to be reduced and selected from many potential series (in our case the news narrative innovation series). Given that the covariances between a stock return and narrative innovations contain instrumental information about the firm's time-varying loadings on the systematic risks, we use IPCA to estimate the instrumental mapping and the factors simultaneously. We devise a Sparse IPCA method for instrument selection. It imposes a group-lasso regularization to filter out the narratives that are irrelevant to systematic risk loadings, which significantly improves the out-of-sample pricing performances. Sparse IPCA is a generalization of lasso-based variable selection to latent factor analysis, which has conceivable potential for many other applications.

From a broader perspective, we view news text as a promising source of data for quantitative model inference. It provides a direct observation of concepts like public information and investor attention, which are central to both rational and behavioral models. Against this backdrop, this paper provides useful tools and demonstrates empirical success of utilizing data from news text in the case of the ICAPM—the central and classical multi-factor asset pricing model.

Appendix

A From *WSJ* Archive to Narrative Innovations (z_t), Details

A.1 Constructing the *WSJ* Document-Term Matrix

We conduct data processing steps in the following order:

1. Remove all articles prior to January 1984 and after June 2017 (data purchased at the beginning of July 2017).
2. Replace all non-alphabetical characters with an empty string and set the remaining characters to lower-case.
3. Parse article text into a white-space-separated word list retaining the article's word ordering. Exclude single-letter words.
4. Exclude articles with page-citation tags corresponding to any sections other than A, B, C, or missing.
5. Exclude articles corresponding to weekends.
6. Exclude articles with subject tags associated with obviously non-economic content such as sports. List of exclusions available from authors on request.
7. Exclude articles with the certain headline patterns (such as those associated with data tables or those corresponding to regular sports, leisure, or books columns). List of exclusions available from authors on request.
8. Concatenate articles with the same accession-number as these are chained articles.
9. Exclude articles with less than 100 words.
10. Remove common "stop" words and URL-based terms. List of exclusions is standard but available from authors on request.
11. Lastly, we conduct light lemmatizing of derivative words. The following rules are applied in the order given, where 'x' is a candidate term. In each case, the stemming is only applied if the multiple terms reduce to the same stem.
    (a) Replace trailing "sses" with "ss"
    (b) Replace trailing "ies" with "y"
    (c) Remove trailing "s"
    (d) Remove trailing "ly"
    (e) Remove trailing "ed." Replace remaining trailing "ed" with "e"
    (f) Replace trailing "ing" with "e". For remaining trailing "ing" that follow a pair of identical consonants, remove "ing" and one consonant. Remove remaining trailing "ing"
    (g) Remove words with less than 3 letters.
12. From the resulting uni-grams, generate the set of bi-grams as all pairs of (ordered) adjacent uni-grams.

13. Exclude terms (both uni-grams and bi-grams) appearing in less than 0.1% of articles. The unique set of terms is the corpus vocabulary. Each column of the DTM corresponds to an element of the vocabulary.

14. Convert an article's word list into a vector of counts for each term in the vocabulary. This vector is the row of the DTM corresponding to the article.

A.2 Topic Model Estimation

We estimate θ and ϕ via the Gibbs sampling procedure proposed by Steyvers and Griffiths (2007). This procedure uses an equivalent form to the DGP given in Equation 2 while introducing an intermediate parameter $y_{m,i}$ corresponding to the topic assignment for each word.

$$\omega_{m,i} \sim \text{Mult}(\phi_{y_{m,i}}, 1), \qquad y_{m,i} \sim \text{Mult}(\theta_m, 1), \tag{15}$$

where $\omega_{m,i}$ is the observed word assignment of the i'th word in article m. The Gibbs sampler generates $\{\theta_m\}, \{\phi_l\}$ as well as the intermediate $\{y_{m,i}\}$ such that,

$$\theta_{m,l} = \frac{\sum_{i=1}^{Len_m} \mathbb{I}(y_{m,i} = l)}{Len_m}, \qquad \phi_{v,l} = \frac{\sum_{m=1}^{M}\sum_{i=1}^{Len_m} \mathbb{I}(\omega_{m,i} = v)\mathbb{I}(y_{m,i} = l)}{\sum_{v=1}^{V}\sum_{m=1}^{M}\sum_{i=1}^{Len_m} \mathbb{I}(\omega_{m,i} = v)\mathbb{I}(y_{m,i} = l)}.^{51,52} \tag{16}$$

Our estimated model contains 180 topics, a specification which is chosen by maximizing the Bayes factor over only the text.

While the topic model is estimated using article level data, we aggregate news attention in the daily level. The daily attention level θ_l ($L \times 1$ vector) is formed as

$$\theta_{\tau,l} = \frac{\sum_{m \in \mathcal{M}_\tau}\sum_{i=1}^{Len_m} \mathbb{I}(y_{m,i} = l)}{\sum_{m \in \mathcal{M}_\tau} Len_m} \tag{17}$$

where $\mathcal{M}_\tau$ is the set of articles published on the next morning of calendar day τ.

