Narrative-Managed Portfolio

Fuwei Jiang Bowen Xue Hong Yan Jianhao Zhan

First Draft: August 2022
This Draft: September 2024

Abstract

We propose a novel narrative-managed portfolio utilizing a large news corpus and machine learning models. The narrative-driven forecasts achieve an out-of-sample R^2 of 2.61%. More importantly, the narrative-managed portfolio, constructed from the forecasts, provides mean-variance-skewness investors with annual utility gains of up to 10% and generates significant FF5 alphas. We also demonstrate that narratives contain incremental information beyond existing economic or financial predictors. The predictive power of narratives stems from their ability to capture information related to disaster risk, and investors exhibit sticky beliefs in response to the information contained in narratives.

JEL classification: C53, G11, G12, G17

Keywords: Managed portfolios, Textual Analysis, Machine Learning, Disaster Risk, Sticky Beliefs

Fuwei Jiang, School of Economics, Wang Yanan Institute for Studies in Economics, Xiamen University, jfuwei@gmail.com. Bowen Xue, School of Economics and Management, University of Chinese Academy of Sciences, xuebowen211@mails.ucas.ac.cn. Hong Yan, Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University, hyan@saif.sjtu.edu.cn. Jianhao Zhan, Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University, jhzhan.21@saif.sjtu.edu.cn. Jiang acknowledges support from National Social Science Fund of China (No.22&ZD063), National Natural Science Foundation of China (No.72072193, No.71872195). All errors and omissions are our own responsibility.

1. Introduction

Forecasting stock returns is a fascinating endeavor and attracts both academics and practitioners in finance. Numerous studies focus on forecasting the stock market risk premium and subsequently manage the market portfolio based on the forecasts[1]. These studies concentrate on various economic variables[2], assuming that investors learn from economic fundamentals and form their trading decisions accordingly. As noted in Cochrane (2011), large variations in risk premia imply significant changes for macroeconomics.

In this paper, we propose a new strategy to manage the market portfolio by utilizing a voluminous news corpus and ensemble machine learning models. The media, as the primary source and distribution channel of information in society, consistently produces and disseminates descriptions, interpretations, and predictions of economic conditions, commonly referred to as "news" or "economic narratives" (Goetzmann, Kim, and Shiller, 2022; Tetlock, 2007). In addition to reacting to economic fundamentals, investors also access and incorporate unstructured news information and, accordingly, shape their beliefs and trading decisions (Shiller, 2017; Veldkamp, 2023). Therefore, the information contained in economic and business news should have predictive power and be relevant to the equity risk premium.

In recent decades, some studies have tried to extract information from narratives and connect it to stock market movements[3], yet a systematic study on the return predictability and economic gains from media narratives is still lacking. In this paper, we focus our analysis on the full text of *The Wall Street Journal*, comprising over 880,000 articles from 1984 to 2020. Specifically, we use a mature NLP algorithm, Latent Dirichlet Allocation (LDA),

[1] see, e.g., Welch and Goyal, 2008; Rapach, Strauss, and Zhou, 2010; Rapach and Zhou, 2013; Rapach, Ringgenberg, and Zhou, 2016; Farmer, Schmidt, and Timmermann, 2022; Hollstein and Prokopczuk, 2022

[2] Such as valuation ratios (Campbell and Shiller, 1988a; Fama and French, 1988; Cochrane, 2008), payout yield (Boudoukh, Michaely, Richardson, and Roberts, 2007), short interest rates and the default and term spreads (Keim and Stambaugh, 1986; Campbell, 1987; Breen, Glosten, and Jagannathan, 1989; Fama and French, 1989), inflation rate (Nelson and Kim, 1993; Fama and Schwert, 1977), consumption-wealth ratio (Lettau and Ludvigson, 2001), and stock market volatility (Ghysels, Santa-Clara, and Valkanov, 2005; Guo, 2006), etc.

[3] see, e.g., Hirshleifer, Mai, and Pukthuanthong, 2024; Calomiris and Mamaysky, 2019; Bybee, Kelly, Manela, and Xiu, 2021.

proposed by Blei, Ng, and Jordan (2003), to comprehensively extract information from unstructured narratives. This topic model facilitates the conversion of high-dimensional, unstructured news data into a relatively low-dimensional set of common "news topics". Through this technique, the unstructured news corpus is transformed into interpretable, time-varying series of news topics.

Although topic modeling aids in reducing the dimensionality of the original news corpus, the number of topic predictors remains excessive for traditional linear forecasting regression. To address this issue, we leverage advances in machine learning methods to fully exploit the predictive power of economic narratives on aggregate stock returns. The set of candidate machine learning models considered in this paper includes linear models (Elastic Net, Complete Subset Regression), dimension reduction techniques (Principal Component Analysis, Partial Least Squares), and regression trees (Random Forests, Gradient Boosted Decision Trees). Furthermore, given that forecast combinations can enhance the performance of individual forecasting models (Rapach et al., 2010; Elliott, Gargano, and Timmermann, 2013; Rapach and Zhou, 2020), we also include a combination model that aggregates information from the six individual machine learning models.

Our main empirical findings are as follows. First, we find that narratives have significant predictive power for future market returns. The R^2_{OS} values for forecasts based on narratives and machine learning models range from 1.35% to 2.08%. This result aligns with expectations, as machine learning models incorporate penalization and nonlinearity, making them more flexible than linear regression models. Meanwhile, the R^2_{OS} value for the combination of the forecasts is 2.61%, outperforming any individual machine learning model. In contrast, forecasts based on economic and financial variables yield R^2_{OS} statistics close to zero, significantly lower than those of narrative-based forecasts.

We demonstrate that managing the market portfolio using economic narratives can enhance investors' utility while reducing their exposure to left-tail risk. The narrative-managed portfolio adopts a time-varying position in the market portfolio to achieve two main objec-

tives: first, to align the position with the forecasted market excess returns, and second, to match the portfolio's volatility with that of the market portfolio, ensuring comparability.[4]. Compared to the market portfolio, the narrative-managed portfolio exhibits right-skewed returns, which are appealing to investors (Markowitz, 1952b; Arrow et al., 1974; Mitton and Vorkink, 2007; Boyer and Vorkink, 2014). A typical mean-variance investor can achieve utility gains ranging from 3.59% to 6.89% by holding the narrative-managed portfolio instead of the market portfolio, while an investor with a preference for positive skewness can achieve gains between 7.26% and 10.57%. The narrative-managed portfolio generates positive and statistically significant alphas across various asset pricing models[5]. For an initial investment of $1 in January 2002, the cumulative wealth of the narrative-managed portfolio was twice that of the market portfolio by December 2020.

We demonstrate that the source of the narrative's predictive power is linked to disaster risk. Following Gu, Kelly, and Xiu (2020), we identify important topics based on the decrement in R^2_{OS}. Over half of these important topics are related to disaster risk, including *Labor Market*, *Latin America*, *Middle East*, *Arabia*, and *War*, among others. The presence of these topics is expected, as news of conflicts, such as the Afghan war, can influence public expectations and confidence, thereby affecting stock returns, which is consistent with Hirshleifer et al. (2024) and Hirshleifer, Mai, and Pukthuanthong (2023). We then test this hypothesis and find that as the disaster index rises[6], the future performance of the narrative-managed portfolio improves. Additionally, we find that narratives have predictive power for future disaster risk.

[4]This approach is widely accepted in portfolio management literature (Moreira and Muir, 2017; Cederburg, O'Doherty, Wang, and Yan, 2020; Hollstein and Prokopczuk, 2022).

[5]The analysis is conducted using several prominent asset pricing models: (i) the Capital Asset Pricing Model (CAPM), (ii) the Fama and French (1992) three-factor model (FF3), (iii) the Carhart (1997) four-factor model (Carhart), (iv) the Fama and French (2015) five-factor model (FF5), and (v) the Hou, Mo, Xue, and Zhang (2021) five-factor model (HXZ).

[6]To capture disaster risk, we consider six measures: Disaster Probability, estimated using option prices from Barro and Liao (2021); VIX, the CBOE Volatility Index; EPU, economic policy uncertainty measured by Baker, Bloom, and Davis (2016); Risk Aversion, generated by Bekaert, Engstrom, and Xu (2022); Unemployment, the unemployment rate from FRED; and Recession, NBER-based Recession Indicators from FRED.

Several financial variables are closely associated with time-varying disaster risk. Changes in disaster risk will drive variations in the following variables: the equity market dividend-to-price ratio (Gabaix, 2012; Wachter, 2013), the term spread (Tsai, 2015), corporate credit spreads (Gabaix, 2012; Wachter, 2013; Gourio, 2013), equity market volatility (Wachter, 2013), and the value premium (Gabaix, 2012; Tsai and Wachter, 2016). If narratives capture information related to time-varying disaster risk, they should exhibit significant predictive power for these variables. Empirical results confirm this: all variables show positive R^2_{OS} values, ranging from 13.44% to 71.10%, and are statistically significant at the 1% significance level.

While investors often exhibit sticky beliefs in response to the disaster information[7], our empirical results align with these findings. Following Mankiw and Reis (2002), Coibion and Gorodnichenko (2015), Bouchaud et al. (2019), and Farmer et al. (2022), we measure the degree of stickiness in investors' beliefs. Our results indicate that investors' beliefs are highly sticky, with over 80% of their expectations carried over from the previous period, despite the gradual learning from narratives. We also find a positive correlation between forecast errors and narrative-based expectations, suggesting that investors' expectations adjust more slowly than the market, which underscores their underreaction to new information embedded in narratives. Furthermore, inspired by Sims (1980) and Moreira and Muir (2017), we present an analysis of the relationship between narratives and stock return dynamics using VAR impulse response functions.

This paper contributes to several important strands of literature. Primarily, it adds to the growing body of research on efficient market timing by leveraging the information available to investors. Notable contributions in this area include Moreira and Muir (2017), Cederburg et al. (2020), Hollstein and Prokopczuk (2022), and Huang, Li, Wang, and Zhou (2020), among others. Our research differentiates itself in two key ways. First, we apply market

[7]see, Mankiw and Reis (2002), Coibion and Gorodnichenko (2015), Bouchaud, Krueger, Landier, and Thesmar (2019), Cohen, Malloy, and Nguyen (2020), Gómez-Cram (2022), Farmer et al. (2022), and Hirshleifer and Sheng (2022)

timing strategies using information embedded in economic narratives, an area that remains underexplored in the existing literature. Second, we emphasize the importance of combining forecasts, a concept often overlooked in studies focusing on machine learning models.

This paper contributes to the disaster risk literature by demonstrating that narratives convey disaster risk information. A substantial body of research has explored the asset pricing implications of disasters (Barro, 2006; Rietz, 1988; Wachter, 2013; Baron, Xiong, and Ye, 2022). Recently, economists have begun quantifying disaster risk from economic news, such as Manela and Moreira (2017) who construct the news-implied volatility (NVIX) to measure concerns about rare disasters, and Hirshleifer et al. (2023, 2024) who show that war-related information forecasts stock returns both cross-sectionally and in time-series. This paper provides new evidence that narratives contain disaster risk information, which can be utilized to manage portfolios.

This paper contributes to the literature on sticky beliefs, which is rooted in Mankiw and Reis (2002)'s macroeconomic model of the business cycle, where beliefs often rely on outdated information. Previous studies support this conclusion, including Coibion and Gorodnichenko (2015), who highlight information rigidity in professional forecasts, and Bouchaud et al. (2019), who link sticky beliefs to the profitability anomaly. Similarly, Gómez-Cram (2022) notes that investors are slow to react to recessions, and Farmer et al. (2022) shows that stock return predictability aligns with sticky expectations. Additionally, Cohen et al. (2020) and Hirshleifer and Sheng (2022) demonstrate underreaction to changes in financial reports and earnings announcements, respectively. This paper supports these findings and provides new evidence that investors underreact to disaster-related information in narratives.

The rest of this paper is organized as follows. Section 2 introduces the data and methodology. Section 3 investigates the predictive power of economic narratives and the performance of narrative managed market portfolio. Section 4 explores the potential mechanisms through which narratives predict stock returns. Section 5 presents the results of several robustness tests. Section 6 concludes with a summary of our findings.

2. Data and Methodology

2.1 *The Wall Street Journal* Data Set

Our original narrative data set consists of all full articles published in *The Wall Street Journal* from January 1984 to December 2020. As one of the largest newspapers, *The Wall Street Journal* faithfully records all major events happening around the world, which may have a significant impact on the stock market. Therefore, it is one of the most used news corpus in the literature, studying the relationship with narrative and stock market (e.g., Tetlock, 2007; Baker et al., 2016; Manela and Moreira, 2017; Bybee et al., 2021). Starting from 1984, which is the earliest available history from Dow Jones's "Factiva" database, the sample covering the history of Dot-Com bubble, The Great Recession, and the COVID-19 Recession.

Table 1: **Summary Statistics of News Data.** Numbers related to the text are rounded to integers. After processing, our corpus consists of 879,820 articles from January 1, 1984 to December 31, 2020.

Variables	Mean	Std Dev.	Min	Max
Number of articles per month	1965	522	932	3012
Number of articles per year	23584	5943	12345	33123
Number of terms per month	521200	136467	209394	837250
Number of uni-gram per month	451878	114958	182933	714343
Number of bi-grams per month	69321	23526	26461	122907
Number of unique terms per month	20752	1162	17399	22450
R_t (%)	0.763	4.454	-23.24	13.65

As is common in Natural Language Processing (NLP), we take a step of text processing to clean our news corpus:

- *Article screening*: exclude articles in Section "Sports", "Leisure & Arts", "Books", "Correction", "Weekend Journal", "Personal Journal"; exclude articles with page-

citation tags corresponding to any sections other than A, B, C, or missing; exclude articles corresponding to weekends; exclude articles less than 100 words.

- *Normalization and lemmatization*: all non-alphabetical characters are removed; every word is in lowercase and transformed to their noninflected form; remove common stop words.
- *Tokenization and selection*: Convert the texts into uni-grams and bi-grams; exclude both uni-grams and bi-grams appearing in less than 0.1% of articles.
- *Document-Term Matrix construction*: based on the remaining terms (uni-grams and bi-grams) and its count numbers, construct the Document-Term Matrix (DTM). Each column of DTM corresponds to one term of the vocabulary and each row of DTM corresponds to one article.

Finally, our cleaned news corpus consists of 879,820 articles, covering over 230 million terms. Table 1 provides summary statistics of news data. On average, we deal with 1965 articles and 521,200 terms per month, which 87% of terms are uni-grams and 13% are bi-grams.

2.2 Topic Model

To further reduce the dimensionality of the Document-Term Matrix, we use a popular topic model called Latent Dirichlet Allocation (LDA) proposed by Blei et al. (2003). As a generative and unsupervised method, the LDA assumes each document d has its own distribution over topics given by $\boldsymbol{\theta_d}$, and each topic k has its own distribution over terms given by $\boldsymbol{\phi_k}$. Therefore, the $D \times V$ Document-Term Matrix, which is highly dimensional and sparse, is converted into a $D \times K$ Document-Topic Matrix. The hyperparameter K represents the number of topics, and following Bybee et al. (2021) and Hong, Jiang, Meng, and Xue (2022), we set $K = 180$.

LDA has two key output, document-topic distribution $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_D]'$ and topic-term distribution $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_K]'$. To construct our market portfolio, we care more about the

topic distribution at time series. Therefore, all articles published in time τ (e.g., a given day, week or month) are aggregated as one article and then use the trained LDA model to infer topic distribution at this point. In this way, we get 180 topic proportion time-series.



Fig. 1. **Word Distribution of Topics**. The figure shows the word clouds of topic *Arabia*, *Financial Oversight*, *Health*, *Labor Market*, *Latin America*, *National Security*, *Party*, *Price* and *War*. Each word cloud only contains a subset of all the most important words, and the font size reflects the weight of each word.

To investigate the interpretability of LDA results, we select nine topics and present their distributions and time-series attention.[8] Figure 1 and Figure 2 shows the word clouds for these topics and their time-series proportions, separately. We find that the words are well-clustered into interpretable topics. For example, the keywords for the topic *Arabia* include many arabia-related words, such as {*saudi, iraq, iran, islamic, ...*}. As shown in Figure 2, this topic received the most attention during Gulf War and Iraq War. We also report another political-related topic *War*, which keys words include {*military, defense,*

[8]In Appendix A, we provide the top ten key terms for each estimated topic.

force, nuclear, official, ...}, news attention rises to over 2% during US-involved war, like Gulf War, Afghanistan War and Iraq War. Moreover, even for the US not directly involved war like Syrian Civil War happened since March 2011, *War* attention rises to a local peak. Recent literature has found evidence that economic "dark matter", such as geopolitical conflicts, can explain the movement of future stock returns (Hirshleifer et al., 2023; Caldara and Iacoviello, 2022; Barro and Liao, 2021). Naturally, these geopolitical topics may also be useful for portfolio management.

