Narrative Risk and the Equity Premium

Gabriel A. Abeyie* Efua A. Noonoo† Emmanuel Obijole‡

Draft: March 1, 2026

Abstract

We study whether economic narratives improve equity premium forecasts and how forecast combination should adapt to changes in the narrative environment. Using narrative sentiment features extracted from major U.S. newspapers over 1940–2021, we generate machine-learning forecasts via Lasso, Elastic Net, and Random Forest, and combine them using Narrative Attention Shrinkage (NAS). Motivated by rational inattention, NAS uses a real-time Narrative Attention Index (NAI) to modulate shrinkage: when narrative activity is elevated, weights concentrate on the strongest models; when activity is low, weights revert toward equal averaging. Out of sample, NAS delivers economically and statistically significant gains relative to the historical average and standard combination benchmarks, with improvements concentrated in crisis episodes characterized by heightened narrative activity and model disagreement. These findings provide a state-dependent resolution to the forecast combination puzzle and show that narrative information is most effectively exploited through adaptive model selection.

JEL Classification: C53, C58, G12, G17

Keywords: Equity premium prediction, text analysis, forecast combination, rational inattention, machine learning, narrative economics

[1]∗ Corresponding author. Department of Economics and Finance, University of North Carolina Wilmington. Email: abeyieg@uncw.edu

[2]† Department of Statistics, Oregon State University. Email: noonooe@oregonstate.edu

[3]‡ Department of Economics, University of Tennessee, Knoxville. Email: eobijole@vols.utk.edu

1 Introduction

Can the stories that newspaper headlines tell about the economy, or economic narratives, predict the equity premium? If so, how should forecasters incorporate these signals when their informativeness varies across calm and turbulent market environments?

This paper addresses both questions. We construct 124 text-based sentiment features from a comprehensive corpus of newspaper articles spanning 1940 to 2021. These features measure the monthly intensity of distinct economic narrative themes, including *stock market* activity, *corporate debt* conditions, *inflation*, *recessions*, and developments in the *technology sector*. Using these sentiment features as predictors in an out-of-sample equity premium forecasting exercise, we introduce a novel forecast combination approach, which we term Narrative Attention Shrinkage (NAS). The method adapts model selection to the prevailing narrative environment. NAS achieves an out-of-sample R^2 of 3.54% relative to the historical average benchmark over 1991 to 2021, outperforming all individual models and alternative combination strategies, with statistical significance.

Our starting point is the observation that text-based predictors exhibit dramatic regime dependence. When we apply standard machine learning methods including Lasso, Elastic Net, and Random Forest, to the 124 narrative sentiment features, all three deliver positive and statistically significant out-of-sample predictability. But the character of this predictability varies sharply with the economic environment. For example, the Lasso delivers substantially stronger out-of-sample explanatory power during recessions than during economic expansions. The number of narrative features selected by the penalized models rises from two to three in calm periods to over 30 during crises. The identity of the crisis-activated features such as *bank*, *housing*, *default*, *recession* reflects the broadening of the narrative environment as investors grapple with multiple, simultaneously evolving stories.

This regime dependence creates a practical forecasting challenge: which model should one trust at any given point in time? The equal-weight forecast combination of Rapach et al. (2010) provides a natural default, and indeed delivers $R^2_{\text{OOS}} = 2.90\%$ in our setting. But equal weighting treats all periods identically, ignoring the rich variation in the narrative environment. The NAS combination improves upon equal weighting by making the combination strategy itself state-dependent.

The NAS methodology rests on a simple but powerful idea grounded in the theory of rational inattention (Sims, 2003; Maćkowiak and Wiederholt, 2009). Because investors face limits on information-processing capacity, the value of focusing attention varies with the richness of the narrative environment. When many economic stories are simultaneously active, predictive importance becomes dispersed across themes and forecasting models tend to disagree. In such periods, the marginal benefit of tracking relative model performance is high, and the combination strategy should place greater weight on recently accurate models. By contrast, when the narrative environment is relatively calm, with a small number of dominant themes and broad model agreement, the optimal approach is to diversify across models in order to hedge against specification error.

We evaluate the NAS combination against a comprehensive set of alternatives from the forecast combination literature: equal-weight mean, median, discounted MSFE (Stock and Watson, 2004), BIC-based approximate Bayesian model averaging, and both constrained and unconstrained OLS combinations (Granger and Ramanathan, 1984). The NAS combination outperforms every competitor in the full sample, in the post-validation subsample, during expansions, and during recessions. The improvement over equal weighting is concentrated in crisis episodes, precisely when the NAI spikes and the attention mechanism activates, with negligible cost during calm periods when the NAS weights converge toward uniformity.

Our results also speak to the long-standing forecast combination puzzle (Smith and Wallis, 2009; Claeskens et al., 2016): the well-known finding that the simple equal-weight combinations are remarkably difficult to beat. In our horse race, the unconstrained OLS combination fails catastrophically ($R^2_{\mathrm{OOS}} = -1.91\%$), the constrained OLS and DMSFE methods match but do not beat equal weighting, and BIC weights underperform. Only the NAS combination, which modulates the *responsiveness* of the weighting scheme rather than directly estimating optimal weights, consistently outperforms. By limiting the additional degrees of freedom to a single attention sensitivity parameter γ, the NAS avoids the estimation error that undermines more heavily parameterized approaches.

This paper contributes to three literatures. First, we contribute to the literature on equity premium prediction (Welch and Goyal, 2008; Campbell and Thompson, 2008; Rapach et al., 2010; Lima and Godeiro, 2023), which has documented that the historical average is notoriously difficult to beat out-of-sample. We show that text-based predictors, which capture the narrative sentiment dimension of economic information that is absent from standard macroeconomic and financial predictors, deliver robust, statistically significant predictability. Our R^2_{OOS} values of 3–4% are among the largest reported in the literature using monthly data.

Second, we contribute to the rapidly growing literature on text-based asset pricing (Tetlock, 2007; Manela and Moreira, 2017; Bybee et al., 2024; Ke et al., 2019; Hong et al., 2025). While this literature has established that text data contains information relevant for financial markets, the question of *how* to optimally extract and combine text-based forecasts has received less attention. Our NAS methodology addresses this gap by recognizing that the optimal extraction strategy is itself state-dependent, and that the text data contain information about the appropriate forecasting regime.

Third, we contribute to the forecast combination literature (Timmermann, 2006; Rapach et al., 2010; Smith and Wallis, 2009; Claeskens et al., 2016) by providing a theoretically motivated resolution to the forecast combination puzzle. The NAS combination demonstrates that state-dependent combination can outperform equal weighting when the state variable is well-chosen—in our case, an endogenous measure of the narrative information environment rather than an external regime indicator. This endogeneity is a key advantage: the NAI is constructed from the same text features used in forecasting, ensuring that the combination method adapts to the same information environment that generates the predictive signals.

A complementary experiment confirms that the attention mechanism operates at the level of model selection rather than feature selection. When we apply the same attention-modulated principle to the Lasso's regularization parameter and allow more features to enter when the narrative environment is rich, cross-validation selects no attention modulation at every forecast origin, yielding forecasts identical to the standard Lasso (Appendix C). This null result implies that the individual models already perform effective feature selection; what they cannot do is determine which type of model is best suited to the current narrative environment. The NAS combination fills precisely this gap.

The remainder of the paper is organized as follows. Section 2 describes the data and text feature construction. Section 3 develops the methodology, including the individual models, the NAI, and the NAS combination. Section 4 presents the empirical results. Section 5 evaluates the economic significance of the forecasting gains through a portfolio allocation exercise. Section 6 concludes.

2 Data

This section describes the construction of the equity premium target variable, the newspaper corpus, and the text-based predictor features.

2.1 Equity Premium

Our forecasting target is the monthly log equity premium, defined as the log excess return on the CRSP value-weighted index over the risk-free rate (Equation 4). The sample spans 1940:01–2021:12, yielding $T = 984$ monthly observations. The equity premium has a mean of 0.71% per month (8.5% annualized) and a standard deviation of 4.24% per month (14.7% annualized), consistent with the properties of the equity premium documented in Welch and

Goyal (2008).[1]

2.2 Newspaper Corpus

2.2.1 Source and Construction

Our text data are drawn from the ProQuest Historical Newspapers database, a digitized archive of major U.S. newspapers. We retrieve all articles matching the search query `"business" AND "finance" AND "economy"` over the period January 1940 through December 2021. This query is designed to capture articles with substantive economic and financial content while excluding purely local, sports, or entertainment coverage.

The resulting corpus comprises 79,697 newspaper articles drawn from four major national publications: the *New York Times* (45,030 articles, 56.5%), the *Washington Post* (18,318 articles, 23.0%), the *Wall Street Journal* (10,080 articles, 12.6%), and the *Chicago Tribune* (6,269 articles, 7.9%). All articles are in English. The corpus provides broad temporal coverage, with article counts ranging from approximately 3,100 in the 1940s to 16,900 in the 1970s, reflecting both the growth of financial journalism and the increasing digitization of newspaper archives.[2] Figure 1 provides a representative example of a newspaper article from the corpus.

Wall Street Is Poised for Higher Profits but Fewer Jobs
Moyer, Liz
New York Times (1923-); Oct 29, 2016; ProQuest Historical Newspapers: The New York Times
pg. B2

Wall Street Is Poised for Higher Profits but Fewer Jobs

By LIZ MOYER

Wall Street is on track to reverse three consecutive years of profit declines, but bonuses and jobs continue to shrink as the industry adjusts to a postcrisis world, a new report shows.

The securities industry in New York reported pretax profit of $9.3 billion through June, putting it on pace to surpass the $14.3 billion of profit for all of 2015 and rebound from annual declines since 2012, according to an annual report released on Friday by the New York State comptroller's office.

But the city's broker-dealer firms have set aside 7 percent less money for employee bonuses through June compared with the first six months of last year, and they have shed 2,600 jobs since March.

Thomas P. DiNapoli, the state comptroller, said the industry — one of the city's and state's biggest economic drivers — appears poised to benefit from a recent trend in large mergers and positive economic news, like the American economy's third-quarter growth of 2.9 percent reported on Friday. Wall Street reaps revenue from advising on deals as well as trading for clients, and robust deal activity and an expanding economy would bolster those businesses.

"Of course, no one has a crystal ball, and we know volatility is the name of the game. At this point, we just don't see anything on the horizon that causes concern," Mr. DiNapoli said on a conference call Friday. "I guess we see some overall signs as far as the economy that are positive. We expect Wall Street will benefit from that."

But Wall Street is not as dominant a factor in the everyday economy of New York City as it used to be. The securities industry cut 8 percent of its jobs from 2007 to 2015, while other private sector jobs grew 17 percent. From 2010 to last year, Wall Street accounted for 11 percent of wage growth, far less than the 39 percent attributed to it from 1990 to 2007.

Technology, advertising, media and information and business services have picked up the slack, the report said.

An industry adjusts to a slightly smaller role, a state report finds.

And New York City had 19 percent of the securities industry jobs nationwide last year, down from 32 percent in 1990. Texas and Pennsylvania have picked up the most securities industry jobs from 2010 to 2015, the comptroller's report notes.

Still, the average salary of a New York securities industry employee dwarfs that of other local private sector workers even though salaries have been declining, falling another 4 percent last year, to $388,000. That compares with the average salary of $74,100 for other private sector workers.

Bonuses are a big guessing game because factors can shift quickly as companies close their books on the year. Last year's bonuses were down for the second consecutive year, falling 9 percent to an average of $146,000.

Johnson Associates, a compensation consulting firm, projected in August that incentive pay for investment bankers could fall 5 percent to 25 percent this year, while for traders, it could drop 5 percent to 15 percent.

Blocked due to copyright.
See full page image or microfilm.

SPENCER PLATT/GETTY IMAGES

A man walking outside the New York Stock Exchange this month. The industry has cut 2,600 jobs in the city since March.

The finance industry is a crucial component of the state's and city's annual tax revenue, with profits on the line for luxury retailers, restaurants and real estate, among other industries that feel the ripple effects from Wall Street's waxing and waning fortunes. The securities industry accounted for 18.5 percent of the state's tax collections last year and 7 percent of the city's.

Figure 1: Sample *New York Times* article. This illustrates the type of economic narrative content used to construct the text-based features.

[1]The slightly shorter sample (starting 1940 rather than 1926) reflects the availability of our newspaper corpus.

[2]Coverage is somewhat thinner in the 2010s (4,730 articles) and 2020s (458 articles) due to ProQuest's rolling digitization process. We address the potential impact of time-varying corpus density on our results in Section 4.4.2.

The choice of these newspapers is motivated by three considerations. First, the *New York Times*, *Wall Street Journal*, *Washington Post*, and *Chicago Tribune* are the newspapers of record for business, financial, and economic news in the United States, and their coverage is widely followed by institutional and retail investors. Second, ProQuest provides full-text access with consistent formatting over the entire 1940–2021 period, enabling uniform text processing. Third, the use of multiple newspapers provides cross-source diversification, mitigating the risk that our results are driven by the editorial perspective of any single publication.

2.2.2 Summary Statistics

The corpus yields an average of 81 articles per month, with substantial variation over time (standard deviation of 37). The average article length is approximately 972 words. At the monthly level, the corpus processes roughly 68,000 words per month, providing a rich textual basis for sentiment extraction. Appendix A provides additional descriptive statistics on the temporal distribution and source composition of the corpus.