A.3 Visualization of the Selected Narratives

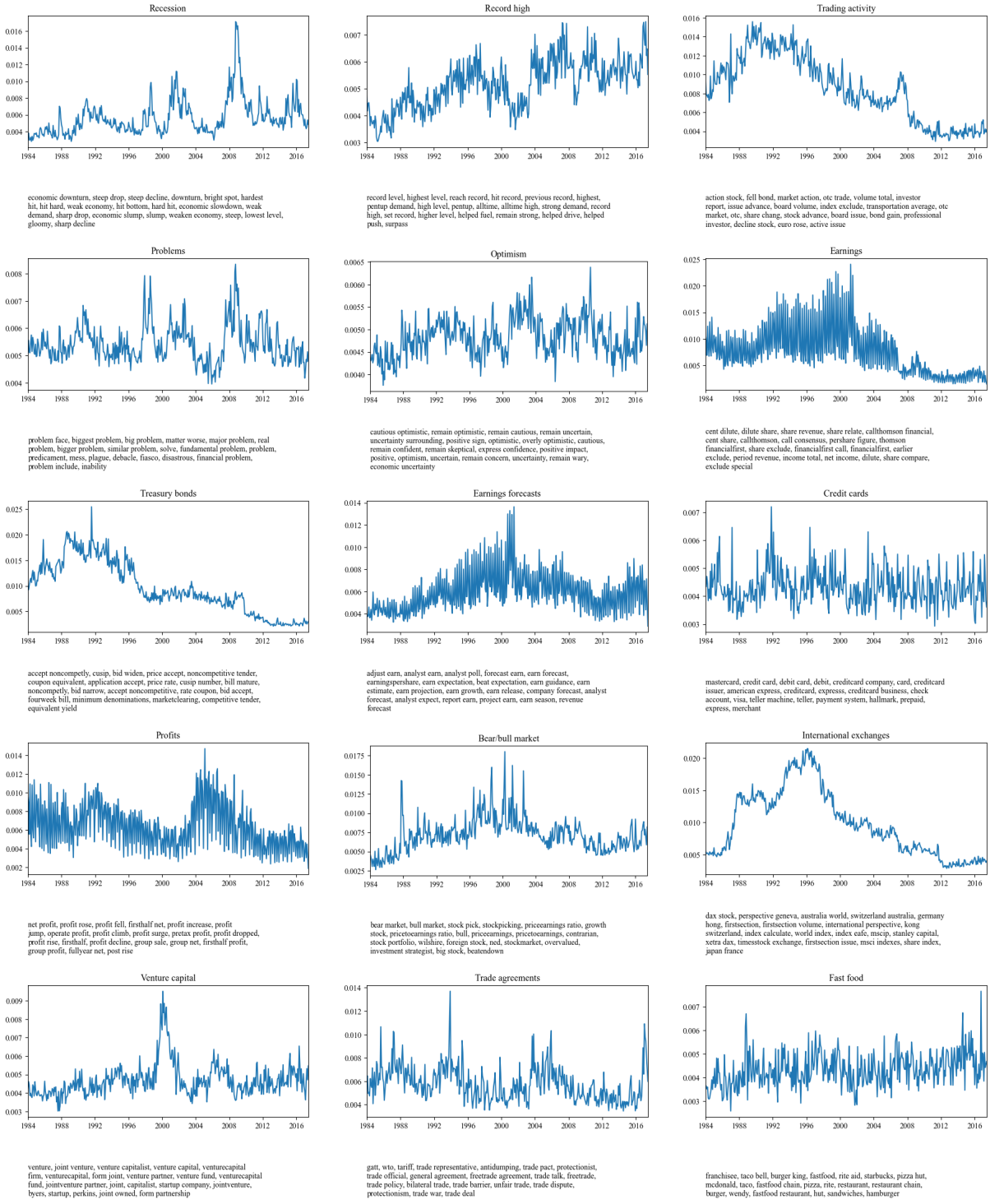
Figure A.1 plots $\theta_{\tau,l}$ time series for the few relevant narratives. The words below each plot lists the ones with large $\phi_{v,l}$ as a display of the content of narrative. The complete visualizations of all the

[52] The denominator in $\phi_{v,l}$'s formula can be equivalently written as $\sum_{m=1}^{M}\sum_{i=1}^{Len_m} \mathbb{I}(y_{m,i} = l)$.

[52] Unlike the standard Gibbs sampling procedure from Steyvers and Griffiths (2007), we do not incorporate the prior terms in our estimates of $\hat{\theta}_{m,l}$ and $\hat{\phi}_{l,v}$. See Bybee et al. (2021) for a fuller discussion of this point.

narratives are in structureofnews.com.

Figure A.1: Selected Narratives



B Result Robustness under Alternative Empirical Procedures

B.1 Instrument Selection Robustness of Sparse IPCA

In order to further justify that Sparse IPCA is effective in distinguishing relevant and irrelevant instruments, we conduct an experiment with simulated placebo instruments. In addition to the 180 observed narratives (z_τ), we randomly generated an equal amount of placebos to "jam" the estimation. In detail, we generate each placebo $z_{l,\tau}$ as an i.i.d. normal sequence that matches the times series variance of a corresponding real $z_{l,\tau}$ sequence. We examine whether the narratives that we know for sure are irrelevant can be successfully filtered out.

Figure B.1 reports the results of this experiment at the benchmark of $K = 3$. The y-axis plots each instrument's $\max_l^\lambda$ (the maximum λ at which l is still included) in log scale. The blue and red bars are for real and placebo instruments, respectively, under the the jammed estimation with $L = 360$. They are sorted according to $\max_l^\lambda$ from left to right. The black dots marks the real instruments' $\max_l^\lambda$ in the original estimation with $L = 180$. The gray horizontal line is the tuned $\lambda_{\mathbb{S}}^*$ under the jammed estimation. Only the bars that are higher than the gray line are eventually selected in the jammed estimation.

The figure that all placebo instruments are correctly excluded. None of the 180 placebos are selected under $\lambda_{\mathbb{S}}^*$. In addition, the real instruments' jammed estimates are very close to the original estimates, at least for the more relevant ones (the ones with high $\max_l^\lambda$). This implies that Sparse IPCA can effectively filter out irrelevant instruments, and the estimates are largely unaffected by the interference of irrelevant ones.

B.2 λ tuning based on leave-one-out (loo) MVE formation

To address the potential concern raised in footnote 23, we provide the results with the λ tuning based on leave-one-out (loo) MVE formation.

Method: For given λ, and the corresponding f_t estimates, take one $t \in \mathbb{S}$ at a time, estimate $\mu_{\text{f}}(t)$ and $\Sigma_{\text{ff}}(t)$ in $\mathbb{S} \setminus \{t\}$, form MVE as $r_t^{\text{loo}} = \mu_{\text{f}}(t)^\top \Sigma_{\text{ff}}(t)^{-1} f_t$, calculate the Sharpe ratio of r_t^{loo} as $\text{SR}^{\text{loo}}(\lambda; \mathbb{S})$. Tune according to $\lambda_{\mathbb{S}}^{*\text{loo}} := \arg\max_\lambda \text{SR}^{\text{loo}}(\lambda; \mathbb{S})$.