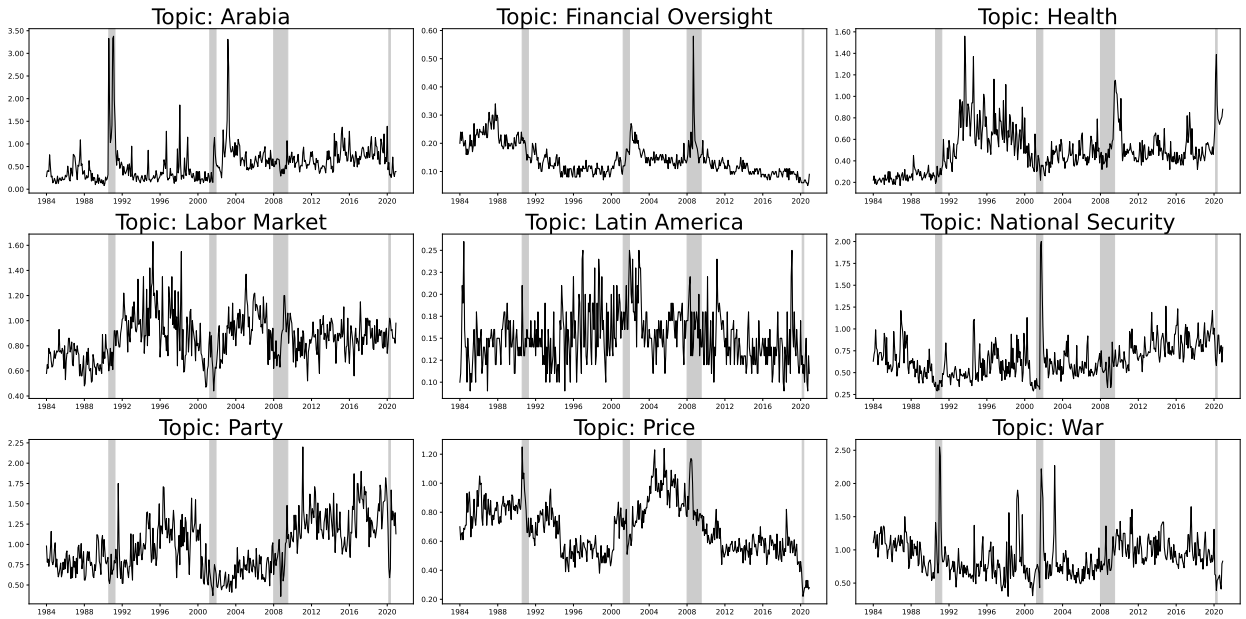


Fig. 2. **Time-series of Topics**. The figure plots the monthly topic proportion for the topic *Arabia*, *Financial Oversight*, *Health*, *Labor Market*, *Latin America*, *National Security*, *Party*, *Price* and *War*. The topic proportions are calculated by aggregating all corpus data for each month into one article and inferred by trained LDA model.

Additionally, we also provide some economic related-topic such as *Labor Maret*, whose key terms include {*benefit, pay, retirement, pension, compensation, employer, ...*}. From the perspective of time series, the highest attention appears in the late 1990s. In fact, the final decade of the 20th century is the bountiful period for pension funds due to the bull capital market. Furthermore, this topic attention tends to rise from a relatively lower trough at the beginning of recession, as labor markets are heavily affected during these periods and become a focal point of discussion. This relationship may help track the development of

economic recessions and guide portfolio management. Overall, these findings suggest that our narrative topics effectively capture various real-world big events.

2.3 Methodology

We estimate the narrative-based forecasts using the following equation:

$$R_{t+1} = G(\boldsymbol{Narr_t}) + \varepsilon_{t+1} \tag{1}$$

where R_{t+1} represents the excess return on a stock market index relative to the risk-free interest rate; $\boldsymbol{Narr_t}$ denotes a 180-vector of narrative topic distribution at time t; and $G(\cdot)$ signifies the mapping from the narrative to the future market excess return. We estimate the mapping relation $G(\cdot)$ utilizing eight models, which include benchmark model kitchen sink (KS) and six machine learning models: elastic net (ENET), complete subset regression (CSR), principal component analysis (PCA), partial least square (PLS), random forest (RF), gradient boosted decision trees (GBDT) and the elastic net combination (COMB) of Rapach and Zhou (2020).

KS. Kitchen Sink model incorporates all variables into a multiple predictive regression framework. The model's $\boldsymbol{\beta}$ can be estimated by:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{t=1}^{T} \left(R_{t+1} - \boldsymbol{\beta}' \boldsymbol{Narr_t} \right)^2 \right) \tag{2}$$

ENET. Elastic net is a penalized linear model introduced by Zou and Hastie (2005), which incorporates both ι_1 and ι_2 penalty functions into the objective function. The model's $\boldsymbol{\beta}$ can be estimated by:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{t=1}^{T} \left(R_{t+1} - \boldsymbol{\beta}' \boldsymbol{Narr_t} \right)^2 + \alpha\lambda \sum_{k=1}^{K} \beta_k^2 + (1-\alpha)\lambda \sum_{k=1}^{K} |\beta_k| \right) \tag{3}$$

CSR. Complete subset regression, proposed by Elliott et al. (2013), combines forecasts

from all feasible predictive regression models while maintaining a fixed number of predictors. CSR can be perceived as a linear model with shrinkage, striking a superior balance between bias and variance.

PCA. PCA is a prevalent factor model employed in macroeconomic and financial forecasting (Stock and Watson, 2002; Baker and Wurgler, 2006; Gu et al., 2020). It posits that $\boldsymbol{Narr}_t$ is influenced by several underlying latent factors, denoted as $\mathbf{f}_t$.

$$\boldsymbol{Narr}_t = \Lambda \mathbf{f}_t + e_t \tag{4}$$

where $\mathbf{f}_t$ represents a k-vector of principal components derived from $\boldsymbol{Narr}_t$, with k being substantially smaller than the dimension of $\boldsymbol{Narr}_t$. The estimation of $\mathbf{f}_t$ can be achieved by maximizing the common variation among the predictors. Subsequently, excess market returns may be predicted as follows:

$$R_{t+1} = \gamma' \mathbf{f}_t + \mu_{t+1} \tag{5}$$

PLS. Contrary to PCA, which maximizes the common variation among predictors while disregarding information in the dependent variable, Partial Least Squares, introduced by Wold (1966), seeks to maximize the covariance between predictors and the dependent variable. PLS can be executed using ordinary least squares (OLS) regressions (Kelly and Pruitt, 2015). The initial step entails estimating the univariate covariance, φ_k, between R_{t+1} and each predictor $Narr_{k,t}$ via time series analysis. Subsequently, a cross-sectional regression of $Narr_{k,t}$ on the corresponding loading φ_k is conducted. The dependent variable and all predictors are then orthogonalized concerning previously established components, and the procedure is reiterated on the orthogonalized dataset until the desired number of PLS components is obtained (Gentzkow, Kelly, and Taddy, 2019).

RF. Random Forest, introduced by Breiman (2001), builds upon the regression tree technique. A single regression tree, as a weak learner, demonstrates inadequate out-of-

sample performance. To improve forecast precision, RF constructs B regression trees, with each tree being informed by R_{t+1} and L features, randomly chosen from $\boldsymbol{Narr}_t$. The ultimate mapping from narrative to future market excess return is obtained by averaging the results of each tree,

$$\hat{G}\left(\boldsymbol{Narr}_t; B, L, N\right) = \frac{1}{B}\sum_{b=1}^{B} \hat{g}_b\left(\boldsymbol{Narr}_t; L, N\right) \tag{6}$$

where L represents the depth of the trees and N signifies the number of features selected at random.

GBDT. Gradient Boosted Decision Trees is a machine learning technique initially introduced by Schapire (1990) and Freund (1995) for classification tasks. Subsequently, Friedman, Hastie, and Tibshirani (2000) and Friedman (2001) expanded its application to regression trees. Contrasting with Random Forest, which constructs trees in parallel, GBDT develops trees sequentially. During each iteration, a regression tree is built based on the residuals of the model consisting of $b-1$ trees. The newly formed tree's prediction is then incorporated into the overall prediction, employing a learning rate ν. This procedure is reiterated until the ensemble comprises B trees. Ultimately, the mapping from narrative to future market excess return is represented by the weighted sum of each tree.

$$\hat{G}\left(\boldsymbol{Narr}_t; B, v, L, N\right) = \sum_{b=1}^{B} \nu\hat{g}_b\left(\boldsymbol{Narr}_t; L, N\right) \tag{7}$$

where L denotes the depth of the trees, N signifies the number of features to be randomly chosen, and v represents the learning rate.

COMB. The combination method, proposed by Rapach and Zhou (2020), uses the elastic net method to select a subset of models, assigning equal weights to the chosen models and setting the weights of the remaining models to zero,

$$\hat{R}_{t+1}^{Narr} = \frac{\sum_{n=1}^{6} \hat{R}_{t+1,n}^{Narr} \cdot \mathbb{1}_{\{\beta_i \neq 0\}}}{\sum_{n=1}^{6} \mathbb{1}_{\{\beta_i \neq 0\}}} \tag{8}$$

where $\boldsymbol{\beta} = [\beta_1, ..., \beta_6]$ is calculated by,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\sum_{t=1}^{T} (R_{t+1} - \boldsymbol{\beta}' \hat{\boldsymbol{R}}_{\boldsymbol{t+1}}^{\boldsymbol{Narr}})^2 + \alpha\lambda \sum_{n=1}^{6} \beta_n^2 + (1-\alpha)\lambda \sum_{n=1}^{6} |\beta_n|) \tag{9}$$

where $\hat{\boldsymbol{R}}_{\boldsymbol{t+1}}^{\boldsymbol{Narr}} = [\hat{R}_{t+1,1}^{Narr}, ..., \hat{R}_{t+1,6}^{Narr}]$ is generated by aforementioned six models.

Among the six machine learning methods examined, several hyperparameters warrant consideration. For the ENET method, the penalty parameter λ is the sole hyperparameter, while α remains fixed at 0.5. In the CSR method, the hyperparameter corresponds to the number of predictors within each predictive regression. For factor models, such as PCA and PLS, the hyperparameter is the number of factors. In the case of regression tree-based models (RF and GBDT), N is set as the logarithm of the feature count, and the random seed is fixed at 0. For the RF model, tree depth L constitutes the hyperparameter, and the number of trees is held constant at 1000. The GBDT model's hyperparameters include tree depth L and the number of regression trees B, with a learning rate v of 0.01. Optimal hyperparameters are determined using 10-fold cross-validation (CV). We report the set of hyperparameters for each machine learning model in Appendix B.

3. Empirical Results

3.1 Out-of-Sample Results

We first investigate the out-of-sample forecasting performance of narratives. We use the period from January 1984 to December 1998 as the initial estimation period for all individual models, and January 1999 to December 2000 for machine learning combination model (COMB), then the out-of-sample period spanning from January 2001 to December 2020. The forecasts are assessed using the widely adopted Campbell and Thompson (2008) R_{OS}^2 statistic and the MSFE-adjusted statistic of Clark and West (2007). The R_{OS}^2 statistic quantifies the proportional reduction in mean squared forecast error (MSFE) for the predictive

regression forecast compared to the historical average benchmark,

$$R^2_{\text{OS}} = 1 - \frac{\sum_{t=p}^{T-1}\left(R_{t+1} - \hat{R}^{Narr}_{t+1}\right)^2}{\sum_{t=p}^{T-1}\left(R_{t+1} - \bar{R}_{t+1}\right)^2}, \tag{10}$$

where R_{t+1} represents the excess return on the stock market index relative to the risk-free interest rate, $\hat{R}^{Narr}_{t+1}$ represents out-of-sample forecasts of stock excess returns by narratives, and $\bar{R}_{t+1}$ represents the historical average benchmark and is calculated by,

$$\bar{R}_{t+1} = \frac{1}{t}\sum_{s=1}^{t} R_s. \tag{11}$$

The R^2_{OS} statistic ranges from $(-\infty, 100\%]$. A positive R^2_{OS} value indicates that the forecast $\hat{R}^{Narr}_{t+1}$ outperforms the historical average $\bar{R}_{t+1}$ in terms of MSFE. We are also interested in the out-of-sample performance during business cycles. Following Rapach et al. (2010), we compute the R^2_{OS} statistics distinctly for economic expansions ($R^2_{\text{OS,up}}$) and recessions ($R^2_{\text{OS,down}}$),

$$R^2_{OS,i} = 1 - \frac{\sum_{t=p}^{T-1} I^i_t\left(R_{t+1} - \hat{R}^{Narr}_{t+1}\right)^2}{\sum_{t=p}^{T-1} I^i_t\left(R_{t+1} - \bar{R}_{t+1}\right)^2} \quad i = up, down, \tag{12}$$

where I^{up}_t $\left(I^{\text{down}}_t\right)$ is a dummy variable that equals one when month t is in an NBER expansion (recession) period and zero otherwise.

The second statistic is the Clark and West (2007) MSFE-adjusted statistic (henceforth referred to as the CW-test). This test examines the null hypothesis that the historical average MSFE is less than or equal to the predictive regression forecast MSFE, against the one-sided (upper-tail) alternative hypothesis that the historical average MSFE is greater than the predictive regression forecast MSFE, corresponding to $H_0 : R^2_{\text{OS}} \leq 0$ against $H_A : R^2_{OS} > 0$. Intuitively, under the null hypothesis asserting that the excess market return originates from a constant expected return model, the historical average benchmark is expected to produce a more accurate forecast than any machine learning model. This is attributed to

Table 2: **Out-of-Sample Forecasts.** Panel A reports out-of-sample forecasting performance of 12 variables used by Welch and Goyal (2008), along with the time-series momentum variable. Panel B reports out-of-sample forecasting performance of the narrative topics. The forecast models includes kitchen sink (KS), elastic net (ENET), complete subset regression (CSR), principal component analysis (PCA), partial least square (PLS), random forest(RF), gradient-boosted decision trees (GBDT) and forecast combination (COMB). R^2_{OS} is the Campbell and Thompson (2008) out-of-sample R^2 statistic. $R^2_{OS,up}$ ($R^2_{OS,down}$) statistics are calculated over NBER business-cycle expansions (recessions). CW-t is the Clark and West (2007) statistic. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively. Out-of-sample period is from January 2001 to December 2020.

Model	$R^2_{OS}(\%)$	CW-t	$R^2_{OS,down}(\%)$	$R^2_{OS,up}(\%)$
Panel A : Economic predictors				
KS	-5.78	0.15	-15.46	-0.79
ENET	0.19	1.22	-0.07	0.33
CSR	0.58	1.13	-0.74	1.26
PCA	0.21	0.96	1.49	-0.45
PLS	-0.07	0.64	-6.80	3.39
RF	-3.04	-1.66	-9.65	0.36
GBDT	-0.13	-0.36	-1.28	0.46
COMB	0.02	0.38	-2.53	1.33
Panel B : Narrative topics				
KS	-217.89	1.21	-177.43	-238.73
ENET	1.35**	1.93	3.41	0.30
CSR	1.78**	1.95	2.17	1.59
PCA	1.52**	1.77	2.08	1.22
PLS	1.57**	1.91	4.15	0.24
RF	2.08*	1.56	4.27	0.96
GBDT	1.67*	1.45	3.50	0.73
COMB	2.61**	2.02	4.26	1.75

the benchmark model's unbiased nature and minimal variance, stemming from its exclusion of noisy predictors. As a result, the benchmark model's MSFE is anticipated to be lower than that of the machine learning model under the null hypothesis. The MSFE-adjusted statistic takes into account the negative expected difference between the historical average MSFE and the predictive regression MSFE under the null, allowing for the rejection of the null hypothesis even if the R^2_{OS} statistic displays a negative value.

For comparison, we first examine the out-of-sample forecasting performance of economic predictors, which include 12 variables used by Welch and Goyal (2008), along with the

time-series momentum variable. Panel A of Table 2 presents the results. We find that the kitchen sink model produces a negative R^2_{OS} , indicating that it underperforms the historical average benchmark in out-of-sample settings. While machine learning methods improve the performance of economic predictors, their results remain statistically insignificant. For example, the combination method produces a positive R^2_{OS} of 0.02% with a CW-t statistic of 0.38. These findings are consistent with Welch and Goyal (2008), which shows that economic variables typically underperform the historical average in out-of-sample forecasts.

Panel B of Table 2 reports the forecasting performance of our narrative topics. Similarly, the Kitchen Sink model shows subpar performance, as indicated by a negative R^2_{OS} statistic of -217.89%, demonstrating that narrative-based forecasts using the KS model still fail to outperform the benchmark. However, narrative-based forecasts generated by machine learning models achieve positive R^2_{OS} statistics around 1.5%. In fact, Campbell and Thompson (2008) demonstrate that, given the substantial unpredictable component embedded in monthly market returns, a monthly out-of-sample R^2_{OS} statistic of 0.5% can yield significant economic value. Additionally, all machine learning models generate MSFEs significantly lower than the benchmark at the 10% significance level or better. Moreover, the machine learning combination method performs particularly well, generating an R^2_{OS} statistic exceeding 2.5%. Finally, unlike economic predictors, which perform better during expansion periods, narrative topics derive their predictive power primarily from recession periods.

To further evaluate the degree to which various models utilize information embedded in narrative topics, we employ a forecast encompassing test. This test, based on the statistical framework developed by Harvey et al. (1998), investigates the null hypothesis that a given forecast effectively contains all the relevant information found in a competing forecast against the alternative hypothesis that the competing forecast holds additional pertinent information beyond what is captured in the given forecast. The associated p-values of this test are reported in Table 3. Our findings suggest that the kitchen sink method fails to encompass narrative forecasts generated by machine learning models, thereby insinuating its inability to

Table 3: **Encompassing Test.** This table reports the p-values of the Harvey, Leybourne, and Newbold (1998) MHLN statistic. The MHLN statistic corresponds to a one-sided (upper-tail) test of null hypothesis that the model forecast for market excess return under the scheme given in the column heading encompasses the forecast under the scheme given in the row heading against the alternative hypothesis that the forecast under the scheme given in the column heading does not encompasses the forecast under the scheme given in the row heading. 0.00 indicates less than 0.005.