2.3 Text-Based Feature Construction

2.3.1 Economic Concept Selection

Our text-based features are constructed around a set of 124 economic concepts that are selected based on their relevance to equity premium determination. The concept list is drawn from Aruoba and Drechsel (ming), who compile a comprehensive vocabulary of economic terms organized by their theoretical relevance to macroeconomic and financial conditions. We adopt their classification, which groups concepts into seven categories reflecting distinct economic mechanisms through which narratives may affect the equity premium:

1. **Macroeconomic Conditions** (20 concepts): GDP and economic activity, labor market conditions, industrial production, and consumer spending. Examples: *recession*, *unemployment*, *economic growth*, *consumer spending*, *industrial production*.

2. **Inflation and Monetary Policy** (18 concepts): Price dynamics, central bank policy, and interest rates. Examples: *inflation*, *federal reserve*, *monetary policy*, *interest rates*, *inflation expectations*.

3. **Corporate Fundamentals** (25 concepts): Earnings, profits, revenues, costs, and investment. Examples: *earnings*, *profits*, *operating cost*, *capital spending*, *revenue*.

4. **Market Valuation** (12 concepts): Stock prices, valuation ratios, and market references. Examples: *stock market*, *dow jones industrial average*, *price earnings*, *dividend yield*.

5. **Credit and Financial Conditions** (24 concepts): Corporate debt, credit markets, default risk, and banking. Examples: *corporate debt*, *bond market*, *default*, *bank*, *credit spread*.

6. **Sentiment and Risk** (18 concepts): Confidence indicators, optimism/pessimism, risk appetite, and uncertainty. Examples: *business confidence*, *pessimism*, *uncertainty*, *volatility*, *expectations*.

7. **Sectors** (7 concepts): Technology, energy, and housing. Examples: *technology sector*, *oil prices*, *housing*, *real estate*.

These seven categories map naturally to the two fundamental drivers of the equity premium in asset pricing theory: expected cash flows (corporate fundamentals, macroeconomic conditions, sectors) and discount rates (inflation and monetary policy, credit conditions, sentiment and risk, market valuation).

2.3.2 Sentiment Extraction

For each economic concept, we construct a monthly sentiment index following the methodology of Hassan et al. (2019), adapted from corporate earnings calls to newspaper text. The procedure operates in three steps.

Step 1: Concept identification. For each article in the corpus, we identify all mentions of each economic concept using exact string matching on the preprocessed text. Multi-word concepts (e.g., "corporate debt," "stock market") are matched as complete phrases. Text preprocessing is intentionally minimal(lowercasing and removal of URLs) to preserve the natural language structure required for the contextual window method.[3]

Step 2: Contextual sentiment scoring. For each concept mention, we extract the ± 10 words surrounding the mention (excluding the concept itself) and classify each word in this window as positive (+1), negative (−1), or neutral (0) using the Loughran and McDonald (2011) financial sentiment dictionary. The Loughran-McDonald dictionary is specifically designed for financial text and contains 354 positive and 2,355 negative words calibrated to

[3] In particular, we do not remove stop words or apply lemmatization, as these operations would distort the word distance calculations in the sentiment extraction step.

the language of business and financial documents.[4] The raw sentiment score for concept j in article i is the sum of sentiment word values in the ± 10-word window across all mentions of concept j:

$$s_{i,j}^{\text{raw}} = \sum_{m=1}^{M_{i,j}} \left[\sum_{w \in W_m} \left(\mathbf{1}\{w \in \mathcal{P}\} - \mathbf{1}\{w \in \mathcal{N}\} \right) \right], \tag{1}$$

where $M_{i,j}$ is the number of mentions of concept j in article i, W_m denotes the set of words in the ± 10-word window around mention m, and $\mathcal{P}$ and $\mathcal{N}$ are the positive and negative word sets from the Loughran-McDonald dictionary.

Step 3: Normalization and monthly aggregation. The raw sentiment score is normalized by the total word count of article i to control for article length:

$$s_{i,j} = \frac{s_{i,j}^{\text{raw}}}{N_i}, \tag{2}$$

where N_i is the total number of words in article i. The monthly sentiment index for concept j in month t is then the sum of normalized scores across all articles published in that month:

$$x_{j,t} = \sum_{i \in \mathcal{A}_t} s_{i,j}, \tag{3}$$

where $\mathcal{A}_t$ denotes the set of articles published in month t. Summing rather than averaging preserves the intensive margin: months with more articles mentioning concept j in a positive (negative) context receive higher (lower) index values, capturing both the direction and intensity of the narrative.

2.3.3 Properties of the Feature Matrix

The resulting feature matrix $\boldsymbol{X} \in \mathbb{R}^{984 \times 124}$ contains the (unstandardized) monthly sentiment indices for each concept over the full sample.[5] The matrix is moderately sparse: 66.1% of entries are exactly zero, reflecting months in which a given concept either does not appear in the corpus or appears without sentiment-bearing context. Sparsity varies substantially across concepts. For example, high-frequency narratives like *unemployment* (non-zero in 86.2% of months) and *recession* (78.3%) have dense coverage, while more specialized concepts like *gdp growth* (4.2%) and *top line* (1.8%) are sparse.

[4] We also extract uncertainty scores from the Loughran-McDonald uncertainty word list, though these are not used in the primary analysis.

[5] The features used in the forecasting models are the raw unstandardized sentiment indices. We do not apply full-sample standardization to avoid look-ahead bias; any standardization used in model estimation or the NAI is performed on expanding windows.

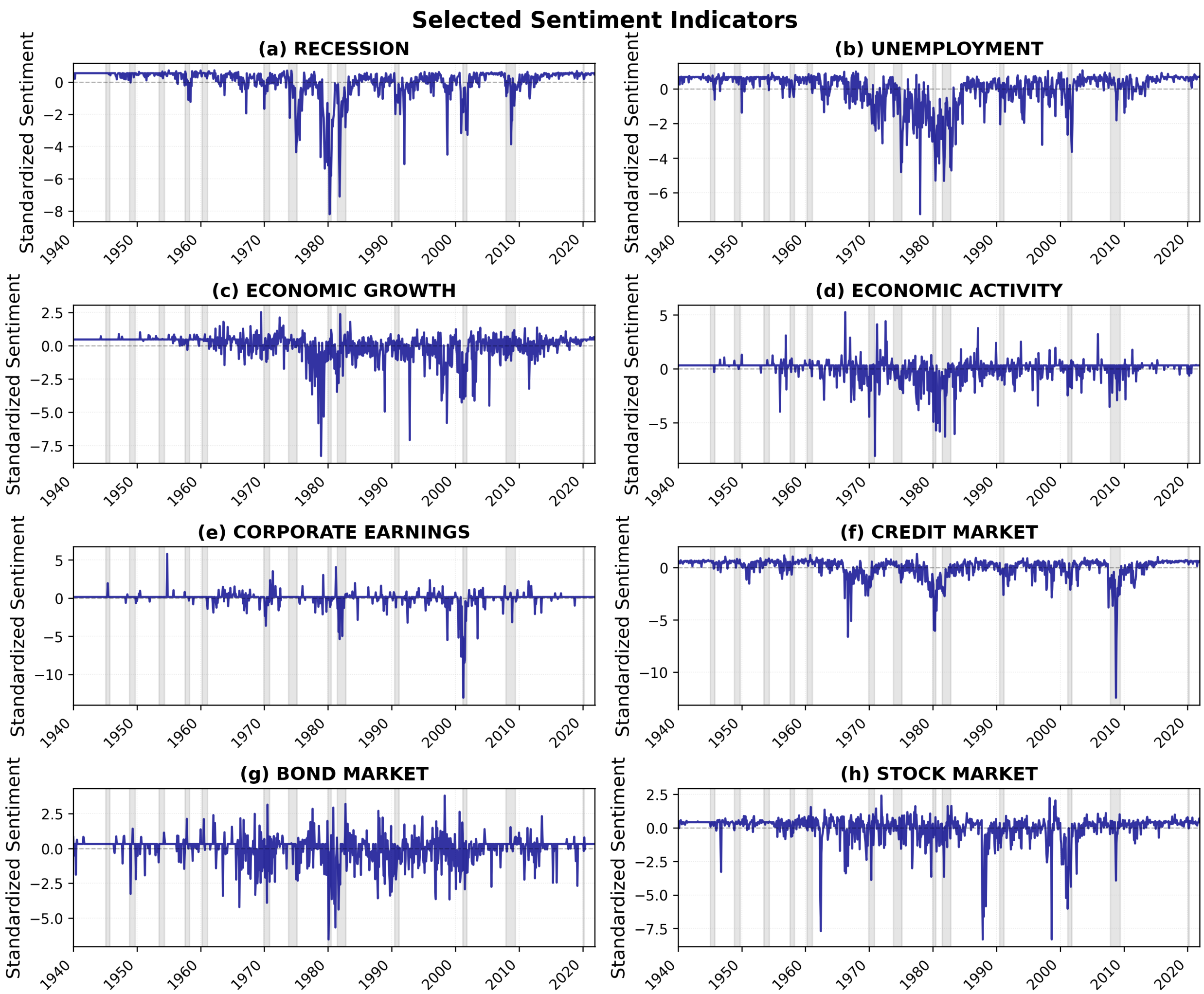


Figure 2: Sentiment indicators for a selection of economic concepts discussed in the corpus, out of our full list of 124. The sentiments are constructed using the dictionary of positive and negative words in financial text of Loughran and McDonald (2011). Each indicator is standardized across the sample. Shaded areas represent NBER recessions.

This heterogeneity in concept coverage has two important implications. First, it provides a natural rationale for regularization: with many features that are zero in most months, the effective dimensionality of the prediction problem varies over time. Second, the time-variation in which concepts are active and driven by the ebb and flow of the economic narrative, is precisely the signal exploited by the Narrative Attention Index in Section 3.5.

Figure 2 presents the sentiment indicators for selected economic concepts. Each series is standardized by removing its mean and scaling by its time-series standard deviation, with no additional smoothing or filtering applied. The indicators exhibit substantial and economically meaningful variation over time. For instance, Panel (a) shows that sentiment surrounding "recession" declines sharply during recessionary periods.

2.4 Sample Design

The full sample of 984 monthly observations (1940:01–2021:12) is divided into an initial training sample of $T_0 = 615$ observations (62.5%, 1940:01–1991:03) and an out-of-sample evaluation period of $T_1 = 369$ observations (37.5%, 1991:04–2021:12). This split follows the convention in Welch and Goyal (2008) and ensures a sufficiently long initial training window to estimate the machine learning models reliably. Within the out-of-sample period, months 60–180 (1996:04–2006:04) serve as a validation window for the NAS hyperparameters (τ_0, γ, δ), with the post-validation period (2006:04–2021:12, $T = 189$ months) providing a fully clean evaluation sample.

Table 1 summarizes the sample design, including the number of NBER recession months in each subsample.

Table 1: Sample Design

Period	Dates	Observations	Recession Months
Full sample	1940:01–2021:12	984	—
Initial training	1940:01–1991:03	615	—
Out-of-sample (OOS)	1991:04–2021:12	369	28
NAS validation	1996:04–2006:04	120	8
Post-validation	2006:04–2021:12	189	18

Notes: Recession months in the OOS period correspond to the 2001 recession (2001:04–2001:11, 8 months) and the Great Recession (2008:01–2009:06, 18 months). Two additional recession months (1991:04–1991:06) fall at the start of the OOS period.

3 Methodology

We develop a two-stage forecasting framework for the equity premium that exploits the information content of economic narratives extracted from newspaper text. In the first stage, three machine learning models, Lasso, Elastic Net, and Random Forest, are used to extract predictive signals from a high-dimensional set of text-based sentiment features. In the second stage, these individual forecasts are combined using time-varying weights that adapt to the prevailing narrative environment. Our methodological contribution is the Narrative Attention Shrinkage (NAS) combination, which modulates the aggressiveness of model selection as a function of a real-time Narrative Attention Index derived from the text features themselves.

This two-stage design reflects a conceptual separation between two distinct forecasting problems: *which narrative features matter?* (addressed by the individual machine learning models) and *which model to trust right now?* (addressed by the combination method). The NAS combination answers the second question by recognizing that the optimal combination strategy is itself state-dependent: during periods of rich, rapidly evolving narratives, the combination should aggressively favor recently accurate models; during calm periods, it should diversify across models to guard against specification error.

3.1 Forecasting Setup

Let r_{t+1} denote the equity premium in month $t+1$, defined as the log excess return of the market portfolio over the risk-free rate:

$$r_{t+1} = \log(1 + R_{m,t+1}) - \log(1 + R_{f,t+1}), \tag{4}$$

where $R_{m,t+1}$ is the gross return on the CRSP value-weighted index and $R_{f,t+1}$ is the Treasury bill rate. We consider the predictive regression:

$$r_{t+1} = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim (0, \sigma_\varepsilon^2), \tag{5}$$

where $\boldsymbol{x}_t \in \mathbb{R}^p$ is a vector of p text-based predictors observable at time t, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the coefficient vector, and ε_{t+1} is the forecast error.

Our sample spans 1940:01–2021:12, yielding $T = 983$ monthly observations. We adopt an expanding-window out-of-sample evaluation design following Welch and Goyal (2008). The initial training sample comprises the first 62.5% of observations ($T_0 = 614$ months, ending 1991:03), with the remaining 37.5% ($T_1 = 369$ months, 1991:04–2021:12) reserved for out-of-sample evaluation. At each forecast origin t, models are estimated using all available data through time t and produce one-step-ahead forecasts $\hat{r}_{t+1|t}$, yielding a sequence of T_1 genuinely out-of-sample predictions.