Figure B.1: Placebo Test

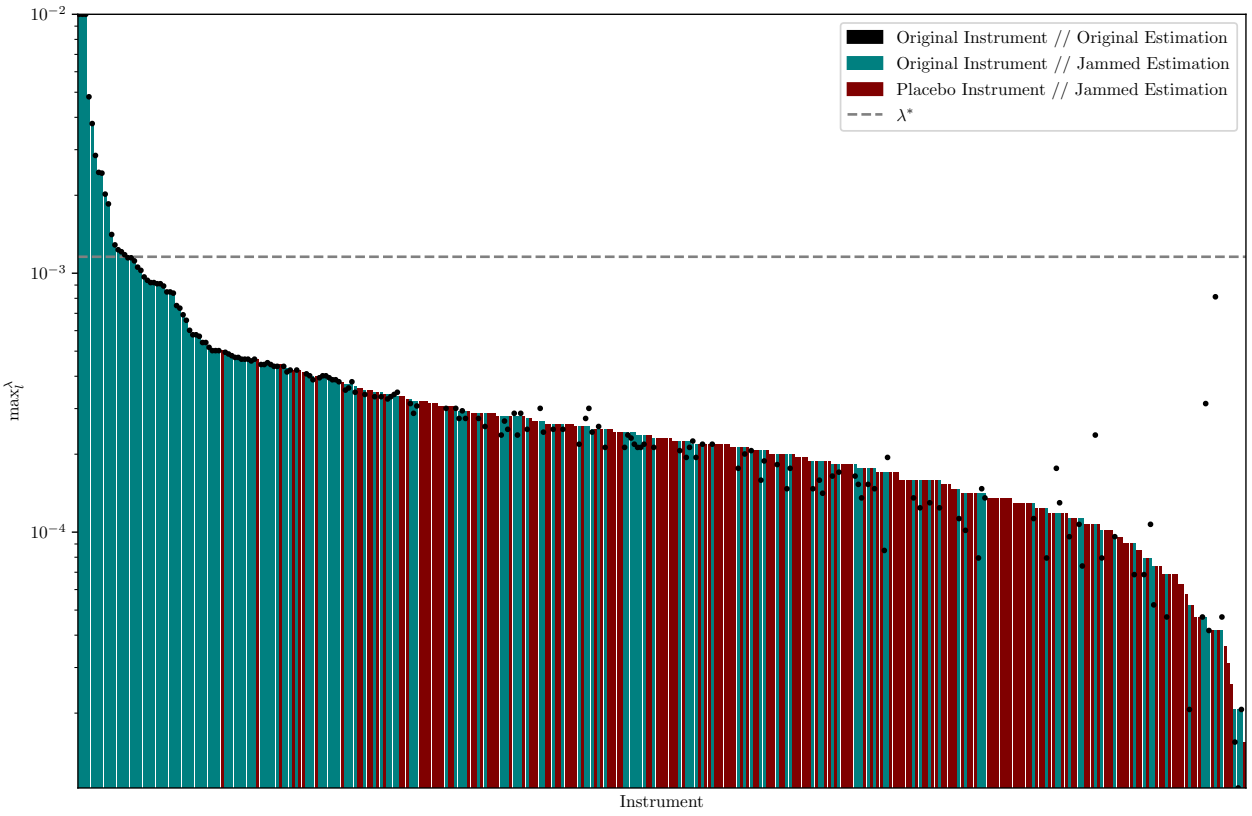


Table B.1: Sharpe ratios of MVE portfolios

		K					
λ selection method	Statistics	1	2	3	4	5	6
$\lambda = \lambda^*_{\mathbb{S}}$	Sharpe ratio	0.48	1.00	1.10	1.26	1.32	1.31
	avg no. instruments	3.88	5.94	13.12	40.12	44.44	62.81
$\lambda = \lambda^{*\text{loo}}_{\mathbb{S}}$	Sharpe ratio	0.48	0.96	1.09	1.26	1.29	1.28
	avg no. instruments	3.88	8.69	16.94	40.12	45.38	61.94

Sharpe ratio results: Table B.1 adds the results with the leave-one-out λ tuning to the results with standard tuning already reported in Table 2 for comparison.

B.3 Asset Pricing Robustness under Different Specifications

Table B.2 repeats the asset pricing tests that are reported in Tables 2 and 3 under different specifications. The specifications are different from the benchmark reported in the main text only in

terms of the instruments ($c_{i,t}$) supplied to the estimation procedure. In calculating $c_{i,t}$, the benchmark specification uses the narrative attention innovation against its trailing 5-day moving average: $z_\tau := \theta_\tau - \frac{1}{5}\sum_{\iota=1}^{5}\theta_{\tau-\iota}$. The items in Panel B change that to "Daily Innovation": $z_\tau := \theta_\tau - \theta_{\tau-1}$; "3-Day Moving-average Innovation" $z_\tau := \theta_\tau - \frac{1}{3}\sum_{\iota=1}^{3}\theta_{\tau-\iota}$; "20-Day Moving-average Innovation" $z_\tau := \theta_\tau - \frac{1}{20}\sum_{\iota=1}^{20}\theta_{\tau-\iota}$, respectively. The fourth specification in Panel B, "Monthly Innovation", works in the monthly frequency instead of daily. It takes the first difference of the monthly attention levels: $z_t := \theta_t - \theta_{t-1}$ (t indexes months). And, the first stage covariance calculation is with monthly stock returns. Panel C completely abandons the news narrative-based approach. It uses a set of 129 macroeconomic series as the z_t inputs in calculating $c_{i,t}$. Panel D appends the daily market return to the 180 narrative innovation series z_τ used in the benchmark configuration.