	KS	ENET	CSR	PCA	PLS	RF	GBDT	COMB
KS		0.21	0.20	0.15	0.18	0.21	0.16	0.19
ENET	0.00		0.28	0.15	0.17	0.36	0.25	0.63
CSR	0.00	0.11		0.11	0.11	0.28	0.20	0.48
PCA	0.00	0.13	0.17		0.23	0.28	0.21	0.62
PLS	0.00	0.13	0.17	0.20		0.24	0.14	0.63
RF	0.00	0.13	0.19	0.13	0.11		0.21	0.50
GBDT	0.00	0.14	0.23	0.16	0.11	0.46		0.65
COMB	0.00	0.06	0.10	0.06	0.04	0.18	0.07	

fully exploit the information within economic narratives. Conversely, the machine learning combination method proves capable of encompassing all individual machine learning model forecasts at conventional significance levels. This result from the encompassing tests suggests that machine learning combination method successfully leverages the entirety of the information found in economic narratives, thus providing a plausible explanation for their superior performance as demonstrated in Table 2.

3.2 Constructing Narrative-Managed Portfolios

In the previous section, we evaluate the forecasting power of narratives in time-series. In this section, we construct the narrative-managed portfolio and compare its performance with the market portfolio. To achieve this, we first construct the narrative-managed portfolios. The primary objective of a managed portfolio is to maintain a substantial long position during favorable periods and a short position during unfavorable times. Consequently, the portfolio's weight should be proportional to the forecasted market excess returns. However, as noted by Hollstein and Prokopczuk (2022), using the forecasts directly as weights in the market portfolio would make the managed portfolio's average market exposure the main

contributor to its returns. To address this, we introduce two key modifications.

First, the weight assigned to the market portfolio is determined based on the demeaned forecasted return derived from the topic data. This procedure is intended to minimize the factor loading of the managed portfolio with respect to the market portfolio. As a result, an investor can achieve their desired market exposure by simultaneously holding the managed portfolio and the initial portfolio with varying positions

Second, we adjust the managed portfolio's weight so that its volatility matches that of the market portfolio (both observed until time t)[9]. This adjustment allows the performance of the managed portfolio to be compared to that of the market portfolio and other factors found in the literature. To construct a feasible portfolio, ω_t is restricted to [0, 1], as leverage constraint. In constructing the narrative-managed portfolios, we use the period from January 2001 to December 2001 as the initial estimation period for volatility, with the narrative-managed portfolios constructed from January 2002 to December 2020.

The narrative-managed market portfolio's return can be calculated as follows:

$$R_{t+1}^{NMP} = \omega_t R_{t+1} \tag{13}$$

where R_{t+1} refers to the market excess return during month $t+1$, and ω_t represents the weight implied by the narrative obtained at the end of month t. The weight ω_t is calculated as follows:

$$\omega_t = c_t \hat{R}_{t+1}^{Narr} \tag{14}$$

where $\hat{R}_{t+1}^{Narr}$ represents the demeaned market return forecasts derived from the narrative, and c_t is chosen such that the conditional volatility of the managed market portfolio equals that of the market portfolio.

[9]To obtain more timely volatility estimates for the narrative-managed portfolios, we employed a 12-month rolling window due to their relatively stable volatility. In contrast, for the market portfolio, we used an expanding window to achieve a more stable estimate, given its highly volatile nature

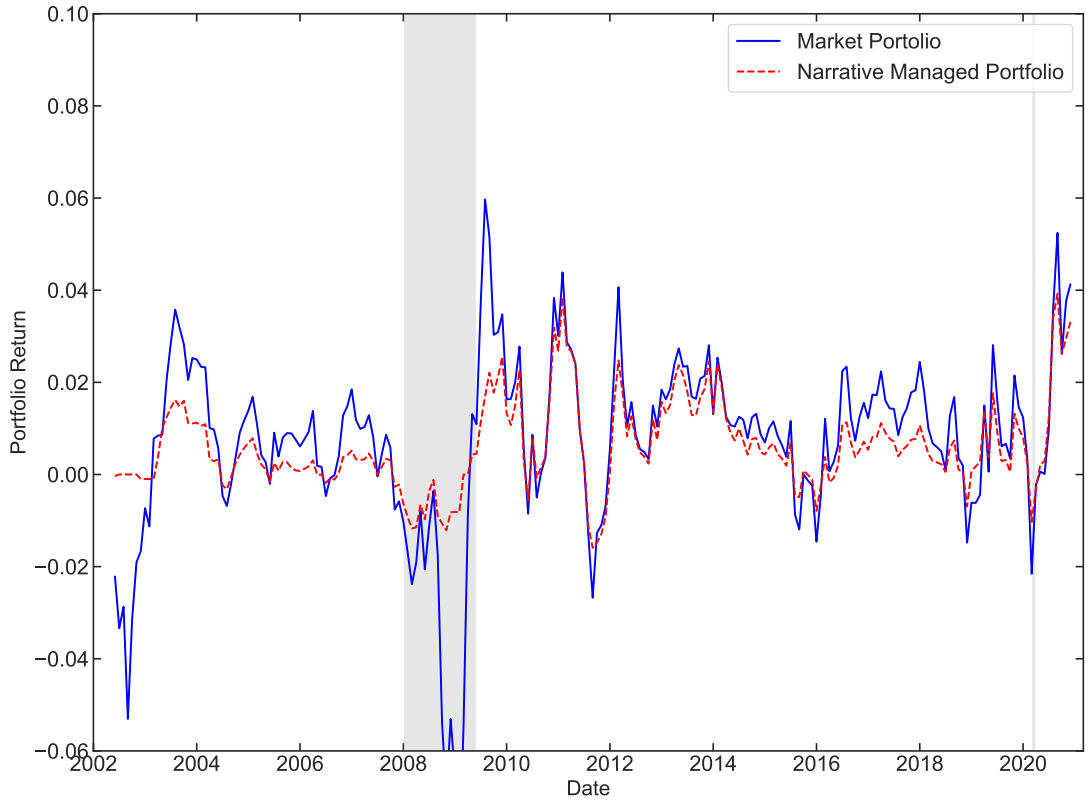


Fig. 3. **Narrative Managed Portfolio Returns**. This figure plots the returns of the narrative-managed portfolio and the market portfolio from 2002 to 2020. To enhance readability, the returns are smoothed using a six-month moving average. The shaded bars represent NBER-dated recession periods.

In the remainder of this section, we compare the performance of the narrative-managed portfolio with that of the market portfolio. Specifically, for the narrative-managed portfolio, we report only the results based on the machine learning combination (COMB). To enhance readability, we smooth the returns of both portfolios using a six-month moving average and plot the time series in Figure 3.

There are two notable observations. First, the narrative-managed portfolio has significantly lower left-tail risk compared to the market portfolio. For example, during the 2008 financial crisis, the market portfolio's minimum returns fell below -6% (even with a six-month moving average), while the narrative-managed portfolio's minimum returns were approximately -1%. Second, the narrative-managed portfolio exhibits right-skewed returns, which

are attractive to investors (Arrow et al., 1974; Markowitz, 1952b; Mitton and Vorkink, 2007; Boyer and Vorkink, 2014). For example, its highest return is about 4%, while the lowest return is about -1%. In contrast, the market portfolio exhibits left-skewed returns.

Then, we examine the economic value of the narrative-managed portfolio compared to the market portfolio. First, we analyze the investors' utility function, particularly when higher moments are considered. Let r be a random variable representing the return of an asset, and $U(r)$ be the utility function of an investor holding this asset. $U(r)$ can be approximated by the third-order Taylor expansion around r's mean $\bar{r}$:

$$E[U(r)] = U(\bar{r}) + \frac{1}{2}U''(\bar{r})E[(r-\bar{r})^2] + \frac{1}{6}U'''(\bar{r})E[(r-\bar{r})^3]$$

If the distribution of r is symmetric, then the third term vanishes, reducing the model to the mean-variance framework. The mean-variance utility function is often used to compute the certainty equivalent return (CER) gain, as seen in prior literature focused on stock return time series prediction (see, e.g., Welch and Goyal, 2008; Rapach et al., 2010; Jiang, Lee, Martin, and Zhou, 2019; Huang, Li, and Wang, 2021).

However, since the market portfolio often experiences more extreme negative returns due to disasters, and the narrative-managed portfolio exhibits right-skewed returns, considering only the mean and variance of the portfolio appears insufficient. Markowitz (1952a) outlines a condition in which mean-variance efficient portfolios may not be optimal, —when an investor's utility is a function of mean, variance, and skewness. De Athayde and Flôres Jr (2004), Satchell (2003), Adcock (2010) and Harvey, Liechty, Liechty, and Müller (2010) use different methods to incorporate the third moment of portfolio return into the utility function. We assume the investor has the following utility function:

$$U = \mu - \lambda\sigma^2 + \tau\gamma$$

where μ and σ^2 are the mean and variance of the asset's return r. Following Ferreira (2024),

γ represents a quantile-based measure of skewness, which can mitigate the impact of outliers on the results:

$$\gamma = (Q_{0.95}(r) - Q_{0.5}(r)) - (Q_{0.5}(r) - Q_{0.05}(r))$$

where $Q_\alpha(r)$ is the αth quantile of the portfolio's return.

Table 4: **Utility Gains With Higher Moments** This table reports the utility for investors holding the narrative-managed portfolio and the market portfolio, as well as the utility gains. The investors have the following utility function:

$$U = \mu - \lambda\sigma^2 + \tau\gamma$$

where μ and σ^2 are the mean and variance of the asset's return r. Following Ferreira (2024), γ represents a quantile-based measure of skewness:

$$\gamma = (Q_{0.95}(r) - Q_{0.5}(r)) - (Q_{0.5}(r) - Q_{0.05}(r))$$

where $Q_\alpha(r)$ is the αth quantile of the portfolio's return. We annualize the utilities by multiplying them by 1,200. For right panel, we set a proportional transaction cost of 10 basis points (bps).

Utility Gains Without Transaction Costs				Utility Gains With Transaction Costs			
	$Utility^{MKT}$	$Utility^{NMP}$	$\Delta Utility$		$Utility^{MKT}$	$Utility^{NMP}$	$\Delta Utility$
Panel A: Mean-variance utility function							
λ=3, τ=0	1.84	5.44	3.59	λ=3, τ=0	1.84	5.29	3.45
λ=5, τ=0	-2.83	4.07	6.89	λ=5, τ=0	-2.83	3.93	6.75
Panel B: Mean-variance-skewness utility function							
λ=3, τ=0.1	-1.48	5.78	7.26	λ=3, τ=0.1	-1.48	5.63	7.11
λ=5, τ=0.1	-6.15	4.41	10.57	λ=5, τ=0.1	-6.15	4.26	10.42

Table 4 reports the results. We annualize the $\Delta Utility$ by multiplying it by 1,200. In Panel A, we find that for investors with a mean-variance utility function, annual utility gains range from 3.59% to 6.89% if they hold the narrative-managed portfolio instead of the market portfolio. The more risk-averse the investors, the higher the utility gains they achieve. In Panel B, for investors with a preference for positive skewness (see, e.g., Arrow et al., 1974; Markowitz, 1952b; Mitton and Vorkink, 2007 and Boyer and Vorkink, 2014), annual utility gains range from 7.26% to 10.57%, nearly double the gains of investors with a mean-variance utility function. Considering that the narrative-managed portfolio requires monthly position adjustments, transaction costs will reduce its performance. Following Moreira and Muir

(2017) and Hollstein and Prokopczuk (2022), we set a proportional transaction cost of 10 basis points (bps). This number is relatively conservative, given the high liquidity of S&P 500 ETFs. We find that transaction costs only slightly reduce $Utility^{NMP}$ and $\Delta Utility$, without affecting the performance of the market portfolio. In summary, the narrative-managed portfolio consistently outperforms the market portfolio across various utility functions, parameters, and whether or not transaction costs are considered.

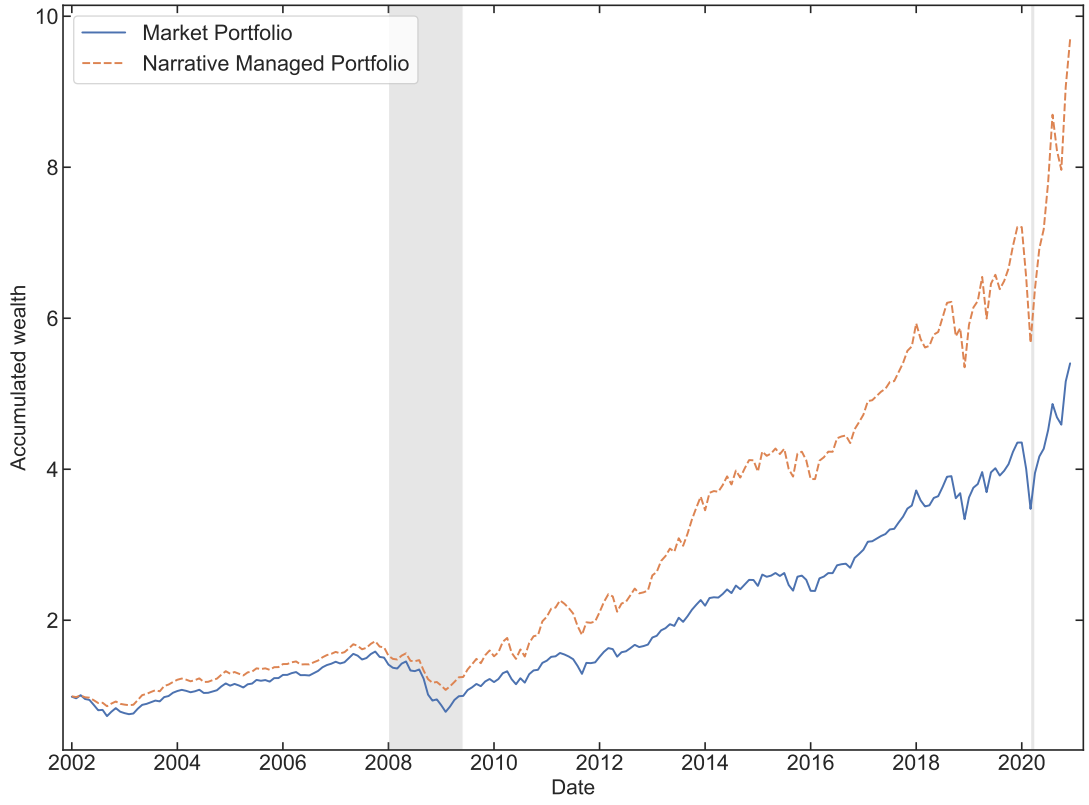


Fig. 4. **Cumulative Wealth**. This figure plots cumulative wealth for the narrative-managed portfolio and market portfolio from 2002 to 2020, both with the same initial $1 investment. The shaded bars are NBER-dated recession periods.

We then compare the cumulative wealth of investors holding the narrative-managed portfolio with those holding the market portfolio in Figure 4. To ensure comparability between the two portfolios, we adjust the position of the narrative-managed portfolio to achieve approximately the same average market exposure as the market portfolio.[10] Unlike the market

[10]For the narrative-managed portfolio, its weight is restricted to a range between 0 and 1. Since stock

portfolio, the narrative-managed portfolio exhibits time-varying exposure to market risk. As shown in Figure 4, the narrative-managed portfolio outperforms the market portfolio. An investment of \$1 in the market portfolio in 2002 yields approximately \$5 by the end of 2020, whereas the narrative-managed portfolio generates a final portfolio value of around \$10, nearly double that of the market portfolio.

3.3 Narrative-Managed Portfolio and Cross-sectional Factor Models

Having constructed the narrative managed portfolio, we proceed to examine whether it can be explained by prominent asset pricing factors. To conduct this test, we employ the following equation:

$$R_t^{NMP} = \alpha + \boldsymbol{\beta}' \boldsymbol{F}_t + \varepsilon_t \tag{15}$$

where R_t^{NMP} is the return of the narrative managed portfolio, and $\boldsymbol{F}_t$ represents the returns of various factors at time t. In this test, a statistically significant α indicates that the superior performance of the narrative managed portfolio cannot be explained by the known asset pricing factors. Barillas, Kan, Robotti, and Shanken (2020) demonstrates that a statistically significant α implies a statistically significant difference in squared Sharpe ratios between the tangency portfolio implied by a model and that of the model augmented by the managed portfolio. Consequently, an investor could enhance their portfolio's performance by incorporating the managed portfolio into their investment opportunity set.

In our primary analysis, we investigate the following factor models: (i) the Capital Asset Pricing Model (CAPM), (ii) Fama and French (1992) three-factor model (FF3), (iii) Carhart (1997) four-factor model (Carhart), (iv) Fama and French (2015) five-factor model (FF5),

returns approximately follow a random walk, we assume that the weights derived from stock return forecasts also follow this pattern. Consequently, the average weight of the narrative-managed portfolio will be around 0.5. To achieve the desired market exposure, we combine the narrative-managed portfolio with a 50% allocation to the market portfolio. This ensures that, on average, both portfolios allocate approximately 100% of their weight to the market portfolio over time.

and (v) Hou et al. (2021) five-factor model (HXZ). The factor data is obtained as following: Data for the Hou et al. (2021) factors are sourced from the Global-Q website, while all other factors are acquired from Kenneth French's website.