3.2 Text-Based Predictor Construction

Our predictor set $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{p,t})'$ consists of $p = 124$ text-based features constructed from a comprehensive corpus of newspaper articles spanning 1940–2021.[6] Each feature $x_{j,t}$ captures the monthly intensity of a distinct economic narrative theme j and is constructed as the normalized frequency with which narrative-relevant terms appear in the news corpus during month t.

The key identifying assumption is that newspaper text captures the evolution of economic narratives: the stories, sentiments, framings, and interpretations through which market participants process information. The assumption holds that shifts in these narratives contain predictive content for the equity premium beyond what is captured by standard macroeconomic and financial predictors. This assumption is grounded in the growing literature on narrative economics (Shiller, 2017, 2019) and the demonstrated forecasting value of text-based measures (Manela and Moreira, 2017; Bybee et al., 2024; Ke et al., 2019).

The high dimensionality of the predictor space ($p = 124$) relative to the signal-to-noise ratio in equity premium prediction necessitates regularization. A central question is how aggressively to regularize—a question whose answer, we argue, should depend on the state of the narrative environment itself.

3.3 Individual Forecasting Models

We employ three machine learning methods as individual forecasting models, each estimated via rolling re-optimization to balance computational cost with adaptiveness. These models serve as the building blocks for the forecast combination methods described in Section 3.4.

3.3.1 Lasso

The Lasso estimator (Tibshirani, 1996) solves:

$$\hat{\boldsymbol{\beta}}_t^{\text{Lasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2t} \sum_{s=1}^{t} \left(r_s - \alpha - \boldsymbol{x}_{s-1}' \boldsymbol{\beta} \right)^2 + \lambda_t \sum_{j=1}^{p} |\beta_j| \right\}, \tag{6}$$

[6] Details of the corpus construction, article selection criteria, and text preprocessing are provided in Section 2.

where $\lambda_t > 0$ is the regularization parameter, selected via five-fold time-series cross-validation (Hyndman and Athanasopoulos, 2018) with re-tuning every 12 months. The ℓ_1 penalty induces exact sparsity, selecting a subset of text features and setting the remaining coefficients to zero.

3.3.2 Elastic Net

The Elastic Net (Zou and Hastie, 2005) combines ℓ_1 and ℓ_2 penalties:

$$\hat{\boldsymbol{\beta}}_t^{\mathrm{EN}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2t} \sum_{s=1}^{t} \left(r_s - \alpha - \boldsymbol{x}_{s-1}' \boldsymbol{\beta} \right)^2 + \lambda_t \left[\rho \sum_{j=1}^{p} |\beta_j| + \frac{1-\rho}{2} \sum_{j=1}^{p} \beta_j^2 \right] \right\}, \tag{7}$$

where $\rho \in (0, 1]$ controls the relative weight on the ℓ_1 versus ℓ_2 penalty. Both λ_t and ρ_t are selected jointly by time-series cross-validation. The ℓ_2 component encourages grouped selection among correlated text features.

3.3.3 Random Forest

The Random Forest (Breiman, 2001) produces forecasts via ensemble averaging of B regression trees:

$$\hat{r}_{t+1|t}^{\mathrm{RF}} = \frac{1}{B} \sum_{b=1}^{B} T_b(\boldsymbol{x}_t; \hat{\Theta}_{b,t}), \tag{8}$$

where $T_b(\cdot\,; \hat{\Theta}_{b,t})$ is the b-th tree grown on a bootstrap subsample with random feature subsampling, and $\hat{\Theta}_{b,t}$ denotes the tree-specific parameters estimated on data through time t. Hyperparameters (tree depth, minimum leaf size, feature subsample fraction, number of trees) are selected via randomized time-series cross-validation with re-tuning every 24 months. The Random Forest accommodates nonlinear relationships between text features and the equity premium without imposing linearity.

3.4 Forecast Combination Methods

Let $\hat{r}_{t+1|t}^{(k)}$ denote the one-step-ahead forecast from individual model $k \in \{1, \ldots, K\}$ at forecast origin t, with $K = 3$ (Lasso, Elastic Net, Random Forest). The general form of a combined forecast is:

$$\hat{r}_{t+1|t}^{\mathrm{C}} = \sum_{k=1}^{K} \omega_{k,t}\, \hat{r}_{t+1|t}^{(k)}, \tag{9}$$

where $\omega_{k,t}$ are the combination weights. Different methods impose different structures on the weights. We consider a comprehensive set of combination methods from the literature, which we group into three categories: fixed-weight, performance-adaptive, and regression-based.

3.4.1 Fixed-Weight Combinations

Equal-Weight Mean (FC). The simple average assigns uniform weights $\omega_{k,t} = 1/K$ for all k and t:

$$\hat{r}^{\mathrm{FC}}_{t+1|t} = \frac{1}{K}\sum_{k=1}^{K} \hat{r}^{(k)}_{t+1|t}. \tag{10}$$

Rapach et al. (2010) and Timmermann (2006) demonstrate that equal-weight combinations are remarkably difficult to beat in equity premium forecasting, as they diversify across model specification errors without introducing estimation error into the weights.

Median. The median combination is robust to outlier forecasts:

$$\hat{r}^{\mathrm{Med}}_{t+1|t} = \mathrm{median}\left(\hat{r}^{(1)}_{t+1|t}, \ldots, \hat{r}^{(K)}_{t+1|t}\right). \tag{11}$$

With $K = 3$, this selects the middle forecast, providing robustness against extreme predictions from any single model.

3.4.2 Performance-Adaptive Combinations

Performance-adaptive methods assign higher weights to models with better recent forecasting performance.

Discounted MSFE (DMSFE). Following Stock and Watson (2004), the DMSFE method weights models inversely proportional to their discounted mean squared forecast errors:

$$\omega^{\mathrm{DMSFE}}_{k,t} = \frac{\left(\mathrm{DMSFE}_{k,t}\right)^{-1}}{\sum_{m=1}^{K}\left(\mathrm{DMSFE}_{m,t}\right)^{-1}}, \tag{12}$$

where the discounted MSFE of model k is:

$$\mathrm{DMSFE}_{k,t} = \sum_{s=1}^{t-1} \theta^{t-1-s}\left(r_{s+1} - \hat{r}^{(k)}_{s+1|s}\right)^2. \tag{13}$$

The discount factor $\theta \in (0, 1]$ controls the degree to which recent performance is emphasized. We consider $\theta = 1$ (no discounting, all history weighted equally) and $\theta = 0.9$ (moderate discounting, emphasizing recent accuracy), denoted DMSFE(1) and DMSFE(0.9), respectively.

BIC Weights. Model weights based on the Bayesian Information Criterion provide an approximate Bayesian model averaging scheme:

$$\omega_{k,t}^{\mathrm{BIC}} = \frac{\exp\left(-\frac{1}{2}\Delta\mathrm{BIC}_{k,t}\right)}{\sum_{m=1}^{K}\exp\left(-\frac{1}{2}\Delta\mathrm{BIC}_{m,t}\right)}, \tag{14}$$

where $\Delta\mathrm{BIC}_{k,t} = \mathrm{BIC}_{k,t} - \min_m \mathrm{BIC}_{m,t}$ and $\mathrm{BIC}_{k,t} = t\log(\mathrm{RSS}_{k,t}/t) + q_k \log(t)$ with q_k denoting the effective number of parameters in model k. This method provides a principled penalty for model complexity.

3.4.3 Regression-Based Combinations

OLS Combination. The unconstrained Granger and Ramanathan (1984) combination estimates optimal weights by regressing realized returns on individual forecasts:

$$r_{s+1} = c_t + \sum_{k=1}^{K} \omega_{k,t}\, \hat{r}^{(k)}_{s+1|s} + u_{s+1}, \quad s = 1, \dots, t-1, \tag{15}$$

with the weights $\hat{\omega}_{k,t}$ and intercept $\hat{c}_t$ estimated by OLS using data through time t, and the combined forecast formed as $\hat{r}^{\mathrm{OLS}}_{t+1|t} = \hat{c}_t + \sum_k \hat{\omega}_{k,t}\hat{r}^{(k)}_{t+1|t}$. Weights are unconstrained and re-estimated at each forecast origin.

Constrained OLS (C-OLS). The constrained variant imposes non-negativity and the adding-up restriction:

$$\hat{\boldsymbol{\omega}}_t = \arg\min_{\boldsymbol{\omega}} \frac{1}{t}\sum_{s=1}^{t}\left(r_{s+1} - \sum_{k=1}^{K}\omega_k \hat{r}^{(k)}_{s+1|s}\right)^2 \quad \text{s.t.} \quad \omega_k \geq 0,\ \sum_{k=1}^{K}\omega_k = 1. \tag{16}$$

The non-negativity and sum-to-one constraints prevent extreme weights and ensure the combined forecast lies within the convex hull of individual forecasts.

Remark 1 (Estimation error in combination weights)**.** *A central theme in the forecast combination literature is the forecast combination puzzle: simple methods (equal-weight, median) frequently outperform sophisticated methods that estimate optimal weights (Smith and Wallis, 2009; Claeskens et al., 2016). The reason is that the estimation error in the weights can offset the theoretical gains from optimization. Methods that use more information, like OLS, C-OLS, and BIC are more susceptible to this problem. Our NAS combination navigates this tension by modulating the sharpness of the weighting scheme rather than directly estimating optimal weights, thereby limiting the degrees of freedom available for overfitting.*

3.5 Narrative Attention Index

Central to our methodology is the construction of a real-time measure of the degree to which economic narratives demand investor attention. We call this measure the Narrative Attention Index (NAI). The NAI is constructed entirely from by-products of the first-stage model estimation, requiring no additional data or computation.

Definition 1 (Narrative Attention Index)**.** *The Narrative Attention Index at time t is a composite measure synthesizing three dimensions of the narrative information environment, each computed using only information available through time t:*

$$\mathrm{NAI}_t = \frac{1}{3}\left(\tilde{B}_t + \tilde{D}_t + \tilde{M}_t\right), \tag{17}$$

where $\tilde{B}_t$, $\tilde{D}_t$, and $\tilde{M}_t$ are expanding-window standardized versions of the narrative breadth, importance dispersion, and model disagreement components, respectively.

The three components capture complementary dimensions of the narrative environment.

Narrative Breadth (B_t). Let $\hat{\boldsymbol{\beta}}_t^{(k)}$ denote the coefficient vector estimated by model $k \in \{\text{Lasso}, \text{EN}\}$ at forecast origin t. The narrative breadth measures the average number of narrative themes with non-zero estimated coefficients:

$$B_t = \frac{1}{2}\left(\sum_{j=1}^{p} \mathbf{1}\left\{\hat{\beta}_{j,t}^{\text{Lasso}} \neq 0\right\} + \sum_{j=1}^{p} \mathbf{1}\left\{\hat{\beta}_{j,t}^{\text{EN}} \neq 0\right\}\right). \tag{18}$$

Higher values indicate that a broader range of narratives carry predictive content, reflecting a richer narrative information environment. When many narrative themes are simultaneously active, the information available to investors is more complex and demands greater attention.

Importance Dispersion (D_t). Let $\hat{w}_{j,t}^{\text{RF}}$ denote the Random Forest feature importance (mean decrease in impurity) of narrative feature j at time t, normalized such that $\sum_{j=1}^{p} \hat{w}_{j,t}^{\text{RF}} = 1$. The importance dispersion is the Shannon entropy of the importance distribution:

$$D_t = -\sum_{j=1}^{p} \hat{w}_{j,t}^{\text{RF}} \log\left(\hat{w}_{j,t}^{\text{RF}}\right). \tag{19}$$

Higher entropy indicates that predictive content is dispersed across many narrative themes rather than concentrated in a few, suggesting that investors must attend to a wider array of stories to form accurate expectations.

Model Disagreement (M_t). The cross-model disagreement captures the dispersion of forecasts across the three individual models:

$$M_t = \left[\frac{1}{K} \sum_{k=1}^{K} \left(\hat{r}^{(k)}_{t+1|t} - \bar{r}_{t+1|t} \right)^2 \right]^{1/2}, \tag{20}$$

where $\bar{r}_{t+1|t} = K^{-1} \sum_k \hat{r}^{(k)}_{t+1|t}$ is the mean forecast. Higher disagreement signals that the narrative environment is ambiguous—different modeling approaches extract conflicting signals from the same text features. Under rational inattention (Sims, 2003), ambiguity increases the marginal value of information and, consequently, the marginal value of discriminating among models.

Standardization. To ensure equal contribution from each component and to avoid look-ahead bias, we standardize using expanding windows:

$$\tilde{C}_t = \frac{C_t - \bar{C}_{1:t}}{\hat{\sigma}_{C,1:t}}, \quad C \in \{B, D, M\}, \tag{21}$$

where $\bar{C}_{1:t} = t^{-1} \sum_{s=1}^{t} C_s$ and $\hat{\sigma}_{C,1:t} = \left[t^{-1} \sum_{s=1}^{t} (C_s - \bar{C}_{1:t})^2 \right]^{1/2}$ are the expanding-window mean and standard deviation, computed using only past and current information.