B.4 State Variables and Self-Reported Forecasts

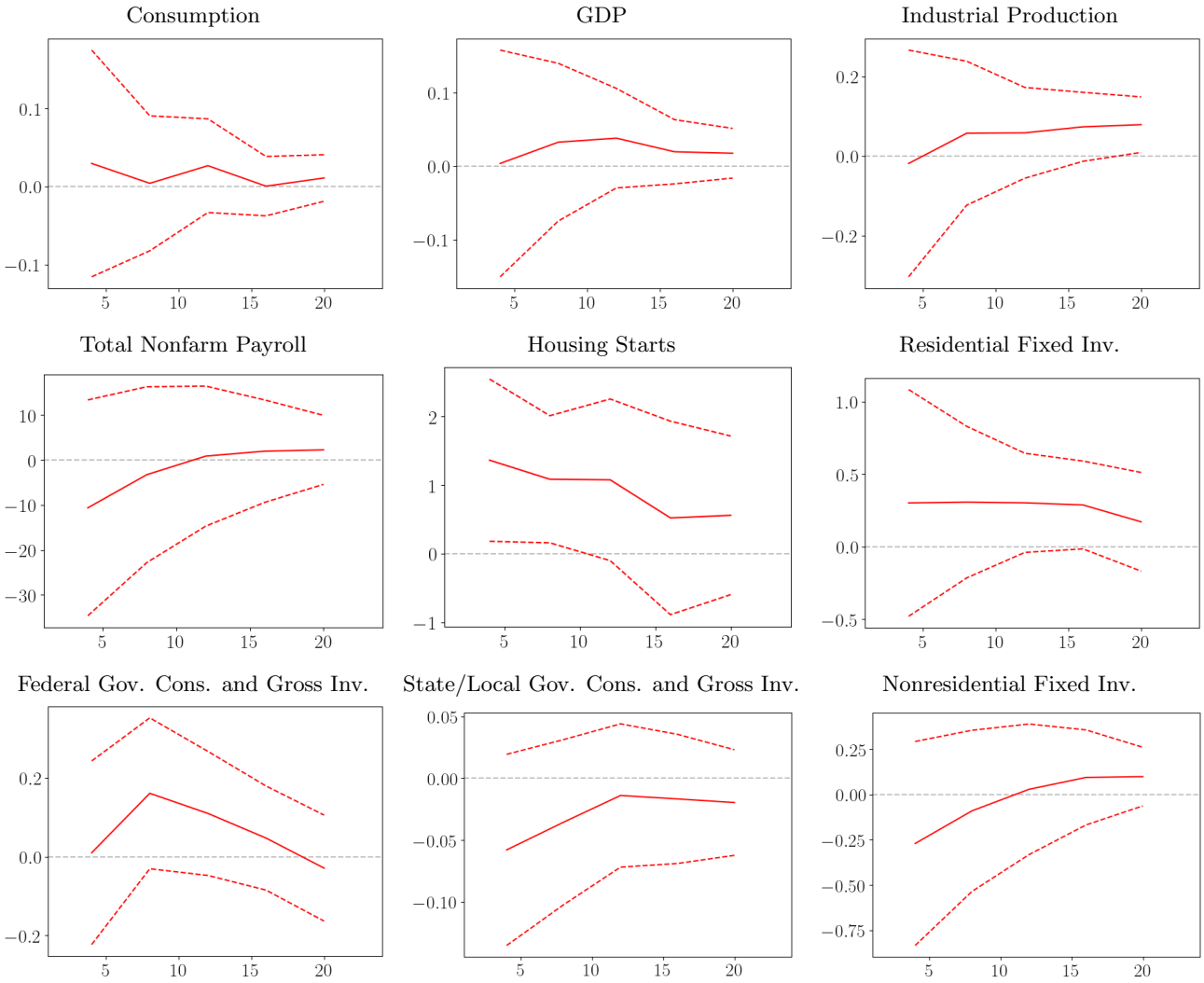
In addition to predicting realized changes in the future, we also examine whether the news-based state variables are associated with the self-reported forecasts from Survey of Professional Forecasters (SPF). We substitute the left-hand-side of regression (14) with Survey of Professional Forecasters (SPF) expectations of some macro-indicators at various future horizons. Figure B.2 shows comparable results with the realized changes. It shows again x_t^{MVE} does a good job predicting the various proxies changes in future investment opportunities.

Table B.2: Asset Pricing Robustness Results

Specification	Statistic	1	2	3	4	5	6
Panel A: Benchmark Specification (repeated from Tables 2 and 3)							
5-Day MA Innov.	OOS Sharpe Ratio	0.48	1.00	1.10	1.26	1.32	1.31
	Avg. Abs. t-Stat. — Anomaly LS	2.11	1.43	1.35	1.46	1.45	1.47
	Avg. Abs. t-Stat. — Size/BM	0.99	1.28	0.90	1.01	1.01	1.05
	Avg. Abs. t-Stat. — Size/Mom	1.26	1.00	0.70	0.81	0.84	0.90
Panel B: Alternative Innovation Specifications							
Daily Innov.	OOS Sharpe Ratio	0.51	1.08	1.05	1.21	1.12	1.11
	Avg. Abs. t-Stat. — Anomaly LS	2.12	1.46	1.38	1.45	1.45	1.46
	Avg. Abs. t-Stat. — Size/BM	1.01	1.51	1.13	0.86	0.87	0.85
	Avg. Abs. t-Stat. — Size/Mom	1.26	1.16	0.88	0.74	0.75	0.74
3-Day MA Innov.	OOS Sharpe Ratio	0.49	0.98	1.08	1.22	1.24	1.19
	Avg. Abs. t-Stat. — Anomaly LS	2.11	1.54	1.38	1.48	1.47	1.47
	Avg. Abs. t-Stat. — Size/BM	0.99	1.32	0.89	0.95	0.94	0.97
	Avg. Abs. t-Stat. — Size/Mom	1.25	1.09	0.73	0.76	0.79	0.81
20-Day MA Innov.	OOS Sharpe Ratio	0.46	0.93	1.00	0.85	1.42	1.44
	Avg. Abs. t-Stat. — Anomaly LS	2.12	1.45	1.35	1.35	1.45	1.43
	Avg. Abs. t-Stat. — Size/BM	0.99	1.23	0.90	0.91	1.02	0.97
	Avg. Abs. t-Stat. — Size/Mom	1.27	0.98	0.74	0.74	0.78	0.80
Monthly Innov.	OOS Sharpe Ratio	0.48	0.72	0.74	0.65	0.77	0.81
	Avg. Abs. t-Stat. — Anomaly LS	1.95	1.69	1.70	1.79	1.73	1.72
	Avg. Abs. t-Stat. — Size/BM	1.01	0.89	1.03	1.22	0.97	0.90
	Avg. Abs. t-Stat. — Size/Mom	1.20	0.96	1.09	1.18	1.04	0.97
Panel C: Macroeconomic Series Instead of News Narrative Innovations							
FRED	OOS Sharpe Ratio	0.46	0.65	0.67	0.63	0.68	0.70
	Avg. Abs. t-Stat. — Anomaly LS	1.96	1.60	1.64	1.66	2.10	1.69
	Avg. Abs. t-Stat. — Size/BM	1.01	0.86	0.89	0.89	0.99	0.89
	Avg. Abs. t-Stat. — Size/Mom	1.23	0.82	0.85	0.89	0.86	0.80
Panel D: Adding Market Return to Narrative Innovations							
LDA+Mkt.	OOS Sharpe Ratio	0.40	1.12	1.14	1.21	1.16	1.21
	Avg. Abs. t-Stat. — Anomaly LS	2.04	1.31	1.41	1.49	1.48	1.50
	Avg. Abs. t-Stat. — Size/BM	1.18	0.71	0.91	0.88	0.89	0.89
	Avg. Abs. t-Stat. — Size/Mom	1.53	0.65	0.72	0.74	0.74	0.80

Note: Panel A repeats the OOS Sharpe ratio and average absolute t-statistic for the benchmark specification from Tables 2 and 3 respectively. Panel B reports the same statistics for a series of different narrative innovation definitions. Panel C reports the same statistics by substituting the set of 129 macro-series from the FRED monthly data series as z_t. For the FRED series, we following the transformations recommended by the dataset documentation (McCracken and Ng, 2015). Panel D appends the daily market return to the 180 narrative innovation series z_τ.