Table 5: **Alphas of Narrative Managed Portfolios.** This table reports the annualized average returns and factor models alphas of narrative managed portfolios. In parentheses are the t-statistics based on robust Newey and West (1986) standard errors using six lags. *, ** and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Average return	α^{CAPM}	α^{FF3}	$\alpha^{Carhart}$	α^{FF5}	α^{HXZ}
ENET	7.04***	3.67**	3.08**	2.76**	2.05*	3.48*
	(3.32)	(2.47)	(2.28)	(2.04)	(1.77)	(1.83)
CSR	4.94***	2.31**	2.23**	1.95*	1.33	2.18*
	(3.26)	(2.19)	(2.18)	(1.92)	(1.37)	(1.79)
PCA	8.5***	2.53*	2.51**	2.14*	1.20	2.52*
	(3.26)	(1.93)	(2.02)	(1.81)	(1.27)	(1.87)
PLS	8.02***	3.18**	3.02**	2.61**	1.70	3.29**
	(3.49)	(2.27)	(2.29)	(2.07)	(1.61)	(1.98)
RF	4.73***	2.09**	1.92*	1.62	1.60*	1.85
	(3.11)	(2.01)	(1.93)	(1.59)	(1.84)	(1.62)
GBDT	4.75***	1.78	1.62	1.49	1.28	2.39*
	(2.72)	(1.5)	(1.48)	(1.34)	(1.33)	(1.66)
COMB	7.49***	3.61***	3.43***	3.02**	2.31**	3.28**
	(3.93)	(2.62)	(2.67)	(2.34)	(2.21)	(2.23)

We begin our analysis with the first column of Table 5. All narrative-managed portfolios demonstrate superior performance, as evidenced by their elevated t-values, leading to a more favorable return-risk tradeoff. Moving to the second row, all CAPM alphas are positive and significant. The combination method, however, stands out, generating CAPM alpha of approximately 3.61% per annum, which is significant at the 1% level. This result is noteworthy as it suggests that index investors could substantially enhance their investment opportunity set by incorporating the narrative-managed portfolio (COMB). The findings are highly consistent across multi-factor models. Although narrative-managed portfolios generated by individual machine learning models often fail to generate significantly positive alphas (especially for FF5), the narrative-managed portfolio (COMB) continues to produce strong results: it delivers alphas exceeding 2% per annum, significant at the 5% level across

various factor models.

Additionally, we observe two notable patterns. First, the CAPM alphas are significantly smaller than the average returns of the managed market portfolios, suggesting that these portfolios have positive exposure to the market portfolio. This outcome is reasonable, given that the narrative-managed portfolios are restricted from short selling. Second, the CAPM alphas are higher than those estimated from multi-factor models. For example, narrative-managed portfolios generate CAPM alphas of around 3%, while their FF5 alphas are only around 1.5%. This observation supports the notion that factors themselves contain conditioning information about future investment opportunities, which aligns with the findings of Dong, Li, Rapach, and Zhou (2022).

We next investigate whether the narrative-managed portfolio can enhance the investment opportunity set of the market portfolio or a set of asset pricing factors. In Figure 5, we illustrate the efficient frontiers of various asset pricing factors, both with and without the narrative-managed portfolio.[11] As observed in the upper-left subfigure, the efficient frontier resulting from the managed market portfolio (COMB) and the market portfolio is located significantly to the top left of the market portfolio. This pattern suggests that incorporating the managed portfolio could substantially improve the initial investment opportunity set. Similar patterns are present in the other four subfigures. In the case of multi-factor models, the gap between the two efficient frontiers is narrower compared to CAPM, which is consistent with the findings in Table 5, where CAPM alphas exceed those derived from multi-factor models. The relationship between the alphas and the differences in the efficient frontiers also aligns with the conclusions of Barillas et al. (2020)

We also show the narrative managed portfolio (COMB) resides outside the mean-variance frontier of well-established asset pricing models. Huberman and Kandel (1987) originally developed a mean-variance spanning test on the hypothesis of whether N assets can be spanned in the mean-variance space by a set of K assets. Subsequent research by De Roon,

[11]For CAPM, which has only one market factor, it is represented as a dot.

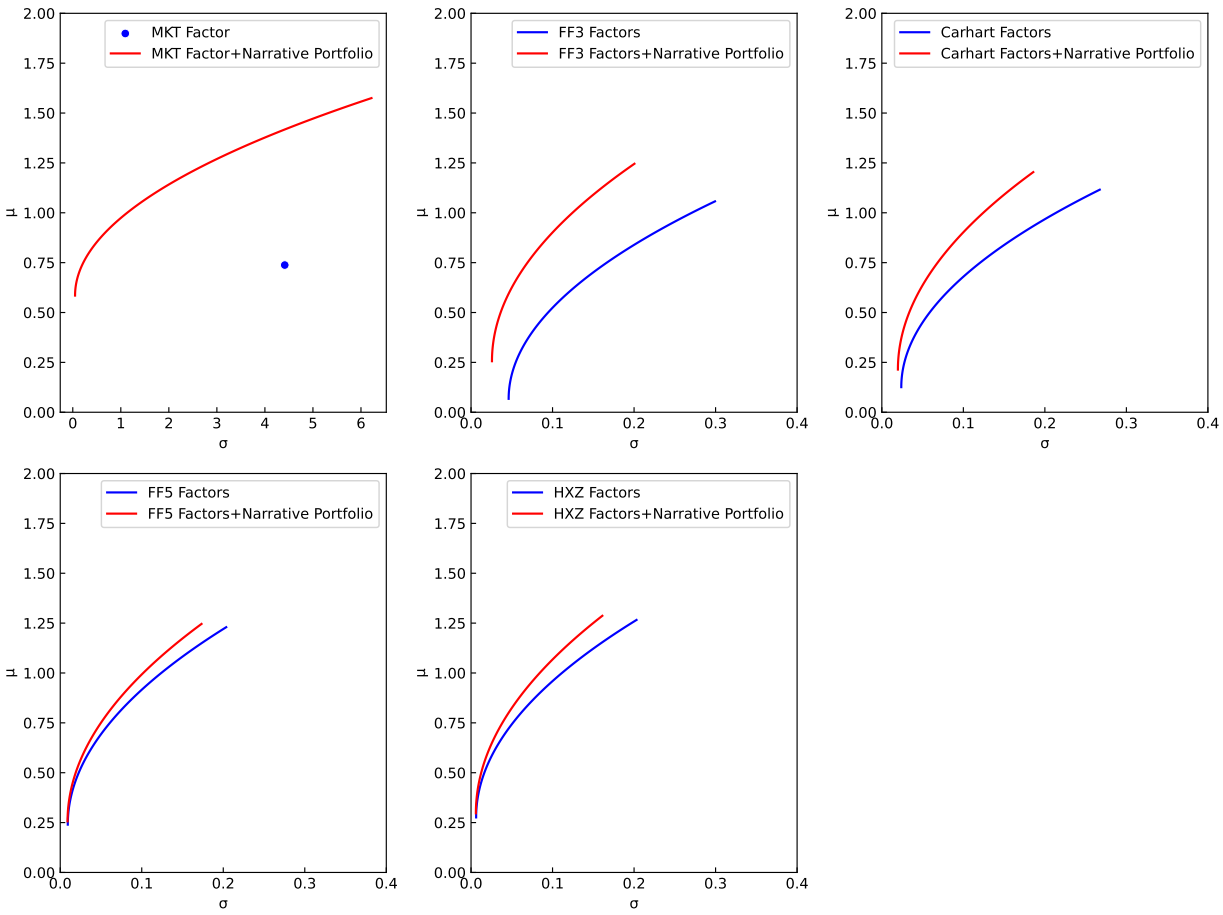


Fig. 5. **Efficient Frontiers of Different Asset Pricing Models and Narrative Factor**. The blue line in each sub-figure plots the efficient frontier of CAPM, FF3 factors, Carhart factors, FF5 factors, or HXZ factors, separately. The red line in each sub-figure plots the new efficient frontier when adding narrative managed portfolio based on COMB forecasts.

Nijman, and Werker (2001) , Korkie and Turtle (2002) , and Kan and Zhou (2008) provided additional tests for the same hypothesis. To execute this test, we utilize the following equation:

$$R_t^{NMP} = \alpha + \boldsymbol{\beta}' \boldsymbol{F}_t + \varepsilon_t \tag{16}$$

where $\boldsymbol{F}_t$ represents the returns of various factors at time t. The spanning hypothesis corresponds to the subsequent parametric constraints on the model,

$$H_0: \quad \alpha = 0, \quad l'\beta = 1 \tag{17}$$

where l represents a vector of ones with length adhering to matrix algebra rules. Following Kan and Zhou (2008), we conduct six spanning tests: (1)Wald test under conditional homoskedasticity, (2)Wald test under independent and identically distributed (IID) elliptical distribution, (3)Wald test under conditional heteroskedasticity, (4)Bekerart-Urias spanning test with errors-in-variables (EIV) adjustment, (5)Bekerart-Urias spanning test without the EIV adjustment and (6)DeSantis spanning test. All six tests have asymptotic chi-squared distribution with $2N$ ($N = 1$ here) degrees of freedom.

Panel A of Table 6 reports the test results for different asset pricing models. The results strongly reject the hypothesis that the narrative managed market portfolio(COMB) lies within the mean-variance frontier across all asset pricing models. This result suggests that the narrative managed market portfolio is a unique portfolio capturing information from economic narratives often neglected in existing asset pricing models.

3.4 Incremental Information of Narrative

In this section, we further investigate whether narrative has additional information in managing the market portfolio when we control well-documented economic predictors.

We compare the predictive power of economic predictors and narrative topics. Table 2 reports the results. We find that forecasts based on narrative topics have significantly higher R^2_{OS} and CW-test statistics than those based on economic predictors. For example, while the COMB model based on narrative topics has an R^2_{OS} of 2.61%, significant at the 5% level, the COMB model based on economic predictors only achieves an R^2_{OS} of 0.02% and is not statistically significant.

To extend our analysis within a portfolio framework, we utilize the method outlined in Equation (14) to formulate a econ-managed portfolio based on the forecasts based on

Table 6: **Mean-Variance Spanning Tests**. This table presents the mean-variance spanning test statistics as outlined in Huberman and Kandel (1987), conducted under various distribution assumptions. Here, W is the Wald test under conditional homoscedasticity, W_e is the Wald test under IID elliptical, W_a is the Wald test under the conditional heteroscedasticity, J_1 is the Bekerart-Urias test with the Errors-in-Variables (EIV) adjustment, J_2 is the Bekerart-Urias test without the EIV adjustment, and J_3 is the DeSantis test. Panel A reports the results of testing whether the narrative-managed portfolio can be spanned by asset pricing factors. Panel B reports the results of testing whether the narrative-managed portfolio can be spanned by asset pricing factors and an Econ-managed portfolio. The Econ-managed portfolio is constructed using 12 economic predictors in Welch and Goyal (2008), along with the time-series momentum variable. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Panel A: Asset Pricing Factors						
	W	W_e	W_a	J_1	J_2	J_3
CAPM	711.26***	282.18***	187.02***	60.11***	60.86***	335.57***
FF3	188.6***	107.93***	81.19***	47.64***	49.89***	94.81***
Carhart	52.87***	28.28***	23.60***	24.25***	25.41***	32.55***
FF5	9.46***	6.24**	8.84***	9.55***	9.89***	10.21***
HXZ	15.45***	9.84***	12.13***	12.83***	12.88***	13.69***
Panel B: Economic Factor and Asset Pricing Factors						
	W	W_e	W_a	J_1	J_2	J_3
CAPM+Econ	277.81***	91.08***	41.99***	39.13***	39.28***	192.64***
FF3+Econ	145.15***	71.09***	39.04***	39.86***	40.42***	90.15***
Carhart+Econ	49.94***	24.57***	13.55***	18.76***	19.46***	24.38***
FF5+Econ	10.38***	6.23**	6.43**	7.55**	7.62**	8.69***
HXZ+Econ	15.85***	9.24***	9.13***	12.47***	12.31***	13.30***

economic predictors and COMB model. Subsequently, we incorporate the econ-managed portfolio as a covariate and estimate the ensuing regression:

$$R_t^{NMP} = \alpha + \boldsymbol{\beta}' \boldsymbol{F}_t + \psi \boldsymbol{Econ}_t + \varepsilon_t \tag{18}$$

where $\boldsymbol{F}_t$ represents the returns of various factors at time t while $\boldsymbol{Econ}_t$ represents the returns of Econ-managed portfolio.

Table 7 presents the results. Two points are worth emphasizing. First, the alphas of narrative-managed portfolios remain positive and significant, indicating that narratives provide incremental information beyond existing economic or financial predictors when man-

Table 7: **Alphas of Narrative Managed Portfolios After Controlling Economic Managed Portfolio.** This table reports the annualized average returns and factor model alphas of the narrative-managed portfolios after controlling for the Econ-managed portfolio. The Econ-managed portfolio is constructed using 12 economic predictors in Welch and Goyal (2008), along with the time-series momentum variable. In parentheses are the t-statistics based on robust Newey and West (1986) standard errors using six lags. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	$\alpha^{CAPM+Econ}$	$\alpha^{FF3+Econ}$	$\alpha^{Carhart+Econ}$	$\alpha^{FF5+Econ}$	$\alpha^{HXZ+Econ}$
ENET	3.65**	3.06**	2.73**	2.04*	3.49*
	(2.48)	(2.30)	(2.05)	(1.80)	(1.86)
CSR	2.29**	2.21**	1.94*	1.32	2.19*
	(2.21)	(2.19)	(1.94)	(1.39)	(1.83)
PCA	2.5*	2.48**	2.11*	1.19	2.53*
	(1.93)	(2.01)	(1.82)	(1.29)	(1.91)
PLS	3.18**	3.01**	2.6**	1.70	3.29**
	(2.28)	(2.29)	(2.07)	(1.62)	(1.99)
RF	2.1**	1.93**	1.63	1.60*	1.84
	(2.03)	(1.97)	(1.61)	(1.85)	(1.54)
GBDT	1.82	1.65	1.53	1.30	2.36
	(1.46)	(1.45)	(1.30)	(1.27)	(1.59)
COMB	3.59***	3.4***	2.99**	2.30**	3.30**
	(2.76)	(2.81)	(2.48)	(2.34)	(2.27)

aging the market portfolio. Second, similar to the previous findings, the COMB model continues to outperform the single machine learning models. This finding further supports that the machine learning combination method is more suitable and stable for capturing information from high-dimensional textual data.

Panel B of Table 6 reports the results for the spanning tests. All six spanning tests are significant at the 1%, rejecting the null hypothesis that the narrative-managed portfolio in within the mean-variance frontier of asset pricing factors and econ-managed portfolio. In summary, we find that narrative contains incremental information beyond existing economic or financial predictors.

4. The Source of Predictive Power

In this section, we investigate which topics contribute most to predictability. As indicated by Gu et al. (2020) and Bali, Goyal, Huang, Jiang, and Wen (2020), we gauge the importance of each topic by measuring the decrease in R^2_{os} , subsequently ranking them. Specifically, we calculate this decrease by setting the values of a particular topic to zero in each of the seven models while keeping the remaining model estimates unchanged.