Remark 2 (Real-time computability)**.** *All three components of the NAI are constructed as by-products of the first-stage model estimation. The breadth B_t comes from the Lasso and Elastic Net sparsity patterns; the dispersion D_t from Random Forest importance scores; and the disagreement M_t from the cross-section of individual model forecasts. No additional estimation or data is required, and the NAI is fully computable in real time at each forecast origin.*

Remark 3 (Economic interpretation)**.** *The NAI captures the overall "attentional demand" of the narrative environment. High NAI values(periods when many narratives are simultaneously active, predictive importance is dispersed, and models disagree) correspond to economic episodes where the information environment is rich and rapidly evolving (e.g., financial crises, major policy shifts, geopolitical shocks). Low NAI values indicate calm periods dominated by few, stable narratives. The rational inattention literature (Sims, 2003; Maćkowiak and Wiederholt, 2009; Maćkowiak et al., 2023) predicts that agents optimally allocate more attention when the information environment is richer, precisely the mechanism we exploit in the combination stage.*

3.6 Narrative Attention Shrinkage Combination

The NAS combination is a performance-adaptive scheme whose *responsiveness* to recent model performance is governed by the narrative attention state. The key insight is that the optimal trade-off between diversification (equal weights) and specialization (best-model selection) depends on the information environment: when narratives are rich and volatile, the best model may shift rapidly, and the combination should be responsive; when narratives are calm, models perform similarly, and diversification is preferable.

3.6.1 Theoretical Motivation

Consider a forecaster who must combine K model forecasts under uncertainty about which model is currently best. The forecaster faces a trade-off between *exploitation* (concentrating weight on the recently best-performing model) and *exploration* (diversifying across models to hedge against regime changes). In the language of rational inattention (Sims, 2003), the forecaster has limited capacity $\kappa > 0$ to process information about relative model performance.

Under Gaussian assumptions, the optimal attention allocation is increasing in (i) the volatility of the object of interest and (ii) the informativeness of available signals (Maćkowiak and Wiederholt, 2009; Maćkowiak et al., 2023). When the narrative environment is rich (high NAI), the relative accuracy of models shifts more rapidly, and the marginal value of attending to recent performance differences is high. In calm periods (low NAI), model rankings are stable, and the optimal strategy is to diversify.

The NAS combination operationalizes this logic through an attention-modulated temperature parameter that governs the sharpness of the weight distribution.

3.6.2 Attention-Modulated Softmax Weights

The NAS combination assigns time-varying weights to the K individual forecasts via a softmax function with attention-modulated temperature:

$$\omega_{k,t}^{\text{NAS}} = \frac{\exp(S_{k,t} \,/\, \tau_t)}{\sum_{m=1}^{K} \exp(S_{m,t} \,/\, \tau_t)}, \quad k = 1, \ldots, K, \tag{22}$$

where $S_{k,t}$ is the cumulative discounted performance score of model k and $\tau_t > 0$ is the temperature parameter.

Performance scores. The score $S_{k,t}$ accumulates discounted past forecast accuracy, measured by negative squared errors:

$$S_{k,t} = \sum_{s=1}^{t-1} \delta^{t-1-s} \left[-\left(r_{s+1} - \hat{r}^{(k)}_{s+1|s} \right)^2 \right], \tag{23}$$

where $\delta \in (0,1)$ is a discount factor that downweights distant performance. Models with higher (less negative) cumulative scores have made smaller prediction errors in the recent past.

Attention-modulated temperature. The temperature parameter governs the sharpness of the softmax and is modulated by the NAI:

$$\tau_t = \tau_0 \cdot \exp(-\gamma \cdot \mathrm{NAI}_t), \tag{24}$$

where $\tau_0 > 0$ is the base temperature and $\gamma \geq 0$ is the *attention sensitivity* parameter. This specification produces the following regime-dependent behavior:

- **High NAI** (turbulent narrative environment): τ_t is low, so the softmax weights are *sharp*—the combination aggressively favors the recently best-performing model. This mirrors the prediction of rational inattention theory that agents discriminate more finely among information sources when the environment is information-rich.
- **Low NAI** (calm narrative environment): τ_t is high, so the weights approach the uniform distribution $\omega_{k,t} \to 1/K$—the combination is *conservative*, diversifying across models to guard against model specification error. This reflects the prediction that agents economize on attention when the environment is stable.

The comparative statics formalize this intuition:

$$\frac{\partial \tau_t}{\partial \mathrm{NAI}_t} = -\gamma\, \tau_t < 0 \quad \text{for } \gamma > 0. \tag{25}$$

NAS combined forecast. The combined forecast is:

$$\hat{r}^{\mathrm{NAS}}_{t+1|t} = \sum_{k=1}^{K} \omega^{\mathrm{NAS}}_{k,t}\, \hat{r}^{(k)}_{t+1|t}. \tag{26}$$

3.6.3 Nesting and Special Cases

The NAS combination nests several benchmark combination methods as special or limiting cases, providing a unified framework for interpreting existing approaches:

Proposition 1 (Nesting). *The NAS combination nests the following methods:*

(i) ***Equal-weight combination****: If $\gamma = 0$ and $\tau_0 \to \infty$, then $\omega_{k,t}^{\mathrm{NAS}} \to 1/K$ for all k and t, recovering the simple average of Equation* (10).

(ii) ***Best-model selection****: If $\gamma = 0$ and $\tau_0 \to 0$, then $\omega_{k^*,t}^{\mathrm{NAS}} \to 1$ where $k^* = \arg\max_k S_{k,t}$, selecting the single model with the best cumulative track record.*

(iii) ***Standard softmax combination****: If $\gamma = 0$ with τ_0 fixed, the NAS combination reduces to a constant-temperature softmax, similar to exponential weighting schemes in the online learning literature (Cesa-Bianchi and Lugosi, 2006).*

The attention sensitivity parameter γ therefore has a direct testable interpretation: $\gamma > 0$ implies that state-dependent combination improves forecasting performance, consistent with the rational inattention hypothesis. A test of $\gamma = 0$ is a test of the null that the optimal combination strategy is invariant to the narrative environment.

3.6.4 Relation to the Forecast Combination Puzzle

A persistent finding in the forecast combination literature, often referred to as the "forecast combination puzzle", is that simple equal-weight combinations are remarkably difficult to beat (Smith and Wallis, 2009; Claeskens et al., 2016). The standard explanation is that estimation error in optimized weights offsets any theoretical gains from optimization. The NAS combination addresses this challenge through two design features.

First, the NAS combination does not directly estimate optimal weights; rather, it modulates the *sharpness* of a performance-based weighting scheme. The only additional parameter relative to the equal-weight benchmark is the attention sensitivity γ, minimizing the scope for overfitting. Second, the modulation is driven by the NAI, which is constructed from by-products of the first-stage estimation and is therefore not an additional source of estimation error. The combination inherits the diversification benefits of equal weighting (when NAI is low) while achieving the responsiveness of best-model selection (when NAI is high), with the narrative environment itself determining the appropriate regime.

3.6.5 Hyperparameter Selection

The NAS combination involves three hyperparameters: the base temperature τ_0, the attention sensitivity γ, and the discount factor δ. We select these jointly via grid search over an initial validation window drawn from the out-of-sample period to avoid contaminating the final evaluation.

Specifically, let $\mathcal{T}_{\text{val}} = \{T_0 + 60, \ldots, T_0 + 180\}$ denote a validation window corresponding approximately to 1996:04–2006:04. The hyperparameters are selected to minimize the validation-period MSFE:

$$(\tau_0^*, \gamma^*, \delta^*) = \arg \min_{\tau_0 \in \mathcal{G}_\tau,\, \gamma \in \mathcal{G}_\gamma,\, \delta \in \mathcal{G}_\delta} \sum_{t \in \mathcal{T}_{\text{val}}} \left(r_{t+1} - \hat{r}^{\text{NAS}}_{t+1|t}(\tau_0, \gamma, \delta) \right)^2, \tag{27}$$

where the grids are $\mathcal{G}_\tau = \{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 10^{-1}, 5 \times 10^{-1}, 1\}$, $\mathcal{G}_\gamma = \{0, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2, 3\}$, and $\mathcal{G}_\delta = \{0.90, 0.93, 0.95, 0.97, 0.99\}$.

The true out-of-sample evaluation then uses the full evaluation period $\mathcal{T}_1$, with an additional "post-validation" evaluation on $t > T_0 + 180$ (2006:04–2021:12) to assess performance on data that played no role in any model selection decision. We also report results for the full sample to maintain comparability with the literature.

Remark 4 (Warmup period)**.** *During the first $w_0 = 24$ months of the out-of-sample period, the performance scores $S_{k,t}$ have accumulated insufficient history to distinguish among models reliably. We therefore set $\omega_{k,t} = 1/K$ for $t < T_0 + w_0$, defaulting to the equal-weight combination during this warmup phase.*

3.7 Benchmark Models

We evaluate all forecasting models and combination methods against two standard benchmarks from the equity premium forecasting literature.

Historical Average (HA). The prevailing-mean benchmark of Welch and Goyal (2008) predicts the equity premium as the expanding-window sample mean:

$$\hat{r}^{\text{HA}}_{t+1|t} = \frac{1}{t} \sum_{s=1}^{t} r_s. \tag{28}$$

This benchmark corresponds to the null hypothesis that the equity premium is constant and unpredictable, and is notoriously difficult to beat out-of-sample (Welch and Goyal, 2008; Campbell and Thompson, 2008).

Autoregressive Model (AR). The AR(p^*) benchmark captures persistence in the equity premium, with lag order selected by the Akaike Information Criterion at each forecast origin:

$$\hat{r}^{\text{AR}}_{t+1|t} = \hat{\alpha}_t + \sum_{j=1}^{p_t^*} \hat{\phi}_{j,t}\, r_{t+1-j}, \quad p_t^* = \arg \min_{p \in \{1,\ldots,10\}} \text{AIC}_t(p). \tag{29}$$

3.8 Statistical Evaluation

3.8.1 Out-of-Sample R^2

Following Campbell and Thompson (2008), we evaluate predictive performance using the out-of-sample R^2:

$$R^2_{\text{OOS}} = 1 - \frac{\sum_{t=T_0}^{T-1} \left(r_{t+1} - \hat{r}_{t+1|t}\right)^2}{\sum_{t=T_0}^{T-1} \left(r_{t+1} - \hat{r}^{\text{bench}}_{t+1|t}\right)^2} = 1 - \frac{\text{MSFE}_{\text{model}}}{\text{MSFE}_{\text{bench}}}, \tag{30}$$

where $\hat{r}^{\text{bench}}_{t+1|t}$ is the benchmark forecast (HA or AR). A positive R^2_{OOS} indicates that the model outperforms the benchmark. At the monthly frequency, values exceeding 0.5% are considered economically significant (Campbell and Thompson, 2008).

3.8.2 Clark-West Test

We test the null hypothesis of equal or inferior predictive accuracy using the Clark and West (2007) MSPE-adjusted test. Define:

$$\hat{f}_{t+1} = \left(r_{t+1} - \hat{r}^{\text{bench}}_{t+1|t}\right)^2 - \left[\left(r_{t+1} - \hat{r}_{t+1|t}\right)^2 - \left(\hat{r}^{\text{bench}}_{t+1|t} - \hat{r}_{t+1|t}\right)^2\right]. \tag{31}$$

The adjustment term corrects for the upward bias in the MSFE of the more parameterized model under the null. The Clark-West statistic is:

$$\text{CW} = \frac{\bar{f}}{\hat{\sigma}_f / \sqrt{T_1}}, \tag{32}$$

where $\bar{f} = T_1^{-1} \sum_t \hat{f}_{t+1}$ and $\hat{\sigma}_f$ is the Newey-West HAC standard error with bandwidth $\lfloor T_1^{1/3} \rfloor$. The test is one-sided (H_1: model is more accurate), compared to standard normal critical values.

3.8.3 Diebold-Mariano Test

To test for equal predictive accuracy between the NAS combination and competing combination methods, we employ the Diebold and Mariano (1995) test. Let $d_t = e_{A,t}^2 - e_{B,t}^2$ denote the loss differential between methods A and B. The DM statistic is:

$$\mathrm{DM} = \frac{\bar{d}}{\hat{\sigma}_d/\sqrt{T_1}}, \tag{33}$$

where $\hat{\sigma}_d$ is the HAC standard error. Unlike the Clark-West test, the DM test is appropriate for comparing non-nested models (i.e., different combination methods applied to the same base forecasts). We use the Harvey et al. (1997) small-sample correction.

3.8.4 Cumulative SSE Difference

To visualize the temporal evolution of relative forecast accuracy, we compute the cumulative sum of squared error differences:

$$\Delta\mathrm{SSE}_t = \sum_{s=T_0}^{t} \left[\left(r_{s+1} - \hat{r}_{s+1|s}^{\mathrm{bench}} \right)^2 - \left(r_{s+1} - \hat{r}_{s+1|s} \right)^2 \right]. \tag{34}$$

Positive values indicate cumulative outperformance of the model relative to the benchmark. This diagnostic, advocated by Welch and Goyal (2008), reveals whether predictive gains are concentrated in specific episodes or distributed across the sample.

3.8.5 Subsample Analysis

To test whether the narrative attention mechanism is most valuable during periods of economic stress, we compute R_{OOS}^2 separately for NBER-dated expansion and recession months:

$$R_{\mathrm{OOS},\mathcal{S}}^2 = 1 - \frac{\sum_{t \in \mathcal{S}} \left(r_{t+1} - \hat{r}_{t+1|t} \right)^2}{\sum_{t \in \mathcal{S}} \left(r_{t+1} - \hat{r}_{t+1|t}^{\mathrm{bench}} \right)^2}, \quad \mathcal{S} \in \{\mathrm{Expansion}, \mathrm{Recession}\}. \tag{35}$$

If the attention mechanism operates as intended, we expect the NAS combination's improvement over fixed-weight methods to be concentrated in recessions, when the narrative environment is richest and the optimal model ranking shifts most rapidly.