Figure B.2: State Variable x_t^{MVE} Associated with SPF Multi-Horizon Expectations



C News Text Content

This appendix section completes the text bodies of the news article titles that appear in the main text of the paper (Table 1 and Figure 5).

C.1 Text Bodies of the Articles in Table 1

The yellow highlighting follows the same rule as that in Table 1. The shades of each term is related to the term's loading to the corresponding narrative ($\phi_{v,l}$). For each article, we provide the excerpt of the first 160 words to save space.

Date: Headline/Text Body
Recession
1993-05-07: Auto Registrations Continued to Slump In Europe Last Month BONN Newcar registrations continued to plunge across Europe last month as a recession in most markets kept consumers away from auto showrooms. In April car registrations in the European Community were off 18.3 from a year earlier according to provisional figures released by the European Automobile Manufacturers Association an industry lobbying group in Brussels. The sharpest declines came in Denmark where 35 fewer cars were registered than in April 1992 and Spain where registrations fell 30.2. The declines were larger than many analysts had expected and bolster the view that Europes auto industry is facing its leanest year in recent memory. These are very hefty declines and will certainly force a lot of us to reexamine our estimates said Bob Barber auto analyst at James Capel Co. in London. Among Europes five biggest car markets Italy Spain Britain France and Germany only Britain is showing signs of life.
2001-04-25: Consumer Confidence Slides on Fears of Layoffs WASHINGTON Consumer confidence is sliding again after stabilizing in March as jobloss fears threaten to undermine what has been surprisingly resilient consumer spending. The Conference Board said its index of consumer confidence fell to 109.2 in April from 116.9 in March. The index is back to its February level which was the lowest since 1996. Consumers gloomier assessment of their present situation accounted for most of the drop. The presentsituation index fell to 155.6 its lowest level since 1997 compared with 167.5 the previous month. The index of expectations slid to 78.2 from 83.1 but remains above its February low of 70.7. The confidence index based on the responses of several thousand households to a monthly questionnaire has fallen 23 since September mostly because consumers have been more pessimistic about the future as layoff announcements have mounted energy costs have risen and stock prices have fallen.
2009-02-19: U.S. News: Housing Starts Hit Lowest Level In Half-Century Housing starts plunged to new lows in January as a large number of vacant homes tight mortgage financing and a deepening recession created the worst housing market in a halfcentury. Meanwhile a report on industrial production showed that a broad collection of manufacturers cut back in January as falling sales and mounting inventories forced them to reduce worker hours and shut down factories. The two reports echoed what has been an emerging theme in the past few months Makers of goods and homes are slashing production as fast as they can to match falling consumer and business demand. There is nothing in these reports that says we are remotely close to turning around said Nigel Gault an economist at forecasting firm IHS Global Insight. Housing starts fell 16.8 in January from a month earlier to a seasonally adjusted annual rate of 466000 units the lowest in at least 50 years according to the Commerce Department.
2011-08-02: World News: Manufacturing Slowdown Adds to Gloom on Economy

LONDON The U.K. manufacturing sector posted an unexpected contraction in July falling to its lowest level in more than two years while activity at eurozone factories slowed to a nearstandstill. The July data released Monday suggested a poor start to the third quarter and damped hopes for a rebound. The U.K. manufacturing purchasing managers index fell to 49.1 in July from 51.4 in June Markit Economics and the Chartered Institute of Purchasing and Supply said. Markit Economics final eurozone manufacturing purchasing managers index fell to 50.4 in July from 52 in June. A reading below 50 indicates activity is contracting. The last time the sector contracted in the U.K. was June 2009 when Britain was still in recession. Eurozone new orders a forwardlooking indicator of activity fell to a reading of 47.6 the lowest since June 2009.

2016-07-08: World News: U.K. Consumer Sentiment Takes Dive
LONDON British consumer confidence suffered its steepest fall in more than two decades after voters decided to take the U.K. out of the European Union an ominous sign that could foreshadow a broader economic slowdown. A longrunning barometer of consumer confidence published by market researchers GfK U.K. Ltd. recorded an 8point fall in early July the biggest monthly drop since 1994 according to results published on Friday. The index fell to minus 9 from minus 1 in June. The survey of 2002 people conducted June 30 to July 5 is the first gauge of household sentiment published since the June 23 referendum. It suggests some consumers have been shaken by the political and market turmoil sparked by the vote including a steep drop in the pound and may rein in spending as uncertainty persists in the months ahead. The Bank of England is bracing for such a slowdown.

Record high

1989-07-05: Japan Vehicle Sales Rise
TOKYO Sales of cars trucks and buses in Japan climbed 15.5 in June from a year earlier to 508319 units the Japan Automobile Dealers Association said. The total was a record for the month surpassing the previous high of 439966 units set in June last year. The brisk June sales were the latest sign of the strength of the domestic auto market which has seen demand surging in recent months. In May and April for instance sales renewed the record for these months. In March they set an alltime high totaling 683299 units. This is totally unexpected one association official said of the June sales. Everybody here is surprised. We didnt think sales would remain so strong for so long.

1994-07-01: Purchasing Managers In U.K. Survey Report Rise for June Orders
LONDON Britains purchasing managers index rose to a record in June the latest monthly survey from the Chartered Institute of Purchasing Supply shows. The index rose from 59.2 in May to 61.4 in June its highest level ever and the fifth month in a row that purchasing managers have reported an upsurge in manufacturing activity. There was significant growth in manufacturing activity during the month overtaking previous record levels and prices were forced up as suppliers failed to meet the increase in demand. The institute said the June index was boosted by record rises in new orders and employment and a strong surge in output. Order books improved across all U.K. industries and regions in June. Increased demand in the domestic market was led by sales promotions and seasonal factors and was supported by a recovery in exports.