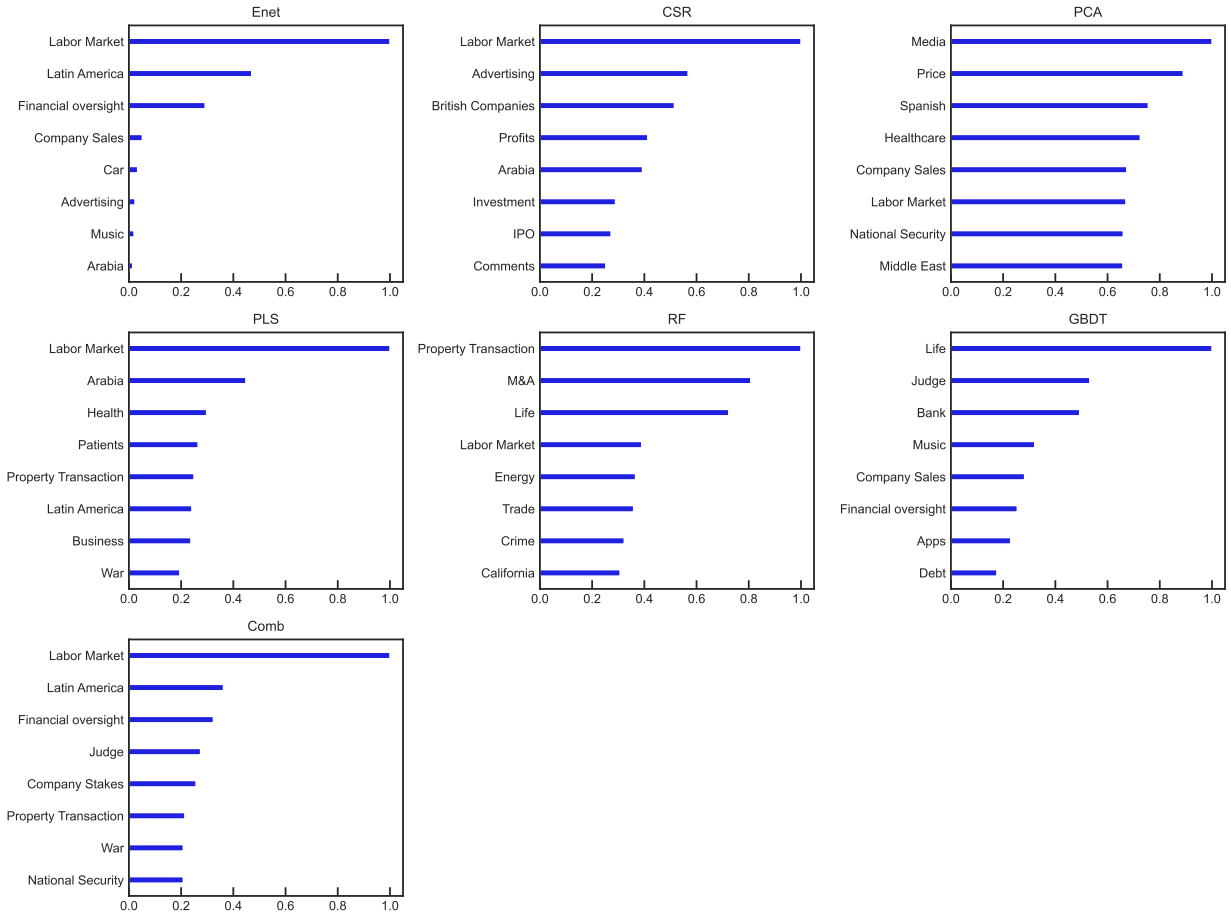


Fig. 6. **Topic Importance**. This figure presents the top eight important topics in each model. We calculate the the reduction in R^2_{os} from setting all values of a given topic to zero in each model to get the given topic importance and scaled it by the most importance topic (topic with maximum R^2_{os} reduction). The importance of the most important topic within each model is normalized to one

Figure 6 outlines the eight most important topics across various forecasting models. First,

more than half of these important topics are linked to disaster risk, including *Labor Market*, *Latin America*, *Middle East*, *Arabia*, and *War*, among others. Notably, the *Labor Market* topic, which predominantly references labor market conditions, plays a significant role in nearly all machine learning models (except for GBDT). This result aligns with the findings of Favilukis and Lin (2016); Belo, Lin, and Bazdresch (2014), which suggest that labor market frictions closely link labor market conditions to asset prices. Second, several geopolitics-related topics also play an important role, such as *Middle East*, *Energy*, *Arabia*, and *War*. The presence of these topics is expected, as news of conflicts, such as the Afghan war, can influence public expectations and confidence, consequently affecting stock returns, consistent with Hirshleifer et al. (2023, 2024). Third, the *Latin America* topic also plays an important role. Its impact on the U.S. stock market can be attributed to its unique geographical position. Lahrech and Sylwester (2011) discovered significant equity co-movements between the U.S. and Argentina, Brazil, Chile, and Mexico. This implies that the predictive power of the narrative for the stock market may stem from capturing latent disaster risk, which drives asset price variations.

4.1 Disaster Risk and Portfolio Performance

In the previous subsection, we find that topics related to disaster risk contribute most to predictability. In this section, we directly examine the relationship between disaster risk and the performance of the narrative-managed portfolio. According to Wachter (2013), Gabaix (2012) and Baron et al. (2022), if disaster risk is time-varying, the equilibrium disaster risk premium will also fluctuate over time. Assets exposed to time-varying disaster risk will be priced with a positive risk premium. Therefore, if the narrative's predictive power over the stock market arises from capturing latent disaster risk, an increase in the disaster index should lead to better future performance of the narrative-managed portfolio.

To test this hypothesis, we estimate the following time series regression:

$$R_t^{NMP} = \beta_0 + \beta_1 DisasterRisk_{t-1} + \boldsymbol{\beta} F_t + \varepsilon_t \tag{19}$$

where R_t^{NMP} is the return of the narrative-managed portfolio. The estimation is performed from January 2001 to December 2020. As in Equation (19), we also use cross-sectional factors as control variables. We employ six proxies to measure disaster risk. The first is Disaster Probability, estimated using option prices as per Barro and Liao (2021). The second is the CBOE Volatility Index (VIX), which reflects the market's expectation of future volatility and is used as a proxy for latent financial disaster. The third is Economic Policy Uncertainty (EPU), measured by Baker et al. (2016). The fourth is a risk aversion index generated by Bekaert et al. (2022), calculated from high-frequency financial variables. The fifth is the unemployment rate from FRED, used to measure labor market conditions. The final disaster index is the NBER-based Recession Indicators from FRED. All disaster risk indices are smoothed using a 12-month moving average and then standardized.

Table 8 reports the results. All estimated slopes are positive, and most are significant at the 5% level. The results imply that as the disaster index rises, the performance of the narrative-managed portfolio improves. For example, a one-standard-deviation increase in EPU is associated with a 0.34% to 0.47% increase in the narrative-managed portfolio's monthly return, controlling for other factors. Overall, Table 8 suggests that the narrative-managed portfolio's performance is closely related to disaster risk.

4.2 Predicting Disaster Risk

In this section, we further investigate whether the narrative contains information related to disaster risk. If it does, disaster risk can be predicted by the narrative. Since news contains information related to uncertainty, Manela and Moreira (2017) use front-page articles of the Wall Street Journal to construct a news-based implied volatility index. Following their

Table 8: **Disaster Risk and Portfolio Performance.** This table reports the results of regressing narrative managed portfolio performance on the disaster index. We use six proxies to measure disaster risk: Disaster Probability, estimated using option prices from Barro and Liao (2021). VIX, the CBOE Volatility Index; EPU, economic policy uncertainty measured by Baker et al. (2016); Risk Aversion, generated by Bekaert et al. (2022); Unemployment, the unemployment rate from FRED; and Recession, NBER based Recession Indicators from FRED. All disaster risk indexes are smoothed using a 12-month moving average and then standardized. Different cross sectional factors are used as control variable. Reported are the regression coefficient and Newey-West t-value. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period spans from January 2001 to December 2020.

$R_t^{NMP} = \beta_0 + \beta_1 DisasterRisk_{t-1} + \boldsymbol{\beta} F_t + \varepsilon_t$

	F:	CAPM	FF3	Carhart	FF5	HXZ
Disaster Prob	β_1	0.25**	0.25**	0.38***	0.19*	0.35***
	t_{NW}	(1.98)	(2.05)	(3.21)	(1.66)	(2.69)
VIX	β_1	0.27**	0.26**	0.38***	0.20*	0.36***
	t_{NW}	(2.29)	(2.35)	(3.59)	(1.86)	(3.04)
EPU	β_1	0.40***	0.38***	0.42***	0.34***	0.47***
	t_{NW}	(3.97)	(4.08)	(4.70)	(4.03)	(4.48)
Risk Aversion	β_1	0.17	0.17	0.31***	0.13	0.26**
	t_{NW}	(1.46)	(1.55)	(2.82)	(1.22)	(2.26)
Unemployment	β_1	0.16**	0.17**	0.17**	0.15*	0.17**
	t_{NW}	(2.25)	(2.23)	(2.18)	(1.83)	(2.15)
Recession	β_1	0.26*	0.27*	0.4***	0.17	0.36***
	t_{NW}	(1.66)	(1.72)	(3.4)	(1.32)	(2.76)

approach, we construct a narrative-based disaster risk index. We use the same proxies to measure disaster risk, excluding the NBER-based Recession Indicators [12], and estimate the narrative-based disaster risk forecasts using the following equation:

$$DisasterRisk_{t+1} = G\left(\boldsymbol{Narr}_t\right) + \varepsilon_{t+1} \tag{20}$$

Due to the limited number of observations, rather than dividing the full sample into three parts—training set, validation set, and test set—as done by Manela and Moreira (2017) for hyperparameter tuning, we use 10-fold cross-validation to determine the hyperparameters. First, we divide the full sample into 10 parts and use the first 9 parts to train the machine

[12]Because the NBER-based Recession Indicator is a 0-1 variable rather than a continuous variable

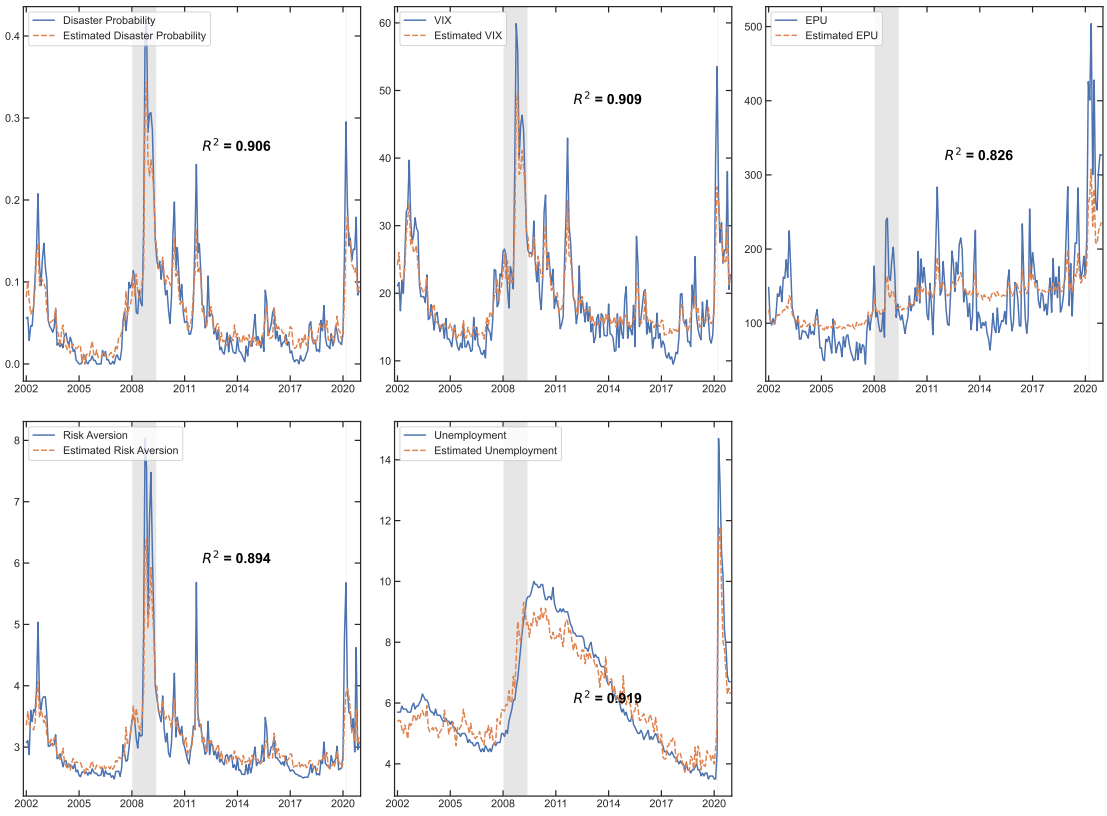


Fig. 7. **Forecasting Disaster Risk by narrative.** This figure plots the time series of narrative-based disaster risk and actual disaster risk. We use five proxies to measure disaster risk: Disaster Probability, estimated using option prices from Barro and Liao (2021). VIX, the CBOE Volatility Index; EPU, economic policy uncertainty measured by Baker et al. (2016); Risk Aversion, generated by Bekaert et al. (2022); and Unemployment, the unemployment rate from FRED. R^2Reg is reported, estimated from a subsequent univariate Ordinary Least Squares (OLS) regression of actual disaster risk on narrative-based disaster risk. The sample period spans from January 2002 to December 2020.

learning models under different parameters. Then, we predict the disaster risk in the remaining part and compute the MSFE. After repeating this process 10 times, the model with the lowest MSFE is selected. In the next step, this model is trained using the full sample, and its forecasts are used as the narrative-based disaster risk. Following Manela and Moreira (2017), we report R^2Reg , which is estimated from a subsequent univariate Ordinary Least Squares (OLS) regression of actual disaster risk on narrative-based disaster risk.

Figure 7 presents the results. We find that the narrative-based disaster risk effectively captures the trend of disaster risk. For example, in the cases of Disaster Probability, VIX,

and Risk Aversion, the narrative-based disaster risk captures the spikes during recessions and the subsequent troughs afterward. For every disaster proxy, the $R^2 Reg$ values are around 90%. These results suggest that the narrative contains information related to latent disaster risk.

4.3 Predicting Discount Rate Related Variables

In this section, we test whether the narrative can predict discount rate related variables. A wide range of discount rate related variables is closely connected to time-varying disaster risk (Gabaix, 2012; Wachter, 2013; Gourio, 2013; Tsai, 2015; Tsai and Wachter, 2016; Baron et al., 2022). These variables can be classified into the following five categories:

Dividend (Earnings)/Price Ratio: as demonstrated by Gabaix (2012) and Wachter (2013), when disaster risk increases, the heightened disaster risk premium leads to a reduction in equity prices, thereby raising the dividend/price ratio of the equity market index. The dividend (earnings)/price ratio is calculated as the difference between the log of dividends (earnings) and the log of prices (Campbell, 1987; Campbell and Shiller, 1988b; Fama and French, 1988; Welch and Goyal, 2008).

Term Spread: Tsai (2015) shows that as disaster risk increases, inflation rises, the prices of nominal bonds fall, and the premium that investors demand for holding nominal bonds increases, resulting in a widening of the term spread. The Term Spread is computed as the difference between the long-term yield on government bonds and the Treasury bill rate (Campbell, 1987; Fama and French, 1989; Welch and Goyal, 2008).

Corporate Credit Spreads: Gabaix (2012), Wachter (2013) and Gourio (2013) show that as disaster risk increases, default risk rises, leading to an increase in corporate credit spreads. Corporate Credit Spreads are measured as the difference between BAA and AAA-rated corporate bond yields (Fama and French, 1989; Welch and Goyal, 2008).

Market Volatility: Wachter (2013) highlights that, under suitable model assumptions, equity volatility increases as a function of time-varying disaster risk. Market volatility is

estimated as the square root of the average daily volatility for the current month.

Value Premium: Gabaix (2012) and Tsai and Wachter (2016) show that value stocks have higher exposure to disaster risk than growth stocks, causing the value premium to increase when disaster risk rises. We use three equity factor long-short portfolios, sorted by the Book/Price (B/P), Dividend/Price (D/P), or Earnings/Price (E/P) ratios, respectively, as proxies for the value premium.

If the narrative contains information related to time-varying disaster risk, it should have significant predictive power for the above variables. Based on narrative information at time t, we predict these variables for the next 12 months as follows:

$$Y_{t+1:t+12} = G\left(\boldsymbol{Narr}_t\right) + \varepsilon_{t+1:t+12} \tag{21}$$

where $Y_{t+1:t+12}$ represents the discount rate-related variables for the next 12 months. Six machine learning models are used for prediction, and we report only the combination (COMB) of these models. Table 9 reports the R^2_{OS} and CW-t results. All variables have positive R^2_{OS} values and are significant at the 1% level. For example, for the three value premium measurements, the R^2_{OS} values range from 13.44% to 17.82%, indicating that, compared with the historical average of value premium measurements, narrative-based forecasts can reduce mean squared forecast errors by nearly 20%. These results are consistent with Table 7, showing that the narrative contains information related to latent disaster risk.

4.4 Sticky Beliefs of Investors

In the previous sections, we find that the narrative contains information related to disaster risk, which explains its ability to predict future stock returns. In this section, we examine an important question: Can investors respond promptly to new information? Mankiw and Reis (2002), Coibion and Gorodnichenko (2015), Bouchaud et al. (2019) and Farmer et al. (2022) show that people typically have sticky beliefs; in other words, they often underreact to new

Table 9: **Predicting Discount Rate Related Variables.** This table reports the R^2_{OS} and CW-t results for discount rate-related variables, which are classified into the following five categories: Dividend (Earnings)/Price Ratio, Term Spread, Corporate Credit Spreads, Market Volatility, and Value Premium. Six machine learning models are used for prediction, and we report only the combination (COMB) of these models in this table. R^2_{OS} is the out-of-sample statistic as defined by Campbell and Thompson (2008). CW-t is the Clark and West (2007) statistic. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The out-of-sample period spans from January 2001 to December 2020.

$Y_{t+1:t+12} = G(\boldsymbol{Narr}_t) + \varepsilon_{t+1:t+12}$		
	R^2_{os}	CW-t
Dividend/Price Ratio	71.10%	12.95***
Earnings/Price Ratio	45.11%	5.02***
Term spread	51.92%	12.88***
Corporate Credit Spreads	53.59%	4.88***
Market Volatility	44.35%	5.83***
Value Premium: DP LS Spread	13.44%	5.43***
Value Premium: BM LS Spread	17.82%	6.73***
Value Premium: EP LS Spread	17.40%	6.75***

information. Specifically, Cohen et al. (2020) and Hirshleifer and Sheng (2022) demonstrate that investors underreact to news due to their limited attention.

Following Mankiw and Reis (2002), Coibion and Gorodnichenko (2015), Bouchaud et al. (2019) and Farmer et al. (2022), we test whether investors have sticky beliefs using the following equation:

$$F_t[R_{t+1:t+12}] = \mu + \lambda F_{t-h}[R_{t-h+1:t-h+12}] + (1-\lambda)\hat{R}^{Narr}_{t+1:t+12} + \varepsilon_{t+1:t+12} \tag{22}$$

where F_t denotes conditional expectations under investors' subjective beliefs at time t, as measured by survey data sourced from Graham-Harvey's survey of CFOs and the Livingston survey. In Graham-Harvey's survey, CFOs provide their expectations for stock returns over the next 12 months in the final month of each quarter. We use the average of these CFOs' expectations to capture the overall investor expectation of market returns. In the Livingston survey, respondents provide their expectations for stock returns over the next 12 months twice

a year, in June and December.[13] The median of respondents' expectations is used to measure the overall investor expectation of market returns. $\hat{R}^{Narr}_{t+1:t+12}$ can be viewed as conditional expectations for the next 12-month market return based on narratives.[14] $F_{t-h}[R_{t-h+1:t-h+12}]$ represents investors' conditional expectations from the previous period, where h equals 3 for Graham-Harvey's survey and 6 for the Livingston survey. In Equation (22), λ represents the stickiness of investors' beliefs.