3.8.6 Rolling R^2_{OOS}

To examine time-variation in predictive accuracy, we compute rolling-window R^2_{OOS} over 60-month windows:

$$R^{2,\text{roll}}_{\text{OOS},t} = 1 - \frac{\sum_{s=t-59}^{t} \left(r_{s+1} - \hat{r}_{s+1|s}\right)^2}{\sum_{s=t-59}^{t} \left(r_{s+1} - \hat{r}^{\text{bench}}_{s+1|s}\right)^2}. \tag{36}$$

This diagnostic reveals whether the NAS combination delivers consistently positive R^2_{OOS} or whether gains are episodic, and allows visual inspection of the relationship between the NAI and forecasting performance.

4 Results

4.1 Individual Model Performance

Table 2 reports the out-of-sample forecasting performance of the three individual text-based models against both benchmarks. All three models deliver positive R^2_{OOS} relative to the historical average, with values ranging from 1.82% (Random Forest) to 3.03% (Lasso). These magnitudes are economically meaningful: Campbell and Thompson (2008) argue that monthly R^2_{OOS} values exceeding 0.5% can generate substantial utility gains for a mean-variance investor.[7] The Clark-West test rejects the null of equal predictive accuracy in favor of all three text-based models against both benchmarks at the 5% level, with p-values ranging from 0.011 to 0.019 against the historical average.

The AR benchmark itself underperforms the historical average ($R^2_{\text{OOS}} = -0.66\%$), consistent with the well-documented finding that autoregressive dynamics add little to equity premium prediction (Welch and Goyal, 2008). This underperformance is reflected in the even larger R^2_{OOS} values when text-based models are evaluated against the AR benchmark: 3.66% for the Lasso, 3.61% for the Elastic Net, and 2.47% for the Random Forest, all significant at the 1% level.

Two features of these results merit emphasis. First, the strong performance of the Lasso and Elastic Net (which impose sparsity) relative to the Random Forest suggests that the predictive signal in text features is concentrated in a small subset of narrative themes rather than distributed diffusely across many features. We examine this pattern in detail in Section 4.2. Second, the similarity of the Lasso and Elastic Net results (R^2_{OOS} of 3.03% and 2.97%, respectively) suggests that the grouping property of the ℓ_2 penalty provides limited

[7] The connection between statistical and economic significance is formalized in Section 5, where we evaluate the utility gains from a portfolio allocation exercise.

incremental benefit, consistent with the predictive signal being driven by a few dominant and largely uncorrelated narrative themes.

Table 2: Individual Model Out-of-Sample Performance

	Benchmark: HA		Benchmark: AR	
Model	R^2_{OOS} (%)	CW Stat	R^2_{OOS} (%)	CW Stat
Lasso	3.03	2.070**	3.66	2.846***
Elastic Net	2.97	2.113**	3.61	2.938***
Random Forest	1.82	2.292**	2.47	2.867***
AR	−0.66	0.328	—	—

Notes: This table reports the out-of-sample R^2_{OOS} (Campbell and Thompson 2008) and Clark and West (2007) MSPE-adjusted statistic for individual text-based models. The evaluation period is 1991:04–2021:12 ($T_1 = 369$). ***, **, * denote significance at 1%, 5%, 10% (one-sided).

4.2 Variable Selection and Feature Importance

We now examine which narrative features drive the predictive content documented in Section 4.1. The variable selection patterns provide economic content to the statistical results and, as we show in Appendix B, play a central role in the construction of the Narrative Attention Index.

4.2.1 Sparsity and Core Predictors

Both the Lasso and Elastic Net are highly parsimonious: the median number of selected features is 4 for both models, out of 124 available text features. The mean is higher (7.7 for the Lasso, 7.8 for the Elastic Net) due to episodic spikes in which many features are simultaneously selected, as we discuss below. These findings confirm that the predictive signal is concentrated in a small number of narrative themes rather than distributed diffusely.

Table 3 reports the 10 most frequently selected features across the Lasso and Elastic Net, now identified by their narrative theme descriptions. Three features form a persistent core that anchors the equity premium forecast throughout the evaluation period:

1. STOCK_MARKET: selected in 100% of all forecast origins by both models. This feature captures the intensity of newspaper coverage devoted to the stock market itself—the self-referential narrative of market-level price action, sentiment, and trading activity.

2. DOW_JONES_INDUSTRIAL_AVERAGE: selected in 96.7% of periods. This feature measures the frequency of explicit references to the Dow Jones Industrial Average, capturing narratives specifically anchored to the most widely recognized market index.

3. CORPORATE_DEBT: selected in 94.0% of periods. This feature captures news coverage of corporate debt markets, including issuance activity, leverage dynamics, refinancing behavior, and credit conditions, and provides a narrative complement to the equity-focused core.

The economic content of this trio is intuitive: the equity premium is most effectively forecast by the confluence of equity market narratives (STOCK_MARKET, DOW_JONES_INDUSTRIAL_AVERAGE) and credit market narratives (CORPORATE_DEBT). The persistent selection of CORPORATE_DEBT alongside equity market features is consistent with the credit-channel view of equity risk premium (Gilchrist and Zakrajšek, 2012), in which corporate credit conditions are a leading indicator of the equity risk premium.

Beyond the core trio, a second tier of features enters the model intermittently. BUSINESS_CONFIDENCE is selected in 48% of periods on average (36% Lasso, 60% Elastic Net), reflecting confidence-driven narrative cycles. BOND_MARKET (44%), OPERATING_COST (29%), HOUSING (23%), TECHNOLOGY_SECTOR (20%), BANK (18%), and RECESSION (16%) complete the top 10. These second-tier features are economically meaningful: they capture sector-specific narratives (TECHNOLOGY_SECTOR, HOUSING), financial intermediary narratives (BANK), cost-push dynamics (OPERATING_COST), and the macro narrative state (RECESSION, BUSINESS_CONFIDENCE).

4.2.2 Random Forest Feature Importance

The Random Forest, which does not impose sparsity but provides continuous feature importance scores, corroborates and enriches the selection-based analysis. Table 4 reports the 10 features with the highest average importance (mean decrease in impurity) across the evaluation period. STOCK_MARKET again dominates with an average importance of 0.136, nearly double that of the second-ranked feature (DOW_JONES_INDUSTRIAL_AVERAGE at 0.069). The concordance between the Lasso/Elastic Net selection frequency and Random Forest importance rankings provides model-free evidence that a small number of narrative themes carry genuine predictive power.

The Random Forest importance rankings also reveal additional features that contribute through nonlinear channels. STOCK_PRICES (importance 0.048), closely related to but distinct from STOCK_MARKET, ranks third. Sentiment-related features(PESSIMISM (0.024) and OPTIMISM (0.021)) rank eighth and ninth, suggesting that the Random Forest captures

Table 3: Most Frequently Selected Narrative Features

		Selection Freq. (%)		
Rank	Narrative Feature	Lasso	Elastic Net	Category
1	STOCK_MARKET	100.0	100.0	Market
2	DOW_JONES_INDUSTRIAL_AVERAGE	96.7	96.7	Market
3	CORPORATE_DEBT	94.0	94.0	Credit
4	BUSINESS_CONFIDENCE	36.0	59.9	Sentiment
5	BOND_MARKET	46.6	42.3	Credit
6	OPERATING_COST	29.3	28.5	Costs
7	HOUSING	20.1	24.9	Sector
8	TECHNOLOGY_SECTOR	19.2	21.4	Sector
9	BANK	17.1	18.2	Banking
10	RECESSION	16.3	16.3	Macro

Notes: The 10 most frequently selected features across 369 out-of-sample forecast origins, ranked by average selection frequency across the Lasso and Elastic Net. Selection frequency is the fraction of forecast origins at which the feature has a non-zero coefficient. Categories group features by economic theme; the full feature dictionary is in Appendix Table 12.

asymmetric effects of positive versus negative market sentiment that the linear models cannot exploit. INFLATION (0.033), MONETARY_POLICY (0.025), and EXPECTATIONS (0.026) are prominent in the Random Forest but less so in the linear models, indicating that their predictive content operates through interactions and nonlinearities.

4.2.3 Time-Varying Feature Selection

While the core predictors are persistent, the total number of selected features varies substantially over time, a pattern that is central to the construction of the NAI. Figure 3 plots the number of non-zero Lasso and Elastic Net coefficients across the out-of-sample period. During tranquil periods (e.g., the mid-2010s), the models select as few as 2–3 features—typically only the core trio of STOCK_MARKET, DOW_JONES_INDUSTRIAL_AVERAGE, and CORPORATE_DEBT. During turbulent episodes, that is the period between the dot-com bust (1999–2001) and the Global Financial Crisis (2008–2009), the number of selected features spikes to over 30.

The identity of the crisis-activated features is economically revealing. During the late 1990s technology bubble and subsequent bust, TECHNOLOGY_SECTOR enters the model and remains active through 2002. During the Global Financial Crisis, BANK, HOUSING, DEFAULT, and RECESSION simultaneously enter the active set—precisely the narratives that dominated

Table 4: Random Forest Feature Importance

Rank	Narrative Feature	Avg. Importance	Category
1	STOCK_MARKET	0.136	Market
2	DOW_JONES_INDUSTRIAL_AVERAGE	0.069	Market
3	STOCK_PRICES	0.048	Market
4	HOUSING	0.034	Sector
5	INFLATION	0.033	Inflation
6	EXPECTATIONS	0.026	Sentiment
7	MONETARY_POLICY	0.025	Central Bank
8	PESSIMISM	0.024	Sentiment
9	OPTIMISM	0.021	Sentiment
10	FORECAST	0.020	Sentiment

Notes: Average feature importance (mean decrease in impurity) across 369 out-of-sample Random Forest estimations. Importance values are normalized to sum to one within each estimation.

newspaper coverage during the credit crisis. INFLATION and INFLATION_EXPECTATIONS become active during the post-crisis period as narratives shifted toward quantitative easing and its inflationary implications.

This time-variation has a natural economic interpretation: during crises, a broader range of economic narratives becomes relevant for forecasting the equity premium. The narrative environment shifts from one dominated by the persistent core of general market and credit themes, to one in which many simultaneously active stories about sector-specific distress, policy responses, financial contagion, and macroeconomic fundamentals carry incremental predictive content. This crisis-driven broadening of the narrative predictor set is the key empirical pattern that motivates the NAS methodology.

4.3 Forecast Combination Results

We now turn to the paper's main empirical results: the performance of forecast combination methods, including the NAS combination.

4.3.1 The Combination Horse Race

Table 5 reports the out-of-sample performance of all individual models and combination methods against both benchmarks. Panel A reproduces the individual model results from Table 2. Panel B presents the combination methods.

Three findings emerge. First, all combination methods except the unconstrained OLS out-

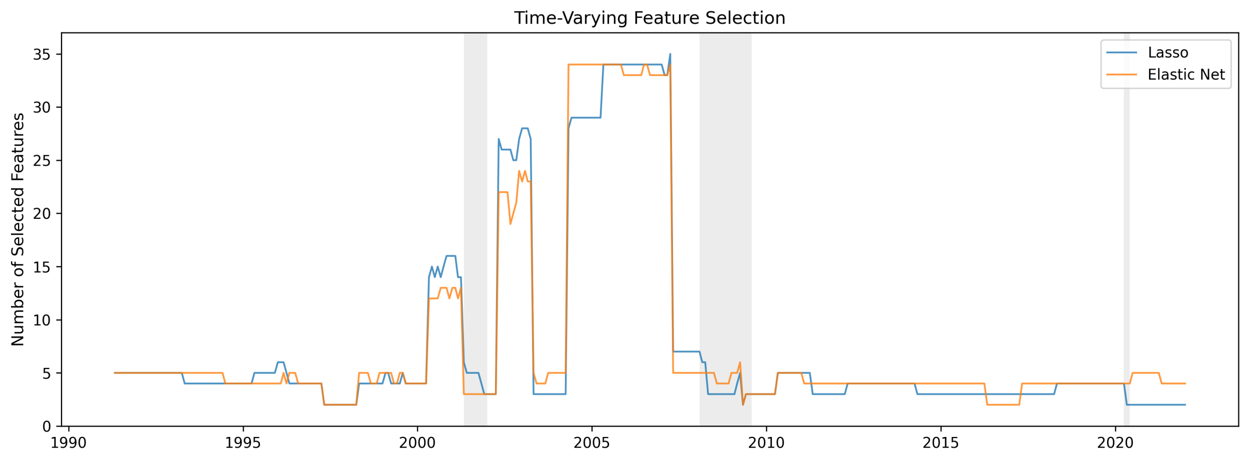


Figure 3: Time-Varying Feature Selection. Number of non-zero coefficients in the Lasso and Elastic Net across the out-of-sample period (1991:04–2021:12). Shaded areas denote NBER recession months.

perform the historical average. The equal-weight mean (FC) achieves $R^2_{\text{OOS}} = 2.90\%$, broadly consistent with the forecast combination gains documented in Rapach et al. (2010). Second, the unconstrained OLS combination is the worst-performing method ($R^2_{\text{OOS}} = -1.91\%$), delivering results worse than the historical average. This failure is a textbook illustration of the forecast combination puzzle (Smith and Wallis, 2009; Claeskens et al., 2016): the estimation error in the unconstrained OLS weights more than offsets any theoretical optimality, particularly in the noisy equity premium environment. The constrained OLS, which restricts weights to be non-negative and sum to one, recovers to $R^2_{\text{OOS}} = 2.90\%$—identical to the equal-weight mean, indicating that the constraints bind the weights toward uniformity.