1995-02-27: Hiring Outlook For Second Quarter Appears Vigorous
MILWAUKEE Hiring activity in the second quarter should be at the strongest pace since mid1989 a quarterly joboutlook survey suggests. The survey by Manpower Inc. indicates that 23 more employers will be increasing their work forces during the second quarter than cutting jobs. That net hiring gain would be the largest since the third quarter of 1989 and is 3 higher than the projected net hiring increase in last years second quarter. In the first quarter the net hiring increase was 10. About 15000 U.S. employers were surveyed by Manpower a leading temporaryhelp concern. Mitchell S. Fromstein Manpowers chief executive contended that the strong hiring activity projected for the second quarter merely reflects a continuation of heavy hiring last year. The recent increase in the unemployment rate is more related to an increase in people seeking work than to an economic change in hiring trends Mr. Fromstein maintained. Hiring plans are still on an upward track.

2006-01-12: Wall Street Bonuses Hit a Record in 2005
NEW YORK Wall Streets collective bonuses climbed to a projected record of 21.5 billion last year as firms revenue grew according to the New York state comptrollers office. Comptroller Alan Hevesi said 2005s bonus tally was 2 billion more than the old record which was set in 2000. In 2004 Wall Street bonuses came to an estimated 18.6 billion. Last years average bonus was pegged at 125500 also a record Mr. Hevesi said. Revenue at Wall Street firms rose 45 through the first three quarters of 2005 climbing to the highest level since 2000 the year when the stock market peaked Mr. Hevesis office said. The mergers and acquisitions business accounted for most of the surge.

2016-07-20: U.S. News: Home Building Continues Recovery as Demand Rises

WASHINGTON Home building in the U.S. rebounded in June a sign demand for housing continues to firm heading into the second half of the year. Housing starts rose 4.8 from a month earlier to a seasonally adjusted annual rate of 1.189 million in June the Commerce Department said on Tuesday. Home building continues to gradually recover from the housing bust that accompanied the great recession said PNC chief economist Stuart Hoffman. Demand for new singlefamily homes is slowly but steadily improving. That rising demand has led to concerns about the low inventory of new and existing homes on the market which is pushing up prices and could weigh on further expansion. But Tuesdays report showed an estimated 1.015 million homes under construction in June the highest level since February 2008. Junes uptick was driven by a jump in starts in the West and the Northeast two of the pricier regions in the country.

Trading activity

1993-12-30: Industrials Rise A Bit to Record; Bonds Decline
The Dow Jones Industrial Average crept to a thirdstraight record. Bond prices fell and the dollar rose. The industrial average added a scant 0.56 point to 3794.33. Standard Poors 500 stock index fell 0.36 to 470.58 and the Nasdaq Composite Index rose 3.92 to 768.48. The industrial average climbed in early trading nearly cracking the 3800 level but then spent most of the day in negative territory until just before the close. Investors were greeted early by some positive economic news the Commerce Departments index of leading indicators rose 0.5 in November and existinghome sales jumped a betterthanexpected 2.9. The Dow Jones Transportation Average declined after hitting a record on Tuesday. But Larry Rice chief market strategist at Josephthal Lyon Ross wasnt surprised that the average slipped. The history of this market lately is that you get marginal new highs in the averages and then they back off he said.

1994-10-20: Profit News Helps Boost Stock Prices — Indexes Gain Ground Despite Weakness Of Bonds and Dollar
Stock prices moved higher on the strength of robust earnings shrugging off declining bond prices and a weak dollar. The Dow Jones Industrial Average rose 18.50 to 3936.04 marching closer to its record high of 3978.36. The bluechip indicator which was up more than 30 points late in the session has gained an impressive 160.48 points or 4.2 in the past nine sessions. Other indexes have failed to keep pace with the Dow industrials recent climb but yesterday they also marched forward. The Standard Poors 500 stock index jumped 2.62 to 470.28 the New York Stock Exchange Composite Index gained 1.07 to 258.32 and the Nasdaq Composite Index bolstered by strong technology earnings rose 5.81 to 770.62. Analysts said another round of solid thirdquarter earnings highlighted by AMRs report yesterday helped drive stock prices higher.

1996-06-21: Nasdaq Sinks Amid Sell-Off Of Tech Stocks
Bluechip stocks continued to benefit from an investor exodus from stocks of emergingtechnology companies that again sank the Nasdaq Composite Index. Bond prices held steady and the dollar strengthened. The Dow Jones Industrial Average wavered throughout the day then surged just before the closing bell. The Dow ended up 11.08 points at 5659.43 its second straight gain after falling in six of the previous seven sessions. Broad market indexes were mixed. Standard Poors 500 stock index edged up 0.14 to 662.10. The New York Stock Exchange Composite Index slipped 0.36 to 354.96. But investors spooked by more warnings about disappointing secondquarter earnings rushed to unload smaller stocks. The Nasdaq Composite plunged more than 20 points early in the day before recovering to a drop of 11.93 points to 1167.34. With the end of the second quarter a week away money managers are dumping the highest fliers in their portfolios along with shares of any company announcing bad news.

1997-12-09: Blue Chips Fall As Dollar's Rise Causes Concern
Bluechip stocks broke a sixday winning streak with a decline prompted in part by concerns over the strengthening dollar. Bonds fell. The Dow Jones Industrial Average dropped 38.29 points or 0.47 to 8110.84 its first decline after six sessions that lifted the average 354.35 points. The Dow Jones industrials fell more than the SP 500 which dropped only 1.42 or 0.14 to 982.37. In part that was because Dow industrials component CocaCola dropped 2 12 to 63 916 as some analysts lowered their 1998 estimates for the company citing the negative currency translation impact of the strengthening dollar on foreign sales. Boeing another component of the Dow average dropped 1516 to 51 716. It said the Asian economic crisis could cause some airlines to request a total of 60 delivery delays over the next three years.

1998-04-21: Drug Stocks Resume Gains; Blue Chips Fall

Drug and technology stocks soared financial and economically sensitive stocks swooned and the stock market finished mixed. The dollar also finished mixed and bonds declined. Pulling back from its Friday record the Dow Jones Industrial Average lost 25.66 to close at 9141.84. But Standard Poors 500 stock index and the Nasdaq Composite Index both bettered their Friday records. The SP 500 gained just 0.93 to 1123.65 but the technologystockheavy Nasdaq surged 20.54 or 1.1 to close at 1887.14. After slipping last week on disappointing earnings announcements drug stocks resumed their rise with news that Pfizers Viagra impotence pill is a huge seller and that Eli Lillys Evista may prevent breast cancer. Tech stocks particularly Internetrelated shares have regained momentum following recent favorable earnings news. KTel International which has announced that it will sell compact disks and other recordings over the Internet rose 12 1516 to 41 58 it traded at 6 58 earlier this month.