Table 10: **Sticky Beliefs Tests.** This table presents the estimated λ values and Newey-West t-values as

$$F_t[R_{t+1:t+12}] = \mu + \lambda F_{t-h}[R_{t-h+1:t-h+12}] + (1-\lambda)\hat{R}^{Narr}_{t+1:t+12} + \varepsilon_{t+1:t+12}$$

where F_t denotes conditional expectations under agents' subjective beliefs at time t, as measured by survey data sourced from Graham-Harvey's survey of CFOs and the Livingston survey. $\hat{R}^{Narr}_{t+1:t+12}$ can be viewed as conditional expectations for the next 12-month market return based on narratives. h equals 3 for Graham-Harvey's survey and 6 for the Livingston survey. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample periods all start in January 2002 but end differently: December 2019 for Graham-Harvey's survey and December 2020 for the Livingston survey.

Data Source	λ	t_{NW}	$1-\lambda$	t_{NW}
Graham-Harvey's survey	0.961***	70.32	0.039***	2.84
Livingston survey	0.805***	11.19	0.195***	2.72

If λ equals one, investors do not react to any new information and always maintain their expectations from the previous period. If λ ranges between 0 and 1, investors' beliefs are sticky, meaning they partially carry over their prior period beliefs and underreact to new information. Table 10 reports the estimated values of λ. Two interesting observations

[13] In the Livingston survey, respondents are asked to provide their expectations for the S&P 500 index 12 months in advance, rather than stock returns directly. The mailed questionnaire includes the previous month's S&P 500 index as a reference value. For example, the questionnaire for the June survey is mailed in May with April's S&P 500 index, and the December questionnaire is mailed in November with October's S&P 500 index. Following the Livingston survey's documentation, we assume that investors base their expectations on the index value provided in the questionnaire and calculate the expected stock return as follows:

$$\text{Stock Return Expectation} = \frac{\text{S\&P 500 Index Expectation}}{\text{S\&P 500 Index Reference Value}}.$$

[14] When using different survey data, we matched the time horizons of $F_t[R_{t+1:t+12}]$ and $\hat{R}^{Narr}_{t+1:t+12}$. For instance, in Graham-Harvey's survey, the predictions are based on the narrative from the past 3 months. For the Livingston survey, predictions are based on the narrative from the past 6 months.

emerge. First, all estimated λ values are above 0.8 and significant at the 1% level. This result is consistent with Mankiw and Reis (2002), Coibion and Gorodnichenko (2015), Bouchaud et al. (2019), Bouchaud et al. (2019) and Farmer et al. (2022), suggesting that investors have very sticky beliefs, carrying over more than 80% of their expectations from the previous period. Second, the estimated $1-\lambda$ values range from 3.9% to 19.5% and are also significant at the 1% level. This finding indicates that although investors have sticky beliefs, they do learn from narratives, albeit slowly.

Following Coibion and Gorodnichenko (2015) and Farmer et al. (2022), we also explore the relationship between forecast errors and narrative-based expectations using the following equation:

$$R_{t+1:t+12} - F_t[R_{t+1:t+12}] = \alpha + \beta \hat{R}^{Narr}_{t+1:t+12} + \varepsilon_{t+1:t+12} \tag{23}$$

where $R_{t+1:t+12}$ denotes the cumulative realized market return from time $t+1$ to time $t+12$. If investors underreact to new information from narratives, their expectations change more slowly than the market, resulting in a positive β. Conversely, if investors overreact to new information from narratives, their expectations change more quickly than the market, and β will be negative. If β is close to zero, it indicates that investors respond to new information appropriately.

Table 11 reports the regression results. The β values are all positive, ranging from 1.19 to 1.40, and significant at the 1% level. The results suggest that investors' expectations change more slowly than the market, indicating that they underreact to new information from narratives, consistent with the findings in Table 10. Thus, we conclude that although narratives contain disaster risk-related information, investors respond slowly to this information; in other words, they underreact to disaster risk-related information.

To reconfirm the previous findings, following Sims (1980) and Moreira and Muir (2017), we also run a VAR model to examine how narrative forecasted returns and market returns respond to a negative narrative shock over time. Figure 8 presents the results. We find that the impact on narrative forecasted returns dies out fairly quickly. However, the movement

Table 11: **Return Forecast Errors and Sticky Beliefs** This table presents the slopes, Newey-West t-values, and R^2 from the regression of

$$R_{t+1:t+12} - F_t[R_{t+1:t+12}] = \alpha + \beta \hat{R}^{Narr}_{t+1:t+12} + \varepsilon_{t+1:t+12}$$

where $R_{t+1:t+12}$ denotes the cumulative realized market return from time $t+1$ to time $t+12$. F_t denotes conditional expectations under agents' subjective beliefs at time t, as measured by survey data sourced from Graham-Harvey's survey of CFOs and the Livingston survey. $\hat{R}^{Narr}_{t+1:t+12}$ can be viewed as conditional expectations for the next 12-month market return based on narratives. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample periods all start in January 2002 but end differently: December 2019 for Graham-Harvey's survey and December 2020 for the Livingston survey.

Data Source	β	t_{NW}	R^2
Graham-Harvey's survey	1.19***	3.16	33.40
Livingston survey	1.40***	6.96	46.40

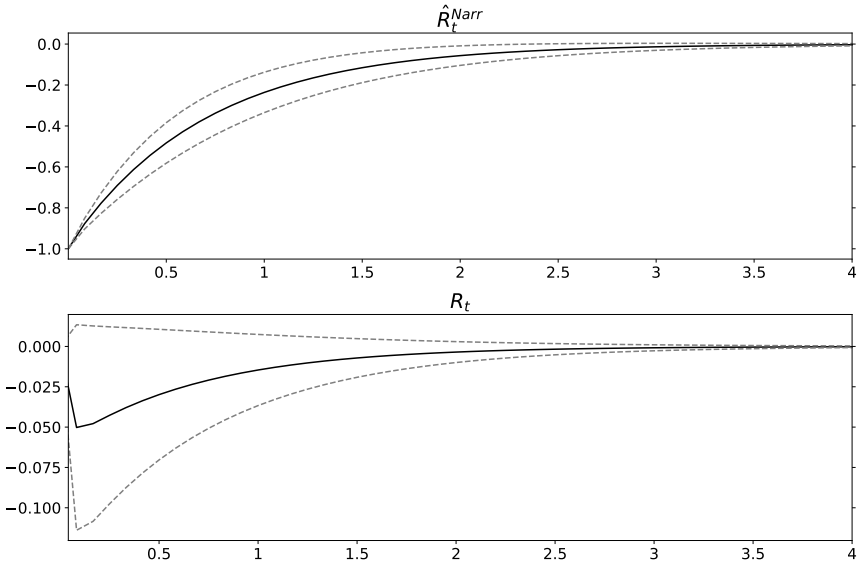


Fig. 8. **Impulse Response of One Negative Narrative Shock for Narrative-based Predicted Returns and Market Returns.** This figure plots the impulse response function to trace the effects of a negative shock to narrative-based predicted returns. The x-axis is in years. We compute impulse responses using a VAR of realized market returns and the narrative-based predicted returns. The dashed lines indicate one-standard deviation confidence intervals.

of market returns is much smaller but more prolonged. The rapid adjustment in narrative returns, along with the slower movement in market returns, suggests that the market is slow to react to the information embedded in narratives.

Table 12: **Sentiment and Portfolio Performance.** We use two proxies to measure investor sentiment: BW, generated using the PCA method by Baker and Wurgler (2006), and HJTZ, generated using the PLS method by Huang, Jiang, Tu, and Zhou (2015), respectively. All sentiment indexes are smoothed using a 12-month moving average and then standardized. Different cross sectional factors are used as control variable. Reported are the regression coefficient and Newey-West t-value. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period spans from January 2001 to December 2020.

$R_t^{NMP} = \beta_0 + \beta_1 Sentiment_{t-1} + \boldsymbol{\beta} F_t + \varepsilon_t$						
	F:	CAPM	FF3	Carhart	FF5	HXZ
BW	β_1	0.14	0.15	0.07	-0.00	0.10
	t_{NW}	(1.2)	(1.32)	(0.63)	(-0.03)	(0.80)
HJTZ	β_1	0.26**	0.27***	0.24**	0.13	0.29***
	t_{NW}	(2.51)	(2.68)	(2.03)	(0.96)	(2.60)

5. Robustness Test

In this section, we conduct several robustness checks to reconfirm our main results. These include examining the relationship between sentiment and the performance of the narrative-managed portfolio, the relationship between labor market conditions and the portfolio's performance, and the effects of leverage constraints and transaction costs on portfolio performance.

5.1 Sentiment and Portfolio Performance

In the previous section, we find that the predictive power of narratives stems from investors' cognitive biases, which are more prevalent among sentiment-driven investors. Sentiment-driven investors' reluctance to take short positions in low-sentiment periods leads them to exert greater influence during high-sentiment periods than in low-sentiment periods (Yu and Yuan, 2011; Stambaugh, Yu, and Yuan, 2012). Therefore, the performance of the narrative-managed portfolio is expected to be better during high-sentiment periods.

Table 8 reports the results. For the BW sentiment index, the estimated slopes are not significantly different from zero, which is not surprising, as Baker and Wurgler (2006)

construct their sentiment index using the first principal component, which may contain information unrelated to stock returns. In contrast, Huang et al. (2015) construct their sentiment index using the PLS method developed by Wold (1966); Kelly and Pruitt (2015), which excludes information unrelated to stock returns. For the HJTZ sentiment index, all slopes are positive and significant at the 5% level (except for FF5). This result supports the notion that the predictive power of narratives stems from investors' cognitive biases, specifically their underreaction to disaster-related information.

5.2 Labor Market Conditions and Portfolio Performance

The topic importance analysis in Figure 6 suggests that labor market conditions may be a key driver of the outperformance of narratives. Therefore, we further examine the relationship between the narrative-managed portfolio and labor market conditions. Inspired by Huang, Jiang, Li, Tong, and Zhou (2023), we plot the narrative-managed portfolio returns and market conditions. Proxies for labor market conditions are the CFNAI sub-index on employment and hours(Employment) and production and income(Income), as well as the Wage Growth Tracker from the Atlanta Fed. To enhance readability, the returns and proxies for labor market conditions are all smoothed using a twelve month moving average. Figure 9 shows the results. We find that narrative-managed portfolio returns are significantly negative in relation to all three labor variables. For example, the correlation between the Wage Growth Tracker and portfolio returns is -0.33. Notably, during the 2008 financial crisis, while the labor market deteriorated, the managed portfolio generated substantial gains. These results further confirm that narratives can help mitigate tail risks during extreme economic downturns

5.3 Alphas under Different Leverage Constraints

In the previous section, the portfolio weight, w_t, was restricted to the range [0, 1]. In this section, we examine portfolios constructed without leverage constraints and with leverage

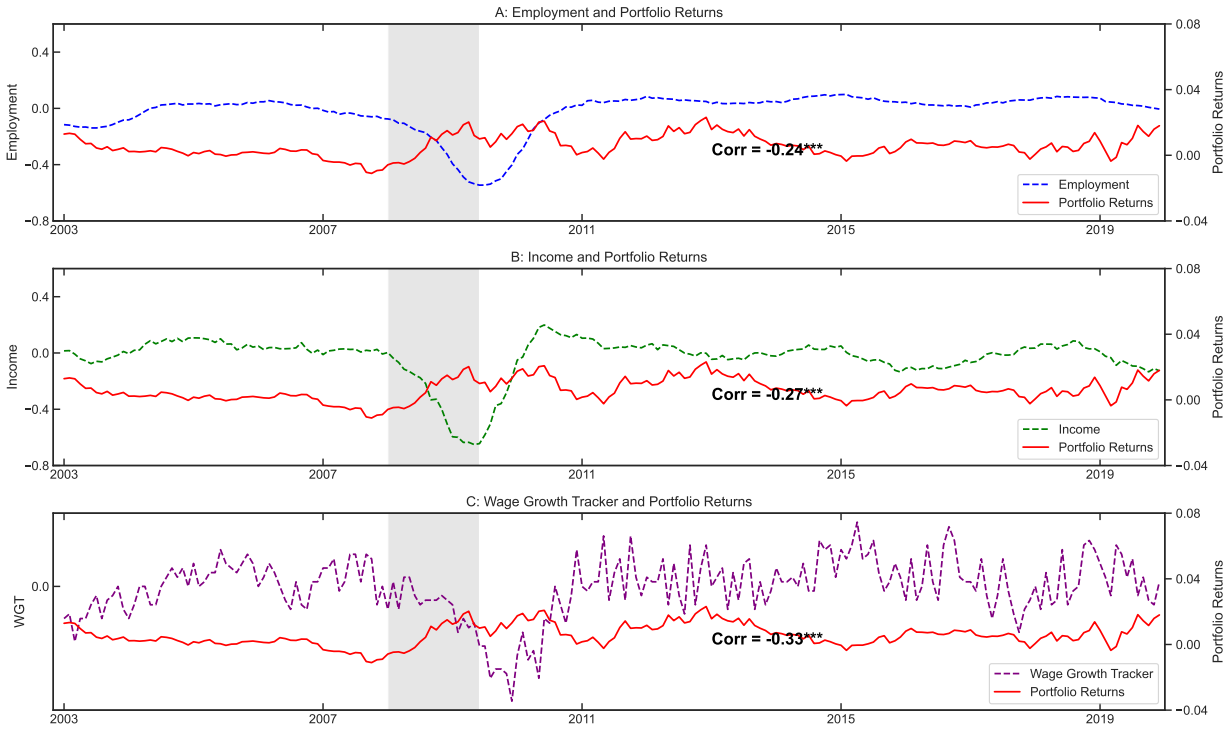


Fig. 9. **Labor Market Conditions and Portfolio Performance.** This figure plots the time series of labor market conditions and the corresponding R_t^{NMP} returns over the next year. Proxies for labor market conditions are the CFNAI sub-index on employment and hours(Employment) and production and income(Income), as well as the Wage Growth Tracker from the Atlanta Fed. Both returns and labor market proxies are smoothed using a twelve-month moving average. The shaded bars represent NBER-dated recession periods.

constraints of [0, 1.5].

Table 13 reports the alphas of narrative-managed portfolios under different leverage constraints. The following conclusions emerge from the table. Compared to the [0, 1] leverage constraint, both the [0, 1.5] leverage setting and the unconstrained specification generate higher average returns. For instance, the annualized average returns exceed 10% when no leverage constraints are applied. Furthermore, under all alternative leverage settings, the alphas of the narrative-managed portfolio remain positive across all factor models. These findings suggest that leverage constraints are unlikely to explain the significantly high alphas of the narrative-managed portfolio.

Table 13: **Alphas of Narrative Managed Portfolios under Different Leverage Constraints.** This table reports alphas of narrative managed portfolios based on combination method when imposing different leverage constraints. The leverage constraints specified in the panel headings adjust weights that exceed the presented interval to align with the boundary of the interval they intersect.

	Average return	α^{CAPM}	α^{FF3}	$\alpha^{Carhart}$	α^{FF5}	α^{HXZ}
No Leverage Constraints	10.12***	7.66**	7.07***	6.29**	4.31**	7.57**
	(4.23)	(2.51)	(2.59)	(2.29)	(2.03)	(2.08)
[0,1.5] Leverage Constraints	8.13***	4.07***	3.86***	3.4**	2.51**	3.67**
	(3.77)	(2.66)	(2.7)	(2.4)	(2.18)	(2.26)
[0,1] Leverage Constraints	7.49***	3.61***	3.43***	3.02**	2.31**	3.28**
	(3.93)	(2.62)	(2.67)	(2.34)	(2.21)	(2.23)

5.4 Alphas Adjusted for Transaction Costs

In this subsection, we further demonstrate that the narrative-managed portfolio remains profitable after accounting for transaction costs. Following Moreira and Muir (2017) and Hollstein and Prokopczuk (2022), we set a proportional transaction cost of 10 basis points (bps). This is relatively conservative, given the high liquidity of S&P 500 ETFs. Table 14 reports the results. As expected, all alphas are smaller compared to the results in Table 5, which do not incorporate transaction costs. However, the alphas generated by narrative and machine learning methods, particularly the COMB, remain positive and significant at the 5% level. Overall, these results support the conclusion that narratives help investors construct a profitable portfolio.

5.5 Alternative Topic Model

In this subsection, we also apply another topic model, the LSI model, to estimate news topics. The LSI (Latent Semantic Indexing) model, proposed by Deerwester, Dumais, Furnas, Landauer, and Harshman (1990), uses singular value decomposition (SVD) to reduce the dimensionality of the document-term matrix, allowing for the identification of underlying topics or concepts. The LSI model has also been applied in economic studies (Ter Ellen, Larsen, and Thorsrud, 2021). We set the number of topics in the LSI model to 180, consis-

Table 14: **Alphas of Narrative Managed Portfolios Adjusted for Transaction Costs.** This table reports alphas of narrative managed portfolios when considering transaction costs of 10bps.