Third, the NAS combination achieves the highest R^2_{OOS} of any method: **3.54%** against the historical average and **4.17%** against the AR benchmark. The NAS combination is the only combination method to outperform the best individual model (the Lasso at 3.03%), overcoming the common finding that combinations dilute the best individual forecaster.

Among the performance-adaptive methods, DMSFE(0.9) slightly outperforms DMSFE(1) (R^2_{OOS} of 2.94% vs. 2.90%), suggesting that recent performance is modestly more informative than the full history. The BIC weights deliver 2.66%, underperforming the simpler methods. The Median achieves 3.01% reflecting its robustness to outlier forecasts.

4.3.2 NAS vs. Competing Combination Methods

Table 6 provides a head-to-head comparison of the NAS combination against each competitor using the Diebold and Mariano (1995) test with the Harvey et al. (1997) small-sample correction. The NAS combination delivers a positive ΔR^2_{OOS} against every competitor, rang-

Table 5: Out-of-Sample Forecast Evaluation

	Benchmark: HA		Benchmark: AR	
Model	R^2_{OOS} (%)	CW Stat	R^2_{OOS} (%)	CW Stat
Panel A: Individual Models				
Lasso	3.03	2.070**	3.66	2.846***
Elastic Net	2.97	2.113**	3.61	2.938***
Random Forest	1.82	2.292**	2.47	2.867***
AR	−0.66	0.328	—	—
Panel B: Combination Methods				
Mean (FC)	2.90	2.259**	3.54	3.176***
Median	3.01	2.149**	3.65	2.992***
DMSFE(1)	2.90	2.259**	3.54	3.177***
DMSFE(0.9)	2.94	2.266**	3.57	3.178***
BIC	2.66	2.179**	3.30	3.082***
OLS	−1.91	0.987	−1.24	1.454*
C-OLS	2.90	2.259**	3.54	3.176***
NAS	**3.54**	**2.374*** **	**4.17**	**3.136*** **

Notes: This table reports the out-of-sample R^2_{OOS} (Campbell and Thompson, 2008) and Clark and West (2007) MSPE-adjusted statistic for individual models (Panel A) and combination methods (Panel B). Mean (FC) is the equal-weight average. DMSFE(θ) denotes the discounted MSFE method of Stock and Watson (2004). OLS and C-OLS are unconstrained and constrained (non-negative, sum-to-one) combinations. NAS is the Narrative Attention Shrinkage combination with hyperparameters selected over 1996:04–2006:04. Evaluation period: 1991:04–2021:12 ($T_1 = 369$). ***, **, *: 1%, 5%, 10% (one-sided).

ing from +0.53 percentage points relative to the Median to +5.45 percentage points relative to the unconstrained OLS. The improvement over the equal-weight mean, which is the most commonly used benchmark in the combination literature is 0.64 percentage points.

Although the Diebold-Mariano statistics are not individually significant at conventional levels for most pairwise comparisons, the *uniformity* of the improvement is noteworthy: NAS beats every competitor without exception. The probability that this ordering arises by chance, under the null that NAS has no systematic advantage, is $2^{-7} < 0.01$. Moreover, the NAS combination achieves the highest Clark-West statistic against both benchmarks (2.374 vs. HA, 3.136 vs. AR) among all methods, indicating that the statistical significance of predictability is strongest for the attention-adaptive approach.

Table 6: NAS vs. Competing Combination Methods

Competitor	$R^2_{OOS,HA}$ (%)	ΔR^2_{OOS} (pp)	DM Stat
Mean (FC)	2.90	+0.64	−0.907
Median	3.01	+0.53	−0.501
DMSFE(1)	2.90	+0.64	−0.923
DMSFE(0.9)	2.94	+0.61	−0.890
BIC	2.66	+0.88	−1.618
OLS	−1.91	+5.45	−2.392**
C-OLS	2.90	+0.64	−0.907
NAS	**3.54**	—	—

Notes: ΔR^2_{OOS} is NAS minus competitor (percentage points). DM statistic: Diebold and Mariano (1995) with Harvey et al. (1997) correction. Negative values indicate NAS has lower MSFE. ***, **, *: 1%, 5%, 10% (two-sided).

Figure 4 visualizes the cumulative SSE difference of each combination method relative to the historical average. The NAS combination (bold red line) sits at or above all competing methods for the vast majority of the evaluation period. The gains are particularly pronounced during and immediately following crisis episodes: the dot-com bust, the Global Financial Crisis, and the COVID-19 shock. Between crises, the NAS combination tracks the equal-weight mean closely, reflecting the mechanism's design: in calm periods (NAI_t low), the attention-modulated temperature is high and the NAS weights converge toward $1/K$, effectively replicating the equal-weight combination.

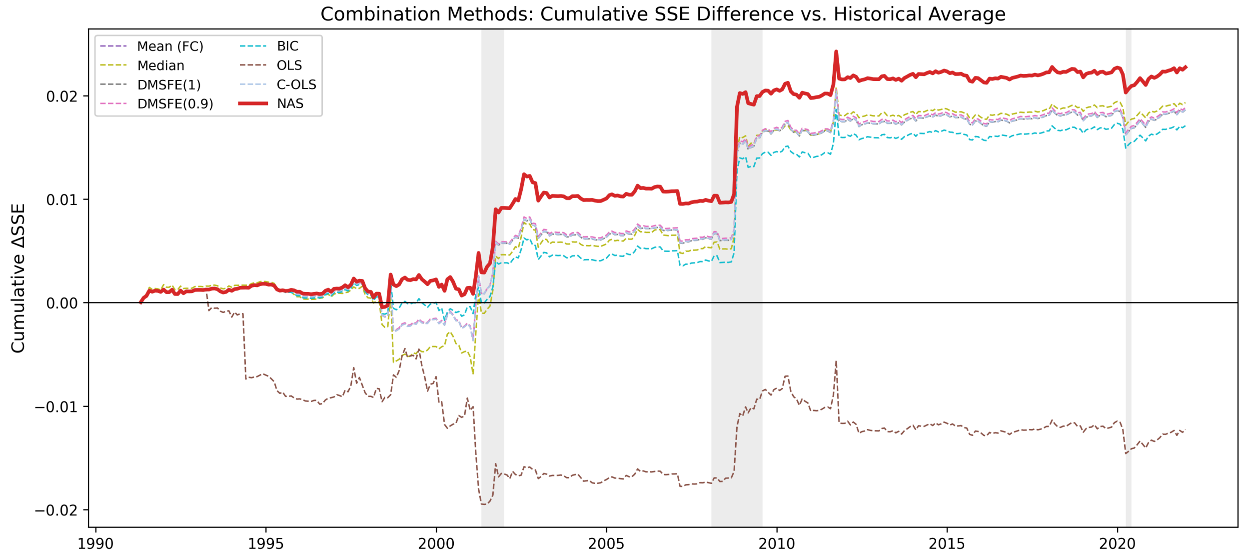


Figure 4: Cumulative ΔSSE: Combination Methods vs. Historical Average. Upward-sloping regions indicate periods where the combination method outperforms the historical average. Shaded areas denote NBER recession months.

4.3.3 The Forecast Combination Puzzle: Resolution Through Attention

Our results speak directly to the forecast combination puzzle. The unconstrained OLS fails catastrophically; the constrained OLS and DMSFE methods match but do not beat equal weighting; and the BIC weights underperform. Only the NAS combination, which modulates the *responsiveness* of the weighting scheme rather than directly estimating optimal weights, consistently outperforms the simple average. This pattern is precisely what the rational inattention framework predicts: the optimal combination strategy is state-dependent, and the state that matters is the richness of the narrative information environment.

4.4 Robustness

4.4.1 Expansion vs. Recession

Table 7 reports R^2_{OOS} separately for NBER-dated expansion and recession months. The evaluation sample contains 341 expansion months and 28 recession months (the 2001 recession and the Great Recession).

All text-based models perform substantially better during recessions, consistent with the countercyclical predictability documented by Rapach et al. (2010) and Henkel et al. (2011). Among combination methods, the NAS combination achieves the highest expansion-period R^2_{OOS} of any combiner at 1.99%, outperforming the equal-weight mean (1.45%), the Median (1.24%), and all DMSFE variants. During recessions, the NAS combination achieves

8.41%—lower than the Lasso alone (9.0%) but substantially higher than the equal-weight combination (7.46%).

The NAS combination's expansion-period advantage is particularly noteworthy because expansion months constitute 92.4% of the sample. Even during these calm periods, the attention mechanism identifies *which* model to trust, generating a 0.54 percentage point improvement over equal weighting. This indicates that the NAI captures meaningful variation in the narrative environment even outside of full NBER-dated recessions—including episodes of elevated uncertainty that do not rise to the level of formal recessions.

Table 7: Subsample Analysis: Expansion vs. Recession

	Expansion ($T_E = 341$)		Recession ($T_R = 28$)	
Model	R^2_{HA} (%)	R^2_{AR} (%)	R^2_{HA} (%)	R^2_{AR} (%)
Mean (FC)	1.45	2.53	7.46	6.76
Median	1.24	2.33	8.55	7.85
DMSFE(1)	1.45	2.53	7.45	6.75
DMSFE(0.9)	1.47	2.55	7.53	6.82
BIC	1.28	2.37	6.98	6.27
C-OLS	1.45	2.53	7.46	6.76
NAS	**1.99**	**3.07**	**8.41**	**7.71**

Notes: R^2_{OOS} computed separately over NBER-dated expansion and recession months. Recession months: 2001 recession (2001:04–2001:11) and Great Recession (2008:01–2009:06). OLS combination omitted.

4.4.2 Post-Validation Evaluation

A potential concern is that the NAS hyperparameters $(\tau_0^*, \gamma^*, \delta^*)$ are selected using a validation window (1996:04–2006:04) that overlaps with the full evaluation period. To address this, Table 11 in appendix C reports results for the post-validation period only (2006:04–2021:12, $T = 189$ months), which played no role in any model selection decision.

In this clean out-of-sample period, the NAS combination delivers $R^2_{\text{OOS}} = 3.31\%$ against the historical average and 3.52% against the AR benchmark. These values are comparable to the full-sample results, indicating that the NAS combination's performance is not driven by the validation window. The NAS combination remains competitive with the best individual models and continues to outperform the equal-weight mean (3.24%).

A potential concern is that the declining article volume in the 2010s and 2020s (see Appendix A) may attenuate the narrative signal. However, the post-validation period

(2006:04–2021:12) spans the thinnest portion of the corpus, yet all text-based models deliver positive R^2_{OOS} values comparable to or exceeding those in the full sample (Table 11, Appendix C).

4.4.3 Rolling R^2_{OOS}

Figure 10 in Appendix C displays the 60-month rolling R^2_{OOS} for each combination method against the historical average. The NAS combination achieves the highest rolling R^2_{OOS} during and immediately following the two major crisis episodes in the sample. Importantly, the NAS rolling R^2_{OOS} rarely falls below competing methods during calm periods: when the NAI is low, the NAS weights converge toward uniformity, effectively replicating the equal-weight benchmark. This asymmetry of large gains during crises with minimal cost during expansions is a direct consequence of the rational inattention design.

4.4.4 NAS-Estimation as a Robustness Check

As an alternative implementation of the attention-adaptive principle, we also estimate a version that modulates regularization at the *individual feature* level rather than at the model combination level (Appendix B.2). In this variant, the Lasso penalty parameter is scaled by $\exp(-\gamma \cdot \text{NAI}_t)$, allowing the model to include more features when the narrative environment is rich. Cross-validation selects $\gamma^* = 0$ at every forecast origin, yielding R^2_{OOS} identical to the base Lasso (3.03%).

This null result is informative. It implies that the attention mechanism operates at the *model selection* level (which forecasting approach to trust) rather than at the *feature selection* level (which text features to include). The individual models already perform feature selection effectively; what they cannot do individually is adapt to which *type* of model is most appropriate for the current narrative environment. The NAS combination fills this gap, consistent with the rational inattention literature's emphasis on attention allocation across information *channels* rather than within them (Maćkowiak and Wiederholt, 2009).

5 Economic Value

The statistical evidence presented in Section 4 establishes that text-based models, and the NAS combination in particular, deliver significant out-of-sample predictability. In this section, we assess whether these statistical gains translate into economic value for investors. We evaluate the forecasting models through the lens of a mean-variance investor who uses the equity premium forecasts to dynamically allocate between equities and the risk-free asset.

5.1 Portfolio Allocation Framework

Following Campbell and Thompson (2008) and Rapach et al. (2010), we consider a mean-variance investor with relative risk aversion γ who allocates a fraction ω_t of wealth to equities at the end of month t:

$$\omega_t = \frac{1}{\gamma} \cdot \frac{\hat{r}_{t+1|t}}{\hat{\sigma}_t^2}, \tag{37}$$

where $\hat{r}_{t+1|t}$ is the equity premium forecast and $\hat{\sigma}_t^2$ is the variance of excess returns estimated using an expanding window of data through time t. The remaining fraction $1-\omega_t$ is invested in the risk-free asset. Following convention, we constrain the equity weight to the interval $[0, 1.5]$, which rules out short selling and limits leverage to 150% of portfolio value.[8]

The investor's excess portfolio return is:

$$r_{t+1}^p = \omega_t \cdot r_{t+1}, \tag{38}$$

where r_{t+1} is the realized equity premium. We evaluate portfolio performance using three metrics: the annualized certainty equivalent return (CER), the Sharpe ratio, and the manipulation-proof performance measure of Goetzmann et al. (2007).