C.2 News Text of Headlines in Figure 5

We highlight each word according to its impact on the market return ($I_{z\to\text{Mkt}}$), with red for negative and blue for positive impacts. The shades of the highlighting reflects the absolute magnitude.

Date (Market Return, %): Headline/Text Body

1986-09-12 (-4.42): Free Fall: Interest-Rate Worries And Program Trading Send Stocks Plunging — Automated Selling Generates Biggest One-Day Decline As Volume Sets a Record — A Fluke or a Possible Omen?
The stock market showed its explosive new character as never before as prices plunged on huge volume yesterday and the Dow Jones Industrial Average fell a record 86.61 points. The selloff was triggered by bad news on interest rates. But it picked up momentum as waves of computerdriven Wall Street trading strategies increased the pressure. Such wide unpredictable price swings have become almost commonplace this year. At such high levels investors have to get used to the fact that stocks have taken on the trading characteristics of commodities which have long been known for swift wide swings says Leon Cooperman the head of research at Goldman Sachs Co. The avalanche of selling came just a few days after the Dow Jones industrials climbed to a record 1919.71 Sept. 4 and only two months after the industrials previous record drop of 61.87 points on July 7.

1987-10-20 (-17.44): The Crash of '87: Stocks Plunge 508.32 Amid Panicky Selling — A Repeat of '29? Depression in '87 Is Not Expected — Banking System Safeguards And Federal Mechanisms Are Viewed as Adeqaute
Can it happen again On Oct. 28 1929 the stock market fell 12.8 ushering in the Great Depression. While the market plunged 22.6 yesterday economists generally dont expect another depression. I dont think the economy looks like it did in 1929 says George Stigler the winner of the 1982 Nobel Memorial Prize in Economics and a University of Chicago economics professor. The most violent and urgent of factors in the great crash was the collapse of the banking system. That cant happen anymore because of the Federal Deposit Insurance Corp. and additional safeguards. Mr. Stigler like other economists stresses that todays financial system and economic policy mechanisms provide considerably more protection against the type of cascading economic collapse that crippled the nation during the Depression which lasted from 1929 to 1933. During that period the value of the nations output contracted by more than 50 and unemployment rates rose to nearly 25.

1989-10-16 (-5.52): Many Executives Hail Market's Slide As Favorable News
HOT SPRINGS Va. Many of the nations highestranking executives saluted Fridays market plunge as an overdue comeuppance for speculators and takeover players. Assuming that the market doesnt head into a bottomless free fall some executives think Fridays action could prove a harbinger of good news as a sign that the leveraged buyout and takeover frenzy of recent years may be abating. This is a reaction to artificial LBO valuations rather than to any fundamentals said John Young chairman of HewlettPackard Co. whose shares dropped 3.125 to 48.125. If we get rid of a lot of that nonsense it will be a big plus. A few of the executives here for the fall meeting of the Business Council a group that meets to discuss national issues were only too happy to personalize their criticism.

1991-11-18 (-3.55): The Outlook: Double-Dip Recession Possible Not Likely

NEW YORK With recoveries like this who needs a recession Last weeks gloomy news from the drop in retail sales to the jump in joblessinsurance claims to Fridays stockmarket plunge hardly inspires confidence that the economy is indeed recovering and will avoid a doubledip recession. Much will depend of course on what occurs in coming weeks in Washington. For now no one knows if the economys modest thirdquarter rise was merely an uptick in a longrunning slump or the start of a sustained recovery. But the evidence on balance still points to the latter eventuality. Doubledip recessions are rare but they do occur. The economy rose briefly amid the yearlong recession of 196970. And many analysts regard the sixmonth recession of 1980 and the 16month recession of 198182 as really a huge doubledipping slump interrupted by a year of economic growth.

1993-02-17 (-2.71): Stocks Slump as Clinton's Plan Sparks Fears
Frightened by the prospect of higher taxes that could choke off the budding economic recovery investors sent the stock market tumbling in a broadbased selloff. Stock prices plunged from the opening bell as investors gave President Clintons economic plans an initial and strong vote of no confidence. Although stock prices recovered slightly from their lows of the day the Dow Jones Industrial Average ended down 82.94 points or 2.44 to 3309.49. It was the biggest oneday point decline since Nov. 15 1991. Standard Poors 500 stock index fell 10.67 or 2.40 to 433.91. Small stocks fared even worse. The Nasdaq Composite Index which has surged in recent months surrendered 25.15 points or 3.64 to 665.39 its worst decline since Oct. 26 1987. Bond investors stayed cool with the Treasurys benchmark 30year bond losing less than a quarter of a point or less than 2.50 for each 1000 face amount. Shorterterm Treasury issues rose modestly.

1994-02-07 (-2.32): Clinton Plans Jobs Summit To Tackle Global Problem
WASHINGTON AP President Clinton will hold an international jobs summit in Detroit next month to tackle the global problem of persistently high unemployment the White House announced. The March 1415 conference will bring together economic labor finance and industry ministers from the Group of Seven industrialized democracies the U.S. Canada Germany Italy Japan France and Britain. The conference will send a message that we intend to confront the challenge of job creation and unemployment not retreat to the economic structures of yesterday the White House announcement said. During a G7 meeting last July in Tokyo Mr. Clinton announced his intention to convene such a conference. He said then the G7 officials would search for the causes and possible answers for this stubbornly high unemployment. The president originally hoped to hold the meeting last fall but it was pushed into 1994 by the crush of other items on his firstyear agenda.

1997-10-28 (-6.58): Drug Makers High-Tech Stocks Head Roster of Day's Big Losers
NEW YORK In a day that saw the largest trading volume ever on the New York Stock Exchange the 30 stocks in the Dow Jones Industrial Average lost a total of 129 billion in market capitalization. From the Dows peak Aug. 6 when the average closed at 8259 and the market capitalization stood at 1.94 trillion the 30 stocks in the industrial average have surrendered 264 billion. Yesterdays Big Board volume totaled 685496330 breaking the previous record of 683800820 set Jan. 23. The outstanding losers in a session chockablock with big losses came from two groups healthcare stocks including both healthcare providers and drug makers and hightechnology stocks. Among the healthcare stocks the biggest loser was Oxford Health Plans which plunged 42 78 to 25 78 after the company said it would post a thirdquarter loss despite expectations that the company would show a profit for the period.