	Average return	α^{CAPM}	α^{FF3}	$\alpha^{Carhart}$	α^{FF5}	α^{HXZ}
ENET	6.71***	3.34**	2.75**	2.43*	1.72	3.15*
	(3.17)	(2.24)	(2.03)	(1.79)	(1.48)	(1.65)
CSR	4.79***	2.16**	2.08**	1.8*	1.17	2.03*
	(3.17)	(2.05)	(2.04)	(1.78)	(1.22)	(1.67)
PCA	8.4***	2.43*	2.4*	2.03*	1.10	2.41*
	(3.22)	(1.85)	(1.93)	(1.72)	(1.16)	(1.79)
PLS	7.9***	3.06**	2.9**	2.49**	1.58	3.17*
	(3.44)	(2.18)	(2.19)	(1.96)	(1.49)	(1.9)
RF	4.52***	1.88*	1.71*	1.41	1.39	1.63
	(2.98)	(1.81)	(1.73)	(1.38)	(1.61)	(1.43)
GBDT	4.56***	1.59	1.42	1.3	1.09	2.20
	(2.61)	(1.33)	(1.3)	(1.16)	(1.13)	(1.53)
COMB	7.35***	3.47**	3.29**	2.87**	2.16**	3.13**
	(3.86)	(2.52)	(2.56)	(2.23)	(2.07)	(2.13)

tent with the LDA model. Table 15 reports the out-of-sample performance of news topics estimated by the LSI model. These results align with the findings presented in the main context. Thus, the forecasting power of narratives is not confined to the specific choice of the LDA model for estimating news topics.

6. Conclusion

This paper examines the value of narratives in managing the market portfolio. By employing a voluminous news corpus and machine learning models, we construct the narrative-managed portfolio. We show that investors can achieve significant utility gains by holding the narrative-managed portfolio instead of the market portfolio. For the same initial investment of \$1 in January 2002, the cumulative wealth of the narrative-managed market portfolio is twice that of the market portfolio by December 2020. Moreover, we demonstrate that narratives offer incremental predictive power for the market equity risk premium, even after accounting for economic and financial predictors.

Table 15: **Out-of-Sample Performance of Latent Semantic Indexing (LSI)**. This table reports out-of-sample performance of forecasts based on the Latent Semantic Indexing (LSI) topic model. The forecast models includes kitchen sink (KS), elastic net (ENET), complete subset regression (CSR), principal component analysis (PCA), partial least square (PLS), random forest(RF), gradient-boosted decision trees (GBDT) and forecast combination (COMB). R^2_{OS} is the Campbell and Thompson (2008) out-of-sample R^2 statistic. $R^2_{OS,up}$ ($R^2_{OS,down}$) statistics are calculated over NBER business-cycle expansions (recessions). CW-t is the Clark and West (2007) statistic. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively. Out-of-sample period is from January 2001 to December 2020.

Model	$R^2_{OS}(\%)$	CW-t	$R^2_{OS,down}(\%)$	$R^2_{OS,up}(\%)$
KS	-389.34	1.03	-309.16	-430.65
ENET	2.33***	2.87	6.54	0.16
CSR	1.62*	1.43	3.83	0.47
PCA	1.55*	1.55	7.51	-1.53
PLS	1.14*	1.49	1.90	0.75
RF	2.95**	1.70	6.42	1.16
GBDT	2.69**	1.68	7.05	0.44
COMB	2.94**	1.81	7.48	0.60

We further explore the underlying sources of this predictive power. Narratives are shown to contain disaster risk-related information and exhibit strong predictive power for future disaster risk and several discount rate-related variables. We also demonstrate that investors have sticky beliefs and underreact to the new information contained in narratives. Additionally, after considering leverage constraints, transaction costs, and alternative topic models, the narrative-managed portfolio's performance remains superior.

In summary, we present a comprehensive demonstration of the investment value of narratives in timing the U.S. stock market. The narrative-based strategy yields significant economic gains for investors.

References

Adcock, Christopher J, 2010, Asset pricing and portfolio selection based on the multivariate extended skew-student-t distribution, *Annals of Operations Research* 176, 221–234.

Arrow, Kenneth Joseph, et al., 1974, *Essays in the theory of risk-bearing*, volume 121 (North-Holland Amsterdam).

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, 2016, Measuring economic policy uncertainty, *Quarterly Journal of Economics* 131, 1593–1636.

Bali, Turan G, Amit Goyal, Dashan Huang, Fuwei Jiang, and Quan Wen, 2020, Predicting corporate bond returns: Merton meets machine learning, *Georgetown McDonough School of Business Research Paper* 20–110.

Barillas, Francisco, Raymond Kan, Cesare Robotti, and Jay Shanken, 2020, Model comparison with sharpe ratios, *Journal of Financial and Quantitative Analysis* 55, 1840–1874.

Baron, Matthew, Wei Xiong, and Zhijiang Ye, 2022, Measuring time-varying disaster risk: An empirical analysis of dark matter in asset prices, Technical report, Mimeo, Princeton University.

Barro, Robert J, 2006, Rare disasters and asset markets in the twentieth century, *Quarterly Journal of Economics* 121, 823–866.

Barro, Robert J, and Gordon Y Liao, 2021, Rare disaster probability and options pricing, *Journal of Financial Economics* 139, 750–769.

Bekaert, Geert, Eric C Engstrom, and Nancy R Xu, 2022, The time variation in risk appetite and uncertainty, *Management Science* 68, 3975–4004.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch, 2014, Labor hiring, investment, and stock return predictability in the cross section, *Journal of Political Economy* 122, 129–177.

Blei, David M, Andrew Y Ng, and Michael I Jordan, 2003, Latent dirichlet allocation, *Journal of Machine Learning Research* 3, 993–1022.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar, 2019, Sticky expectations and the profitability anomaly, *Journal of Finance* 74, 639–674.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.

Boyer, Brian H, and Keith Vorkink, 2014, Stock options as lotteries, *Journal of Finance* 69, 1485–1527.

Breen, William, Lawrence R Glosten, and Ravi Jagannathan, 1989, Economic significance of predictable variations in stock index returns, *Journal of Finance* 44, 1177–1189.

Breiman, Leo, 2001, Random forests, *Machine Learning* 45, 5–32.

Bybee, Leland, Bryan Kelly, Asaf Manela, and Dacheng Xiu, 2021, Business news and business cycles, *Journal of Finance* .

Caldara, Dario, and Matteo Iacoviello, 2022, Measuring geopolitical risk, *American Economic Review* 112, 1194–1225.

Calomiris, Charles W, and Harry Mamaysky, 2019, How news and its context drive risk and returns around the world, *Journal of Financial Economics* 133, 299–336.

Campbell, John Y, 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Campbell, John Y, and Robert J Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y, and Robert J Shiller, 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cederburg, Scott, Michael S O'Doherty, Feifei Wang, and Xuemin Sterling Yan, 2020, On the performance of volatility-managed portfolios, *Journal of Financial Economics* 138, 95–117.

Clark, Todd E, and Kenneth D West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of econometrics* 138, 291–311.

Cochrane, John H, 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Cochrane, John H, 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2020, Lazy prices, *Journal of Finance* 75, 1371–1415.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644–78.

De Athayde, Gustavo M, and Renato G Flôres Jr, 2004, Finding a maximum skewness portfolio—a general solution to three-moments portfolio choice, *Journal of Economic Dynamics and Control* 28, 1335–1352.

De Roon, Frans A, Theo E Nijman, and Bas JM Werker, 2001, Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets, *Journal of Finance* 56, 721–742.

Deerwester, Scott, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman, 1990, Indexing by latent semantic analysis, *Journal of the Association for Information Science* 41, 391–407.

Dong, Xi, Yan Li, David E Rapach, and Guofu Zhou, 2022, Anomalies and the expected market return, *Journal of Finance* 77, 639–681.

Elliott, Graham, Antonio Gargano, and Allan Timmermann, 2013, Complete subset regressions, *Journal of Econometrics* 177, 357–373.

Fama, Eugene F, and Kenneth R French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Fama, Eugene F, and Kenneth R French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and G William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Farmer, Leland, Lawrence Schmidt, and Allan Timmermann, 2022, Pockets of predictability, *Journal of Finance, forthcoming* .

Favilukis, Jack, and Xiaoji Lin, 2016, Does wage rigidity make firms riskier? evidence from long-horizon return predictability, *Journal of Monetary Economics* 78, 80–95.

Ferreira, Thiago RT, 2024, Cross-sectional financial conditions, business cycles and the lending channel, *Journal of Monetary Economics* 103597.

Freund, Yoav, 1995, Boosting a weak learning algorithm by majority, *Information and Computation* 121, 256–285.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, 2000, Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors), *Annals of Statistics* 28, 337–407.

Friedman, Jerome H, 2001, Greedy function approximation: a gradient boosting machine, *Annals of Statistics* 1189–1232.

Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–74.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2005, There is a risk-return trade-off after all, *Journal of Financial Economics* 76, 509–548.

Goetzmann, William N, Dasol Kim, and Robert J Shiller, 2022, Crash narratives, Technical report, National Bureau of Economic Research.

Gómez-Cram, Roberto, 2022, Late to recessions: Stocks and the business cycle, *Journal of Finance* 77, 923–966.

Gourio, Francois, 2013, Credit risk and disaster risk, *American Economic Journal: Macroeconomics* 5, 1–34.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Guo, Hui, 2006, On the out-of-sample predictability of stock market returns, *Journal of Business* 79, 645–670.

Harvey, Campbell R, John C Liechty, Merrill W Liechty, and Peter Müller, 2010, Portfolio selection with higher moments, *Quantitative Finance* 10, 469–485.

Harvey, David I, Stephen J Leybourne, and Paul Newbold, 1998, Tests for forecast encompassing, *Journal of Business & Economic Statistics* 16, 254–259.

Hirshleifer, David, Dat Mai, and Kuntara Pukthuanthong, 2023, War discourse and the cross section of expected stock returns, Technical report, National Bureau of Economic Research.

Hirshleifer, David, Dat Mai, and Kuntara Pukthuanthong, 2024, War discourse and disaster premium: 160 years of evidence from the stock market, *The Review of Financial Studies* hhae081.

Hirshleifer, David, and Jinfei Sheng, 2022, Macro news and micro news: complements or substitutes?, *Journal of Financial Economics* 145, 1006–1024.

Hollstein, Fabian, and Marcel Prokopczuk, 2022, Managing the market portfolio, *Management Science* .

Hong, Yongmiao, Fuwei Jiang, Lingchao Meng, and Bowen Xue, 2022, Forecasting inflation with economic narratives and machine learning, *Available at SSRN 4175749* .

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An augmented q-factor model with expected growth, *Review of Finance* 25, 1–41.

Huang, Dashan, Fuwei Jiang, Kunpeng Li, Guoshi Tong, and Guofu Zhou, 2023, Are bond returns predictable with real-time macro data?, *Journal of Econometrics* .

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *Review of Financial Studies* 28, 791–837.

Huang, Dashan, Jiangyuan Li, and Liyao Wang, 2021, Are disagreements agreeable? evidence from information aggregation, *Journal of Financial Economics* 141, 83–101.

Huang, Dashan, Jiangyuan Li, Liyao Wang, and Guofu Zhou, 2020, Time series momentum: Is it there?, *Journal of Financial Economics* 135, 774–794.

Huberman, Gur, and Shmuel Kandel, 1987, Mean-variance spanning, *Journal of Finance* 42, 873–888.

Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou, 2019, Manager sentiment and stock returns, *Journal of Financial Economics* 132, 126–149.

Kan, Raymond, and GuoFu Zhou, 2008, Tests of mean-variance spanning, in *AFA 2001 New Orleans Meetings, OLIN Working Paper*, number 99-05.

Keim, Donald B, and Robert F Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kelly, Bryan, and Seth Pruitt, 2015, The three-pass regression filter: A new approach to forecasting using many predictors, *Journal of Econometrics* 186, 294–316.

Korkie, Bob, and Harry J Turtle, 2002, A mean-variance analysis of self-financing portfolios, *Management Science* 48, 427–443.

Lahrech, Abdelmounaim, and Kevin Sylwester, 2011, Us and latin american stock market linkages, *Journal of International Money and Finance* 30, 1341–1357.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.

Manela, Asaf, and Alan Moreira, 2017, News implied volatility and disaster concerns, *Journal of Financial Economics* 123, 137–162.

Mankiw, N Gregory, and Ricardo Reis, 2002, Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve, *Quarterly Journal of Economics* 117, 1295–1328.

Markowitz, Harry, 1952a, Portfolio selection, *Journal of Finance* 7, 77–91.

Markowitz, Harry, 1952b, The utility of wealth, *Journal of Political Economy* 60, 151–158.

Mitton, Todd, and Keith Vorkink, 2007, Equilibrium underdiversification and the preference for skewness, *Review of Financial Studies* 20, 1255–1288.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Nelson, Charles R, and Myung J Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.

Newey, Whitney K, and Kenneth D West, 1986, A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix .

Rapach, David, and Guofu Zhou, 2013, Forecasting stock returns, in *Handbook of Economic Forecasting*, volume 2, 328–383 (Elsevier).

Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Rapach, David E, Jack K Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Rapach, David E, and Guofu Zhou, 2020, Time-series and cross-sectional stock return forecasting: New machine learning methods, *Machine Learning for Asset Management: New Developments and Financial Applications* 1–33.

Rietz, Thomas A, 1988, The equity risk premium a solution, *Journal of Monetary Economics* 22, 117–131.

Satchell, Stephen, 2003, *Advances in portfolio construction and implementation* (Elsevier).

Schapire, Robert E, 1990, The strength of weak learnability, *Machine Learning* 5, 197–227.

Shiller, Robert J., 2017, Narrative economics, *American Economic Review* 107, 967–1004.

Sims, Christopher A, 1980, Macroeconomics and reality, *Econometrica: Journal of the Econometric Society* 1–48.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Stock, James H, and Mark W Watson, 2002, Forecasting using principal components from a large number of predictors, *Journal of American Statistical Association* 97, 1167–1179.

Ter Ellen, Saskia, Vegard H. Larsen, and Leif Anders Thorsrud, 2021, Narrative monetary policy surprises and the media, *Journal of Money, Credit and Banking* .

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

Tsai, Jerry, 2015, Rare disasters and the term structure of interest rates, *Available at SSRN 2669992* .

Tsai, Jerry, and Jessica A Wachter, 2016, Rare booms and disasters in a multisector endowment economy, *Review of Financial Studies* 29, 1113–1169.

Veldkamp, Laura L, 2023, *Information choice in macroeconomics and finance* (Princeton University Press).

Wachter, Jessica A, 2013, Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* 68, 987–1035.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Wold, Herman, 1966, Estimation of principal components and related models by iterative least squares, *Multivariate analysis* 391–420.

Yu, Jianfeng, and Yu Yuan, 2011, Investor sentiment and the mean–variance relation, *Journal of Financial Economics* 100, 367–381.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: series B* 67, 301–320.

Appendix A.

Table A1: **Topic Key Terms**. This table lists top 10 key terms for each topic. We manually assign a label to each topic based on our reading of the key term lists.