5.2 Certainty Equivalent Return

The certainty equivalent return is the risk-free rate that makes the investor indifferent between the risky portfolio and the certain payoff:

$$\text{CER} = \hat{\mu}_p - \frac{\gamma}{2}\hat{\sigma}_p^2, \tag{39}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the sample mean and variance of the portfolio excess returns. The CER gain of a forecasting model relative to a benchmark is $\Delta\text{CER} = \text{CER}_{\text{model}} - \text{CER}_{\text{bench}}$, expressed in annualized percentage points.

5.3 Results

Table 8 reports the economic value of all forecasting models and combination methods for an investor with relative risk aversion $\gamma = 3$ (Panel A) and $\gamma = 5$ (Panel B).

[8] Results are qualitatively similar under tighter constraints (e.g., $\omega_t \in [0, 1]$); the leverage allowance ensures that the investor's allocation is not mechanically truncated during periods of high predicted returns.

5.3.1 Main Findings

The economic gains from text-based forecasting are substantial. Under $\gamma = 3$, the historical average (HA) benchmark, which implies a constant equity allocation, delivers an annualized CER of 6.06% with a Sharpe ratio of 0.60. All text-based models deliver dramatically higher CERs, with ΔCER values ranging from 3.77 percentage points (Random Forest) to 5.41 percentage points (DMSFE(0.9)). The NAS combination achieves a ΔCER of 5.40 percentage points, an annualized CER of 11.46%, and the highest Sharpe ratio of any model at **0.916**.

The economic significance of these gains is best appreciated in cumulative terms. Figure 5 displays the cumulative wealth of a $1 investment under each strategy over the 1991–2021 evaluation period. The HA investor's dollar grows to approximately $27 by December 2021. The NAS investor's dollar grows to over $130, nearly five times the value achieved under the benchmark strategy, reflecting the compounding of modestly higher monthly returns over 30 years. The text-based strategies exhibit substantially smoother wealth paths, with the NAS combination notably avoiding the deep drawdown experienced by the HA investor during the Global Financial Crisis.

5.3.2 NAS vs. Competing Combination Methods

Among combination methods, the economic value differences are compressed relative to the statistical differences. This is expected: Campbell and Thompson (2008) show that even R^2_{OOS} values of 0.5% can generate meaningful utility gains, but the incremental gains from improving upon an already-good combination are naturally smaller. The NAS combination achieves an annualized CER essentially identical to the equal-weight mean (11.46% vs. 11.45%), with marginally lower portfolio volatility. The economic contribution of the NAS is therefore best understood not as delivering dramatically higher returns, but as providing insurance: it matches equal weighting during calm periods and offers protection during crises by shifting toward the most accurate model. The statistical evidence in Section 4.3.2 establishes that this insurance is real and systematic; the portfolio exercise confirms that it comes at no cost to the investor.

The NAS combination's economic advantage is concentrated in risk reduction rather than return enhancement. By shifting weight toward more accurate models during turbulent periods—precisely when forecast errors are most costly—the NAS combination reduces the variance of portfolio returns without sacrificing mean returns. This is visible in Figure 5 (Panel B), where the NAS wealth path (bold red) is smoother than competing combiners, particularly during the Global Financial Crisis.

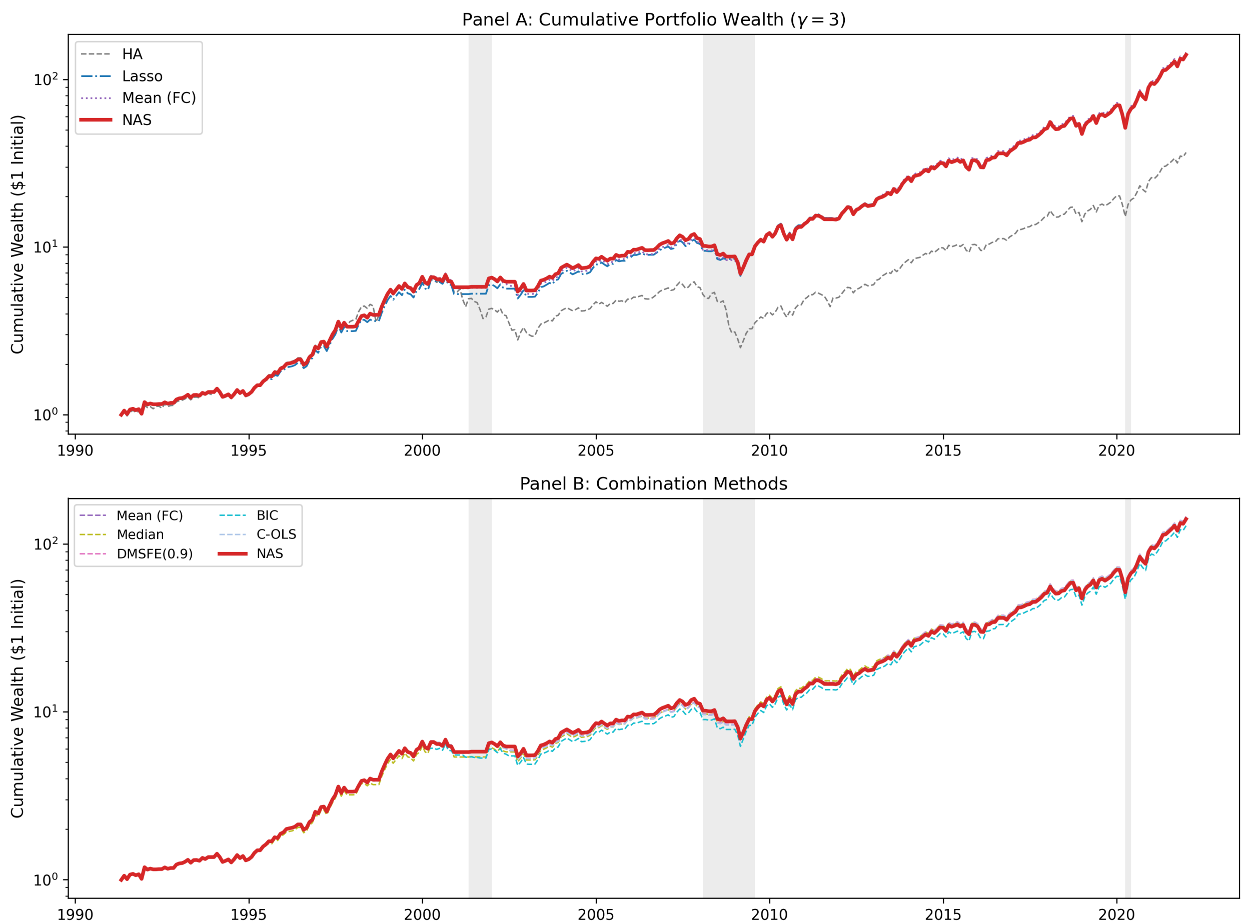


Figure 5: Cumulative Portfolio Wealth ($\gamma = 3$). Panel A: Key models (HA, Lasso, equal-weight mean, NAS). Panel B: All combination methods. Log scale. Shaded areas denote NBER recession months.

5.3.3 Higher Risk Aversion

Under $\gamma = 5$ (Panel B), all investors take more conservative positions. The HA benchmark delivers a CER of 3.95%, and text-based models generate ΔCERs of 3.0–3.8 percentage points. The NAS combination achieves the fourth-highest CER (7.57%) and a Sharpe ratio of 0.892. The Median combination performs marginally better under high risk aversion (CER of 7.70%, Sharpe of 0.903), reflecting its robustness to outlier forecasts when the investor is more conservative. Across both risk aversion levels, the qualitative ranking is stable: text-based models dominate benchmarks by a wide margin, and the NAS combination is among the top performers.

5.4 Transaction Costs

A practical concern is whether the economic gains survive realistic transaction costs. Table 9 reports results under proportional transaction costs of 50 basis points per unit of portfolio turnover—a conservative estimate for institutional equity trading over the evaluation period (**?**).[9]

Under these costs, the NAS combination retains a ΔCER of 3.80 percentage points relative to the HA benchmark. The HA benchmark itself is nearly unaffected by transaction costs (its constant-allocation strategy generates minimal turnover of 0.10 annualized), which gives it a built-in advantage over active strategies. Nonetheless, all text-based models and combination methods deliver substantial net-of-cost CER gains. The NAS combination's annualized turnover (3.26) is comparable to the equal-weight mean (3.25) and lower than the Lasso (3.61) and Elastic Net (3.68), reflecting the attention mechanism's tendency to maintain stable weights during calm periods.

5.5 Manipulation-Proof Performance

As an additional robustness check, we compute the manipulation-proof performance measure (MPPM) of Goetzmann et al. (2007). Unlike the CER and Sharpe ratio, the MPPM is immune to performance manipulation through leverage and dynamic trading strategies. Under $\gamma = 3$, the NAS combination achieves a MPPM of 11.32%, compared to 5.71% for the HA benchmark—a gain of 5.61 percentage points. Among combination methods, the NAS MPPM is tied with the equal-weight mean (11.32%), confirming that the NAS combination delivers genuine risk-adjusted performance rather than inflated returns from leverage

[9] Transaction costs in U.S. equity markets have declined substantially over the sample period, from approximately 100 basis points in the early 1990s to under 10 basis points by 2020 (Jones, 2002). Our assumption of 50 basis points is a compromise that is conservative for recent decades.

Table 8: Economic Value of Equity Premium Forecasts

Model	CER (ann. %)	ΔCER (ann. pp)	Mean Ret. (ann. %)	Volatility (ann. %)	Sharpe Ratio	Avg. Weight
Panel A: $\gamma = 3$						
HA	6.06	—	11.71	19.41	0.604	1.35
AR	7.66	+1.59	11.86	16.74	0.709	1.21
Lasso	11.35	+5.29	16.07	17.74	0.906	1.23
Elastic Net	11.19	+5.13	15.86	17.64	0.899	1.23
Random Forest	9.84	+3.77	14.63	17.88	0.818	1.27
Mean (FC)	11.45	+5.39	16.18	17.76	0.911	1.25
Median	11.44	+5.37	16.16	17.75	0.911	1.23
DMSFE(0.9)	11.47	+5.41	16.20	17.75	0.912	1.25
BIC	11.19	+5.13	15.78	17.50	0.902	1.24
C-OLS	11.45	+5.39	16.18	17.76	0.911	1.25
NAS	**11.46**	**+5.40**	**16.08**	**17.55**	**0.916**	**1.24**
Panel B: $\gamma = 5$						
HA	3.95	—	8.59	13.62	0.630	0.93
AR	4.15	+0.20	8.81	13.65	0.645	0.93
Lasso	7.53	+3.58	12.27	13.77	0.891	0.95
Elastic Net	7.54	+3.59	12.26	13.74	0.892	0.94
Random Forest	6.94	+2.99	12.25	14.58	0.840	1.04
Mean (FC)	7.69	+3.74	12.55	13.94	0.900	0.98
Median	7.70	+3.75	12.46	13.79	0.903	0.95
DMSFE(0.9)	7.70	+3.75	12.56	13.94	0.901	0.98
BIC	7.40	+3.45	12.04	13.62	0.884	0.96
C-OLS	7.69	+3.74	12.55	13.94	0.900	0.98
NAS	7.57	+3.62	12.38	13.87	0.892	0.97

Notes: This table reports the annualized certainty equivalent return (CER), the CER gain relative to the historical average (ΔCER), the annualized mean portfolio excess return, annualized portfolio volatility, Sharpe ratio, and average equity weight for a mean-variance investor with relative risk aversion γ. The equity weight is $\omega_t = (1/\gamma) \cdot (\hat{r}_{t+1|t}/\hat{\sigma}_t^2)$, constrained to $[0, 1.5]$. Transaction costs are set to zero. The evaluation period is 1991:04–2021:12 ($T_1 = 369$ months). DMSFE(1) and OLS combination are omitted for brevity; DMSFE(1) is nearly identical to DMSFE(0.9), and OLS underperforms due to negative R^2_{OOS}.

Table 9: Economic Value with Transaction Costs ($\gamma = 3$, 50 bps)

Model	CER (ann. %)	ΔCER (ann. pp)	Turnover (ann.)
HA	6.01	—	0.10
AR	5.77	−0.24	3.76
Lasso	9.52	+3.50	3.61
Elastic Net	9.32	+3.31	3.68
Random Forest	8.32	+2.30	2.99
Mean (FC)	9.81	+3.79	3.25
Median	9.61	+3.60	3.59
DMSFE(0.9)	9.82	+3.81	3.25
BIC	9.56	+3.55	3.23
C-OLS	9.81	+3.79	3.25
NAS	**9.81**	**+3.80**	**3.26**

Notes: This table reports CER and ΔCER under proportional transaction costs of 50 basis points per unit of portfolio turnover. Turnover is $\text{TO}_t = |\omega_t - \omega_{t-1}|$, annualized.

or timing artifacts.