2007-02-28 (-3.43): The Evening Wrap: From Bad to Worse
The worlds stock markets took a dive today beginning with a sharp plunge in Chinas stock market that pulsed through global trading floors and culminated with one of the worst days for U.S. markets in recent memory. The Dow Jones Industrial Average posted a staggering loss of 416.02 points or 3.3 to end at 12216.24. For anyone caught in the turmoil the close of the session couldnt come soon enough. Blue chips began the session deeply lower and continued to step down until about an hour before the end of the session. Then the industrials plunged in a heartbeat. Weaker by about 280 points the index suddenly was down more than 500 points its deepest intraday swoon since the markets reopened after days of inactivity following the Sept. 11 2001 terrorist attacks.

2011-08-05 (-5.04): Stocks Nose-Dive Amid Global Fears — Weak Outlook Government Debt Worries Drive Dow's Biggest Point Drop Since '08

Stocks spiraled downward Thursday as investors buckled under the strain of the global economic slowdown and the failure of policy makers to stabilize financial markets. The selling began in Europe and continued in the U.S. where stocks plunged from the opening bell. The Dow Jones Industrial Average posted its worst point drop since the financial crisis in December 2008 falling 512.76 points or 4.31 to 11383.68. Oil and other commodities were also hammered. Even gold was a safe haven no more as prices fell. Asian markets slid on Friday morning with benchmark indexes in Tokyo Australia South Korea and Hong Kong all falling more than 3 by midday. It was an absolute bloodbath said John Richards head of strategy at RBS Global Banking Markets. There was no one single catalyst for the downdraft traders said. Rather it reflected multiple concerns that have mounted over the past month and came to a head this week. Worries about a U.S.

2016-06-27 (-3.70): EU Tumult Ripples Through Markets — Europe's battered lenders face new risks from investors and an uncertain economy
Just a few years ago Europes banks managed to stagger out of crisis brought on by the Continents debt woes. Britains looming exit from the European Union analysts and investors fear could push them back in. A wide swath of European financial institutions are at risk Hobbled behemoths like Deutsche Bank AG and Credit Suisse Group AG that are limping through difficult turnarounds clusters of regional banks pressured by negative interest rates and banks across Europes weak periphery that are reeling under piles of bad loans. Still wounded from the eurozone debt crisis European banks need investor confidence and steady economic growth to prosper. Brexit risks both. Perhaps most acutely Britains breakaway calls into question the durability of the European Union and the euro. All of a sudden the prospects of Europes political framework disintegrating at its core considered farfetched just days ago have edged up.

References

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *Q. J. Econ.*, 131(4):1593–1636.

Bali, T. G. and Engle, R. (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics*, 57(4):377–390.

Bansal, R. and Yaron, A. (2004). Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *The Journal of Finance*, 59(4):1481–1509.

Bryzgalova, S. (2015). Spurious factors in linear asset pricing models. *LSE manuscript*, 1(3).

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2021). Business News and Business Cycles. *NBER*.

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

Chen, N.-F., Roll, R., and Ross, S. (1986). Economic Forces and the Stock Market. *Journal of Business*, 59(3):383–403.

Cochrane, J. H. (1996). A Cross-Sectional Test of an Investment-Based Asset Pricing Model. *Journal of Political Economy*, 104(3):572–621.

Cochrane, J. H. (2009). *Asset Pricing: Revised Edition*. Princeton University Press.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging Climate Change News. *Rev. Financ. Stud.*, 33(3):1184–1216.

Fama, E. F. and French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, 51(1):55–84.

Fama, E. F. and French, K. R. (2016). Dissecting Anomalies with a Five-Factor Model. *Rev. Financ. Stud.*, 29(1):69–103.

Fama, E. F. and MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3):607–636.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *Journal of Finance*, 75(3):1327–1370.

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

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *Rev. Financ. Stud.*, 33(5):2223–2273.

Hansen, L. P. and Richard, S. F. (1987). The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica*, 55(3):587.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects. *Q. J. Econ.*, 134(4):2135–2202.

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.

He, Z. and Krishnamurthy, A. (2013). Intermediary Asset Pricing. *American Economic Review*, 103(2):732–770.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, 28(3):650–705.

Jagannathan, R. and Wang, Z. (1996). The Conditional CAPM and the Cross-Section of Expected Returns. *The Journal of Finance*, 51(1):3–53.

Jeon, Y., McCurdy, T. H., and Zhao, X. (2021). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics*.

Julliard, C. (2007). Labor Income Risk and Asset Returns. Technical report.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2020). Predicting Returns with Text Data. SSRN Scholarly Paper ID 3389884, Social Science Research Network, Rochester, NY.

Kelly, B. T., Pruitt, S., and Su, Y. (2017). Instrumented Principal Component Analysis. SSRN Scholarly Paper ID 2983919, Social Science Research Network, Rochester, NY.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2):175–194.

Liu, Y. (2021). Labor-Based Asset Pricing. SSRN Scholarly Paper ID 3081364, Social Science Research Network, Rochester, NY.

Liu, Y. and Matthies, B. (2021). Long Run Risk: Is It There? SSRN Scholarly Paper ID 2592814, Social Science Research Network, Rochester, NY.

Lopez-Lira, A. (2020). Risk Factors That Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns. SSRN Scholarly Paper ID 3313663, Social Science Research Network, Rochester, NY.

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. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4):1187–1230.

Maio, P. and Santa-Clara, P. (2012). Multifactor models and their consistency with the ICAPM. *Journal of Financial Economics*, 106(3):586–613.

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

McCracken, M. W. and Ng, S. (2015). Fred-Md: A Monthly Database for Macroeconomic Research. SSRN Scholarly Paper ID 2646151, Social Science Research Network, Rochester, NY.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica*, 41(5):867–887.

Pelger, M. and Xiong, R. (2020). Interpretable Sparse Proximate Factors for Large Dimensions. SSRN Scholarly Paper ID 3175006, Social Science Research Network, Rochester, NY.

Rossi, A. G. and Timmermann, A. (2015). Modeling Covariance Risk in Merton's ICAPM. *The Review of Financial Studies*, 28(5):1428–1461.

Shiller, R. J. (2017). Narrative Economics. *Am. Econ. Rev.*, 107(4):967–1004.

Steyvers, M. and Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7):424–440.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics*, 68(1):47–73.

Yang, Y. and Zou, H. (2015). A fast unified algorithm for solving group-lasso penalize learning problems. *Stat. Comput.*, 25(6):1129–1141.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zou, H., Hastie, T., and Tibshirani, R. (2012). Sparse Principal Component Analysis. *J. Comput. Graph. Stat.*