Topic number	Topic label	Key terms
1	Taxes	tax, income, taxe, rate, tax_cut, tax_rate, cut, deduction, corporate, return
2	List	list, listie, task, listing, national, task_force, force, gray, basket, symbol
3	Company	company, big, customer, start, firm, work, busines, spend, executive, offer
4	Fast Food	food, restaurant, brand, coffee, wine, bar, bag, eat, fashion, dog
5	US States	rock, bondholder, twin, mississippi, indiana, marriott, tennessee, flagship, arkansas, theme
6	Physical Examination	commentary, brain, peer, pain, analysis, risk, moore, feedback, physical, breast
7	Blue Chip	index, blue, session, lake, cup, salt, utah, blue_chip, nonstop, chip
8	Balance Sheet	accountie, balance, sheet, guidance, subway, transit, balance_sheet, rider, financial, rule
9	Investigation	cloud, scandal, probe, suspend, trade_tension, investigation, allegation, theft, scheme, fallout
10	*Others*[10]	smith, cigarette, prescription, pharmacy, tier, smokie, bonuses, prescription_drug, merck, roger
11	Natural Resource	natural, resource, florida, shale, sand, cameron, rig, field, paso, natural_resource
12	Business	busines, business, small, owner, corporate, big, large, corporate_tax, corporation, smaller
13	Share	share, cent, earlier, rose, cent_share, net, earning, compar, period, income
14	American	american, brazil, expres, brazilian, america, thomson, east_side, rais_concern, collar, american_company
15	Bankruptcy	bankruptcy, chapter, filie, creditor, protection, court, fil, restructurie, company, fil_bankruptcy
16	Bank	bank, financial, bankie, asset, capital, fargo, bank_america, chase, banker, big

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
17	German	german, germany, deutsche, deutsche_bank, berlin, beverage, drink, cola, coca, coca_cola
18	Books	book, story, sexual, page, post, author, write, read, wrote, search
19	Sales	sale, sold, sellie, sale_growth, sotheby, store_sale, total, pickup, sale_rose, rebate
20	Loss	los, mueller, loss, net, result, charge, post, company, cost, relat
21	Company Sales	company, market, product, consumer, brand, industry, competition, busines, sale, maker
22	Advertising	advertisie, carbon, eric, van, agency, advertiser, organic, fannie, jon, madison
23	Securities Transaction	security, client, broker, brokerage, firm, lynch, merrill, steele, financial, account
24	Department	department, account, transfer, application, municipality, labor_department, vest, dakota, claim, department_official
25	Photos	brooklyn, photo, camera, image, color, portrait, picture, pet, cat, oscar
26	European	european, europe, french, france, pari, ita, union, country, germany, italian
27	Microchips	chip, valley, silicon, silicon_valley, electronic, intel, maker, semiconductor, memory, component
28	Power	power, project, plant, infrastructure, utility, cost, construction, nuclear, electricity, wind
29	*Others*[29]	letter, relat, oak, interest_taxe, brook, tribune, net_worth, barne, worth, milwaukee
30	*Others*[30]	online, aid, vietnam, updat, clinical_trial, soccer, rabbi, censor, patriot, margaret
31	*Others*[31]	sear, dean, audio, merchant, china_sea, south_china, redesign, advancie, friedman, downside
32	Loan	loan, credit, lender, lendie, bank, borrower, credit_card, risk, crisi, finance
33	Entertainment	film, movie, fox, studio, theater, star, television, audience, channel, production
34	Systems	system, control, server, abu, gaza, monitor, dual, bug, install, secure
35	*Others*[35]	clinton, ton, metric, mountain, metric_ton, hungarian, hungary, profit_revenue, view, syracuse
36	California	california, san, francisco, san_francisco, angele, calif, bay, area, kansa, southern
37	Disaster	water, area, storm, river, flood, local, town, mile, land, resident

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
38	US Senate	house, bill, senate, congres, republican, democrat, lawmaker, president, committee, budget
39	Financial Oversight	oversight, brother, mccarthy, lehman, deloitte, bond_investor, loeb, volcker, longest, aggregate
40	Stock Markets	market, investor, stock, buy, high, price, financial, gain, analyst, sector
41	Earnings Forecasts	analyst, earning, estimate, company, forecast, outbreak, footage, snack, expectation, projection
42	Money Market	money, fiat, yield, average, virginia, money_market, earlier, market, eligibility, west_virginia
43	Apps	app, user, digital, software, device, computer, microsoft, machine, technology, screen
44	Commodity Price	price, demand, commodity, supp, silver, exchange_trad, market, global, future, copper
45	Oil Industry	oil, company, energy, fuel, oil_company, exxon, industry, gasoline, production, field
46	Country	country, global, mexico, india, world, border, international, emergie, climate, mexican
47	Energy	energy, pipeline, fort, william, pipe, reid, natural, transmission, charleston, energy_price
48	*Others*[48]	good, work, world, problem, idea, great, long, real, reason, important
49	Spanish	spanish, world_biggest, santa, spain, san, calif, diego, san_diego, jose, maria
50	Law	lawsuit, claim, settlement, fil, case, court, complaint, lawyer, suit, legal
51	Arts	work, art, museum, song, night, start, long, piece, room, life
52	Middle East	syria, israel, turkey, state, islamic, israeli, syrian, middle, east, middle_east
53	Long/Short Term	term, long, short, long_term, short_term, elliott, galaxy, analyst_poll, longer, policy_maker
54	Price	price, cost, higher, increase, pricie, average, discount, cut, expensive, high
55	Retail	store, retailer, chain, retail, customer, shopper, shoppie, item, open, grocery
56	Food	food, farmer, farm, crop, meat, animal, agriculture, agricultural, corn, plant
57	Securities	due, note, secur, moody, senior, sept, investor_service, moody_investor, notch, convertible
58	Apple	apple, model, pan, accessory, franklin, mac, schwartz, steve, copyright, blockbuster

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
59	Material	material, chemical, production, aluminum, facility, ireland, plastic, manufacturie, glas, metal
60	Donald Trump	trump, donald, bad, interest, taylor, loan, roman, total, loss, portfolio
61	Election	republican, campaign, election, democrat, democratic, voter, party, political, candidate, vote
62	Agency	agency, rule, official, safety, ukraine, federal, review, problem, concern, administration
63	Policy	policy, american, america, political, public, freedom, donor, social, editor, catholic
64	Bear Market	fell, share, stock, rose, tradie, nasdaq, share_fell, market, gain, company
65	Police	police, kill, officer, attack, war, gun, death, fire, violence, shoot
66	*Others*[66]	walgreen, cole, cigna, workout, keith, busines_unit, slowie_growth, wilbur, drugstore, asset_sale
67	Regulatory	firm, law, sec, security, investigation, letter_letter, document, information, disclosure, probe
68	Travels	travel, trip, ticket, christie, fare, aviation, coach, sky, destination, lewi
69	Children	children, parent, game, child, kid, adam, toy, boy, baby, jonathan
70	*Others*[70]	award, greece, narrative, greek, damage, breach, graham, shield, polish, romantic
71	Fund	fund, investor, asset, investment, money, manager, invest, return, hedge, portfolio
72	Futures	tradie, trader, future, trade, chicago, market, epstein, thompson, hedge, exchange
73	Electricity	electric, general, derivative, conn, malaysia, general_electric, manual, xerox, list_price, malaysian
74	Conferences	question, public, told, conference, speech, event, message, activist, meet, statement
75	Judge	judge, justice, department, justice_department, federal, attorney, antitrust, hearie, general, attorney_general
76	Party	party, political, election, government, leader, country, protest, vote, opposition, muslim
77	Interest Rate	rate, inflation, interest, interest_rate, policy, central, central_bank, reserve, bank, federal
78	*Others*[78]	cohen, gold, charter, minie, airbus, fla, seattle, oklahoma, orlando, minie_company
79	Korea	north, korea, south, north_korea, korean, kim, south_korea, north_korean, pyongyang, north_america

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
80	Arabia	iran, saudi, sanction, official, iranian, iraq, arabia, saudi_arabia, state, islamic
81	*Others*[81]	martin, boat, executive_order, swift, heritage, unilever, dock, deal_makie, territorial, scalia
82	Economic Growth	growth, economy, economic, spend, recession, economic_growth, slow, economist, deficit, global
83	Russia	russian, official, moscow, meet, state, foreign, country, summit, leader, western
84	Debt	debt, payment, cash, financial, default, pay, borrow, flow, interest, cash_flow
85	Data	data, information, collect, acces, computie, bas, track, sharie, provider, center
86	*Others*[86]	operation, core, battery, exit, division, wolf, spinoff, tire, focus, shedd
87	Media	media, publish, publisher, chri, newspaper, magazine, ros, print, dai, post
88	Comments	comment, declin, declin_comment, spokesman, request, respond, reach, representative, reach_comment, official
89	*Others*[89]	alliance, stewart, usa, arthur, mason, craig, fluid, gannett, seek_comment, weis
90	Profits	profit, company, growth, analyst, margin, earning, busines, result, cost, operatie
91	Health	health, care, health_care, medicare, cost, medicaid, medical, service, health_insurance, program
92	Disney	site, disney, content, environmental, waste, epa, walt, walt_disney, environmental_protection, protection
93	Exchange	exchange, stock, stock_exchange, tradie, share, york, york_stock, filie, clos, drug_company
94	Real Estate	estate, real, property, real_estate, developer, market, office, trust, commercial, tower
95	*Others*[95]	church, religious, christian, faith, dougla, philippine, cruise, vatican, god, hughes
96	Court	court, law, supreme, supreme_court, case, decision, justice, appeal, rulie, legal
97	World	world, wide, equipment, world_wide, world_largest, largest, maker, clark, anderson, caterpillar
98	Currency	currency, foreign, dollar, exchange, trade_war, qatar, central, foreign_exchange, central_bank, outflow
99	Government in Crisis	government, crisi, country, official, economic, sector, economy, finance, fund, financial
100	UK Politics	minister, prime, prime_minister, perry, parliament, portland, cabinet, supportive, conservative_party, cox

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
101	Patients	patient, hospital, doctor, medical, ice, mental, physician, surgery, treatment, medicine
102	Presidents	president, candidate, nominee, poll, warren, nomination, presidential, vice, vice_president, convention
103	Life	wireles, australia, gate, australian, tablet, milk, antonio, dairy, victoria, cow
104	M&A	deal, company, acquisition, buy, merger, transaction, share, bas, shareholder, cash
105	Patent	patent, kel, pair, indonesia, shoe, cambridge, monsanto, royalty, consolidat, simplify
106	Pharma	allen, pfizer, pittsburgh, penn, philadelphia, bankruptcy_court, henry, freeze, earning_interest, rex
107	Crime	investigation, prosecutor, criminal, charge, case, attorney, fbi, arrest, official, crime
108	Traffic	seat, traffic, mile, fly, southwest, route, corridor, load, horizon, passenger
109	Subsidiary	unit, parent, parent_company, unit_state, subsidiary, busines, affiliate, metro, franchise, bas
110	Treasury Department	financial, treasury, account, deposit, saving, federal, bill, number, money, treasury_department
111	Latin America	america, venezuela, latin, venezuelan, argentina, latin_america, region, colombia, spokeswoman_declin, country_includie
112	Environment	cap, emission, clean, air, earth, statue, environmental, small, pollution, shower
113	Holiday	holiday, season, hudson, christmas, mart, wal, wal_mart, macy, target, penney
114	Housing Market	home, house, housie, bedroom, agent, market, homeowner, buyer, family, mortgage
115	Automotive	program, financial_analysis, carolina, michigan, car_maker, bronx, georgia, north_carolina, walker, ohio
116	Airlines	airline, flight, airport, air, plane, pilot, passenger, carrier, traveler, delta
117	South Africa/Sweden	south, african, africa, chamber, coin, diamond, sullivan, swedish, south_africa, sweden
118	Property Transaction	buyer, owner, property, sold, land, buy, sell, mall, bought, auction
119	Labor	social, job, labor, wage, worker, unemployment, population, income, rate, employment
120	Executive	executive, chief, chief_executive, officer, company, chairman, financial, president, vice, top
121	*Others*[121]	respond_request, scott, shipp, coal, wang, miner, cargo, yard, transport, trailer

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
122	China	china, chinese, hong, kong, hong_kong, asia, shanghai, asian, singapore, global
123	Credit Ratings	standard, rating, ratie, rat, credit, grade, doug, poor, triple, rand
124	Shipping	platform, ship, sea, drill, ocean, island, vessel, coast, offshore, water
125	Service	service, provider, financial_service, tobacco, ramp, service_company, russell, knight, pregnant, service_firm
126	*Others*[126]	johnson, statement, audit, auditor, warmie, johnson_johnson, rodriguez, wheeler, financial_statement, financial
127	*Others*[127]	door, ben, ukrainian, camp, hailie, tunnel, chinese_government, open_door, receipt, detain
128	States	state, law, federal, county, governor, gov, immigration, york, local, jersey
129	Russia Leader	president, russia, support, leader, political, power, policy, country, fight, force
130	Media Industry	mobile, fiscal, radio, media_company, revenue_growth, frequency, harrison, cellular, recording, current_fiscal
131	Organizations	member, organization, foundation, nonprofit, money, fee, district, public, gift, association
132	Treasury Bond	bond, yield, treasury, corporate, bond_yield, bond_market, government_bond, investor, corporate_bond, market
133	NY city	city, york, york_city, hotel, mayor, neighborhood, manhattan, area, resident, local
134	Technology	technology, company, tech, venture, research, development, founder, develop, innovation, work
135	Employment	job, employee, work, company, worker, cut, factory, cost, staff, hirie
136	IPO	public, goldman, offerie, initial, sach, goldman_sach, initial_public, public_offerie, offering, bear
137	Contract	contract, contractor, grid, supplier, helicopter, aerospace, radar, generator, defense, lockhe
138	Japan	japan, japanese, nissan, tokyo, volkswagen, yen, toyota, honda, dealership, march
139	School	school, student, university, college, education, teacher, professor, high, high_school, graduate
140	Research	study, research, researcher, cancer, disease, university, test, health, risk, scientist
141	Labor Market	pay, benefit, retirement, paid, compensation, cost, employee, pension, employer, earn
142	Survey	survey, universal, conduct, borough, respondent, busines_school, percent, grace, service_provider, poll

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
143	Space Program	space, solar, panel, satellite, sensor, launch, rocket, moon, bayer, station
144	Wood	paper, wood, crown, forest, stone, kraft, mill, columbia, lauren, pharmaceutical_company
145	*Others*[145]	london, british, plc, britain, telecom, rio, zealand, london_bas, european_parliament, anglo
146	Regulator	commission, regulator, regulatory, regulation, proposal, propos, approval, commissioner, rule, regulat
147	Beers	brown, beer, miller, sport, super, lincoln, prize, comic, spirit, morri
148	MENA Geopolitics	kie, prince, jew, egypt, chinese_company, mohamm, jordan, jerusalem, emirate, egyptian
149	Private	private, equity, public, buyout, firm, leverag, financie, investor, deal, clayton
150	Rental Properties	apartment, rent, park, rental, beach, england, landlord, condo, tenant, documentary
151	Fiat Money	dollar, swis, european_central, switzerland, currency, franc, pound, sterl, unwind, weaker
152	National Security	intelligence, security, government, official, letter_editor, national, protester, national_security, attack, agency
153	Card	card, island, visa, long_island, align, lion, campus, identity, tourism, extradition
154	Investment	firm, investment, capital, management, morgan, global, partner, stanley, managie, client
155	Music	music, video, artist, record, sony, sound, recordie, display, camera, hill
156	Oil Market	oil, barrel, crude, production, price, oil_price, hurricane, output, export, vladimir
157	Car	car, vehicle, auto, maker, driver, ford, truck, motor, auto_maker, drivie
158	Insurance	insurance, insurer, coverage, premium, policy, life, cover, risk, insurance_company, financial
159	Mortgage	rate, mortgage, ryan, interest_rate, interest, high, junk, fix, home_price, high_yield
160	Family Members	family, women, black, men, young, son, white, live, life, woman
161	Canada	canada, train, canadian, transportation, rail, railroad, murphy, free_trade, track, pacific
162	US Leader	president, white, house, white_house, administration, official, office, adviser, secretary, team
163	Takeover	offer, bid, takeover, tender, jacob, reject, share, company_offer, bidding, noon

Continued on next page

Table A1 – continued from previous page

Topic number	Topic label	Key terms
164	Drug	drug, pharmaceutical, fda, treatment, patient, puerto, trial, rico, medicine, puerto_rico
165	Company Board	board, shareholder, director, member, vote, company, meet, chairman, proxy, corporate
166	Stock Options	stock, option, trad, market, tradie, trade, volume, exchange, seller, market_capitalization
167	Negotiations	deal, agreement, negotiation, president_donald, reach, sign, agre, pact, side, term
168	Company Stakes	company, familiar, matter, stake, person, investor, deal, busines, potential, includie
169	Stock	share, stock, company, dividend, shareholder, share_price, price, preferr, capital, company_share
170	Communication Industry	phone, network, customer, service, company, communication, carrier, subscriber, cable, acces
171	War	military, force, defense, nuclear, missile, official, weapon, troop, drone, army
172	*Others*[172]	samsung, spectrum, collin, cooper, phoenix, mitchell, myer, samsung_electronic, ariz, recruitment
173	Consumption	data, consumer, level, rose, number, increase, decline, earlier, fell, figure
174	Cities	texa, west, billionaire, houston, coast, disruption, dalla, louisiana, east, charlotte
175	Company Revenue	revenue, share, company, generate, company_revenue, generat, earn, share_revenue, revenue_company, period_earlier
176	Boeing	boe, order, delivery, sander, jet, engine, plane, aircraft, dynamic, deliver
177	Worker Unions	worker, union, labor, request_comment, wage, minimum, strike, workforce, employee, work
178	Steel	steel, merger, iron, hub, industry, ore, consolidation, helm, hamilton, hunt
179	Trade	trade, tariff, import, export, commerce, good, free, industry, administration, american
180	Architecture	square, space, buildie, startup, foot, design, office, floor, room, feet

Appendix B.

Table B1: **Hyperparameters For All Machine Learning Models**

Model	Hyperparameters
ENET	α=0.5 The list of penalty λ are set automatically
CSR	The number of predictors within each regression=1$\sim$ 10
PCA	The number of factors=1$\sim$ 5
PLS	The number of factors=1$\sim$ 5
RF	Trees=1000 Predictors in each split = log_2(total number of predictors) Depth= 1 $\sim$ 6
GBDT	Trees$\in$ {50,100,150,200,250,300} Predictors in each split = log_2(total number of predictors) Learning Rate =0.01 Depth= 1 $\sim$ 4

Appendix C.

Table C1: **ECON Data**

Variable	Abbreviation	Brief description
Dividend-price ratio	DP	Log difference between dividends and stock prices of the S&P 500 index.
Dividend yield	DY	Log difference of dividends and previous month's stock prices.
Dividend-payout ratio	DE	The difference between the log of dividends. and the log of earnings
Stock variance	SVAR	Sum of squared daily returns on the S&P 500 index.
Book-to-market ratio	BM	Book to market value ratio of the Dow Jones Industrial Average.
Net equity expansion	NTIS	The ratio of the 12-month moving sum of net issues for NYSE-listed stocks to the NYSE's year-end market capitalization.
Treasury bill rate	TBL	The secondary market rate for 3-month U.S. Treasury bills.
Long-term yield	LTY	Long-term government bond yield.
Long-term return	LTR	Long-term government bond return.
Default yield spread	DFY	Yield difference between BAA- and AAA-rated corporate bonds.
Default return spread	DFR	Return difference between long-term corporate bond. and long-term government bond.
Inflation	INFL	The Consumer Price Index (All Urban Consumers).
Momentum	MOM	An dummy variable that equals to one (zero) if the S&P 500 price index is greater than or equal to (less than) its value six months ago.