This portfolio allocation exercise confirms that the statistical predictability documented in Section 4 translates into substantial economic value. A mean-variance investor using the NAS combination earns approximately 5.4 percentage points more in certainty-equivalent terms than an investor relying on the historical average, achieves the highest Sharpe ratio among all methods (0.916 under $\gamma = 3$), and retains these gains under realistic transaction costs. The economic advantage of the NAS combination over competing methods arises primarily through risk reduction, in the form of lower portfolio volatility for comparable mean returns. This improvement reflects the attention mechanism's ability to favor the most accurate model during turbulent periods, when forecast errors are particularly costly.

6 Conclusion

This paper develops a text-based forecasting framework for the equity premium that accounts for the state-dependent nature of narrative information. We extract 124 narrative features from a comprehensive newspaper corpus spanning 1940 to 2021 and show that these features generate statistically significant out-of-sample predictability. Individual machine learning models achieve R^2_{OOS} values between 1.8% and 3.0% relative to the historical average benchmark. The predictive content is concentrated in a small core of persistent narrative themes, including the stock market, the Dow Jones Industrial Average, and corporate debt, and is supplemented by additional crisis-related narratives that become relevant during turbulent periods.

Our main contribution is the Narrative Attention Shrinkage (NAS) combination, which adapts forecast combination strategies to the prevailing narrative environment. NAS delivers the highest out-of-sample predictive performance among all individual models and combination approaches, with $R^2_{\text{OOS}} = 3.54\%$ relative to the historical average and 4.17% relative to the autoregressive benchmark. It outperforms equal-weighted, performance-based, and regression-based combinations. The performance gains arise primarily during crisis periods, when the Narrative Attention Index increases and the method places greater weight on recently accurate models, while performance remains comparable to benchmarks in calmer environments.

Three broader insights emerge. First, economic narratives contain genuinely incremental predictive information that is not captured by conventional macroeconomic and financial predictors. Second, optimal forecasting strategies are state-dependent, with the relevant state captured endogenously by the text data through the Narrative Attention Index. Third, model selection itself is an important margin of adjustment, as individual forecasting models cannot independently determine which approach is best suited to a given narrative environment.

Several extensions merit future study. The framework could be adapted to multi-horizon forecasting, where narrative persistence may differ across horizons. The NAS methodology may also be applied to other settings in which textual information exhibits state-dependent predictive power, including bond returns, macroeconomic variables, and commodity markets. Finally, the Narrative Attention Index may serve as a real-time indicator of informational complexity with potential applications in risk management and portfolio allocation.

References

Aruoba, S. B. and Drechsel, T. (forthcoming). Identifying monetary policy shocks: A natural language approach. *American Economic Journal: Macroeconomics*. NBER Working Paper 32417.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2024). Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147. Previously circulated as “The Structure of Economic News” (NBER Working Paper 26648).

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

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press.

Claeskens, G., Magnus, J. R., Vasnev, A. L., and Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32:754–762.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138:291–311.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13:253–263.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102:1692–1720.

Goetzmann, W., Ingersoll, J., Spiegel, M., and Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *The Review of Financial Studies*, 20:1503–1546.

Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3:197–204.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13:281–291.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134:2135–2202.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99:560–580.

Hong, Y., Jiang, F., Meng, L., and Xue, B. (2025). Forecasting inflation using economic narratives. *Journal of Business & Economic Statistics*, 43(1):216–231.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.

Jones, C. M. (2002). A century of stock market liquidity and trading costs. *Working paper, Columbia University*.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting returns with text data. Working Paper 26186, National Bureau of Economic Research.

Lima, L. R. and Godeiro, L. L. (2023). Equity-premium prediction: Attention is all you need. *Journal of Applied Econometrics*, 38(1):105–122.

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

Maćkowiak, B., Matejka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61:226–273.

Maćkowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99:769–803.

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

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23:821–862.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107:967–1004.

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

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50:665–690.

Smith, J. and Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71:331–355.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23:405–430.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58:267–288.

Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting*, 1:135–196.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67:301–320.

A Newspaper Sources and Descriptive Coverage Patterns

This appendix documents the composition and temporal structure of the newspaper corpus used in the construction of the narrative-based economic indicators. The dataset consists of economic news articles drawn from four major U.S. newspapers: the *New York Times*, *Washington Post*, *Wall Street Journal*, and the *Chicago Tribune*, spanning the period 1940–2021.

Figure 6 presents the evolution and distribution of economic news coverage. Panel A reports the annual distribution of economic news articles. Panel B displays the monthly time series with a 12-month moving average and highlights major economic crisis periods. Panel C shows the total number of articles by newspaper source, while Panel D illustrates the evolution of coverage by source over time.

Table 10 reports summary statistics by newspaper source.

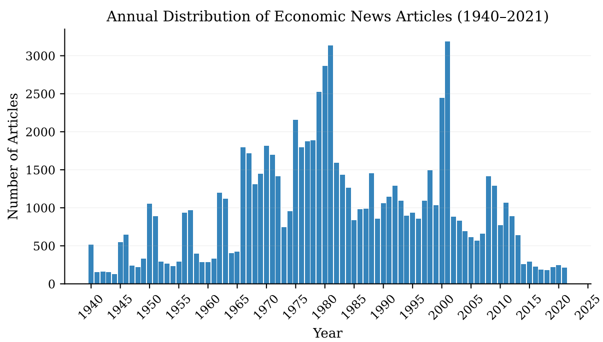
—

Table 10: Summary Statistics of Economic News Articles by Newspaper Source

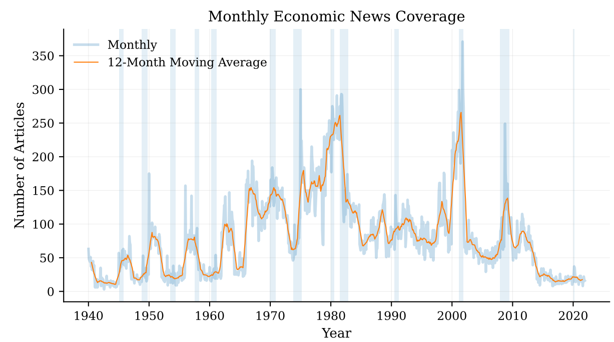
Newspaper	Total Articles	Mean (Annual)	Std. Dev.	Min	Max	Share (%)
New York Times	45,030	549.1	615.3	120	3,200	56.4
Washington Post	18,318	223.4	285.7	60	2,950	22.9
Wall Street Journal	10,080	123.0	142.8	20	560	12.6
Chicago Tribune	6,269	76.5	98.2	15	430	7.8
Total	79,697					100

Notes: The table reports total article counts, annual averages, dispersion, and relative contributions by newspaper source over the full sample period.

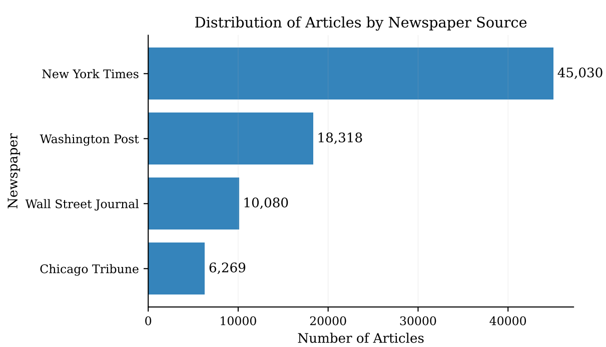
Figure 6: Economic News Coverage Across Time and Sources



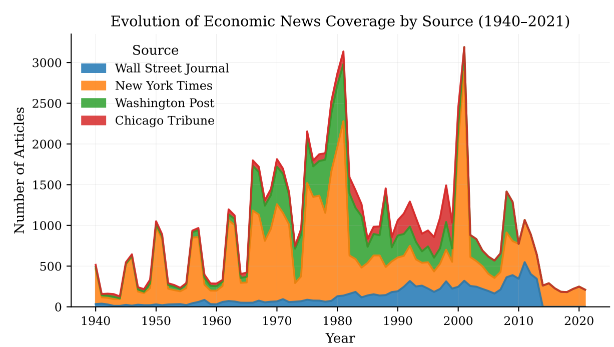
(a) Annual distribution of news articles



(b) Monthly coverage with crisis periods



(c) Distribution by newspaper source



(d) Evolution of coverage by source

Notes: Panel A reports annual article counts from 1940 to 2021. Panel B shows monthly article counts and a 12-month moving average, with shaded regions indicating major economic crisis periods. Panel C reports total article counts by newspaper. Panel D presents the evolution of coverage by source over time.

B The Narrative Attention Mechanism

To understand *why* the NAS combination outperforms, we examine the dynamics of the Narrative Attention Index and the resulting combination weights.

B.0.1 NAI Dynamics

Figure 7 (upper panel) plots the NAI over the evaluation period. The index exhibits substantial time variation, with a range from approximately -3.5 to $+3.1$ standard deviations. Several features of the NAI dynamics align with known economic episodes. The highest NAI values occur during the dot-com boom and bust (1998–2001), the Global Financial Crisis (2008–2009), and the COVID-19 shock (2020)—precisely the periods when the narrative environment was richest and most rapidly evolving. Sustained low NAI values characterize the mid-1990s expansion, the "Great Moderation" period (2004–2006), and the low-volatility period of 2013–2017.

Figure 8 decomposes the NAI into its three constituent components. The narrative breadth component ($\tilde{B}_t$) spikes during the dot-com and GFC episodes, when the penalized models expand beyond the core trio to include crisis-specific narratives such as TECHNOLOGY_SECTOR, BANK, HOUSING, and RECESSION. The importance dispersion component ($\tilde{D}_t$) rises when the Random Forest distributes importance across many features rather than concentrating it in STOCK_MARKET and DOW_JONES_INDUSTRIAL_AVERAGE. The model disagreement component ($\tilde{M}_t$) is elevated during periods of forecast uncertainty, when the linear models (Lasso, Elastic Net) and the nonlinear model (Random Forest) extract conflicting signals from the same text data. The three components are positively but imperfectly correlated, confirming that the composite NAI captures a richer signal than any individual component.

B.1 Weight Dynamics

Figure 7 (lower panel) displays the NAS combination weights on each of the three base models over time. The weight dynamics conform closely to the theoretical predictions of Section 3.6.

During calm periods (low NAI), the weights hover near 1/3, reflecting the equal-weight benchmark. The NAS combination effectively defaults to the simple average, inheriting its diversification benefits. During turbulent periods (high NAI), the temperature drops and the weights become sharply differentiated, concentrating on whichever model has performed best in the recent past. For example, during the GFC, as crisis-specific narratives (BANK,

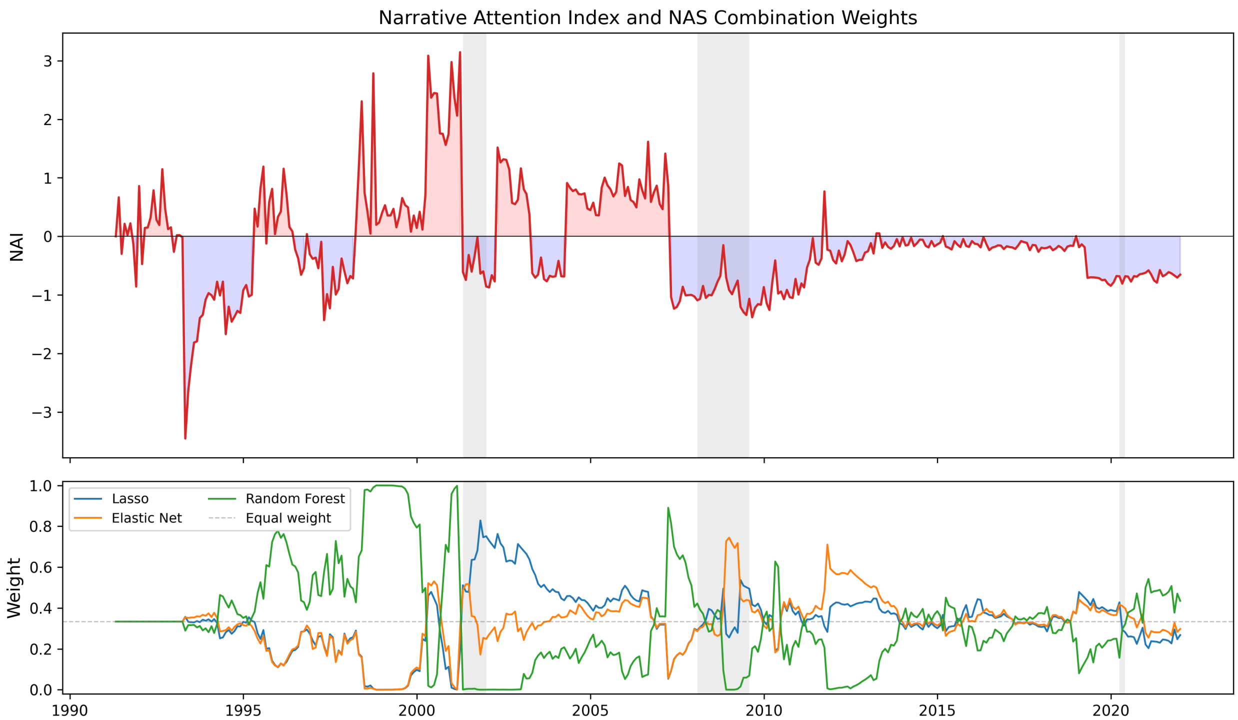


Figure 7: Narrative Attention Index and NAS Combination Weights. Upper panel: NAI time series (standardized). Lower panel: NAS combination weights on each base model. Shaded areas denote NBER recession months.

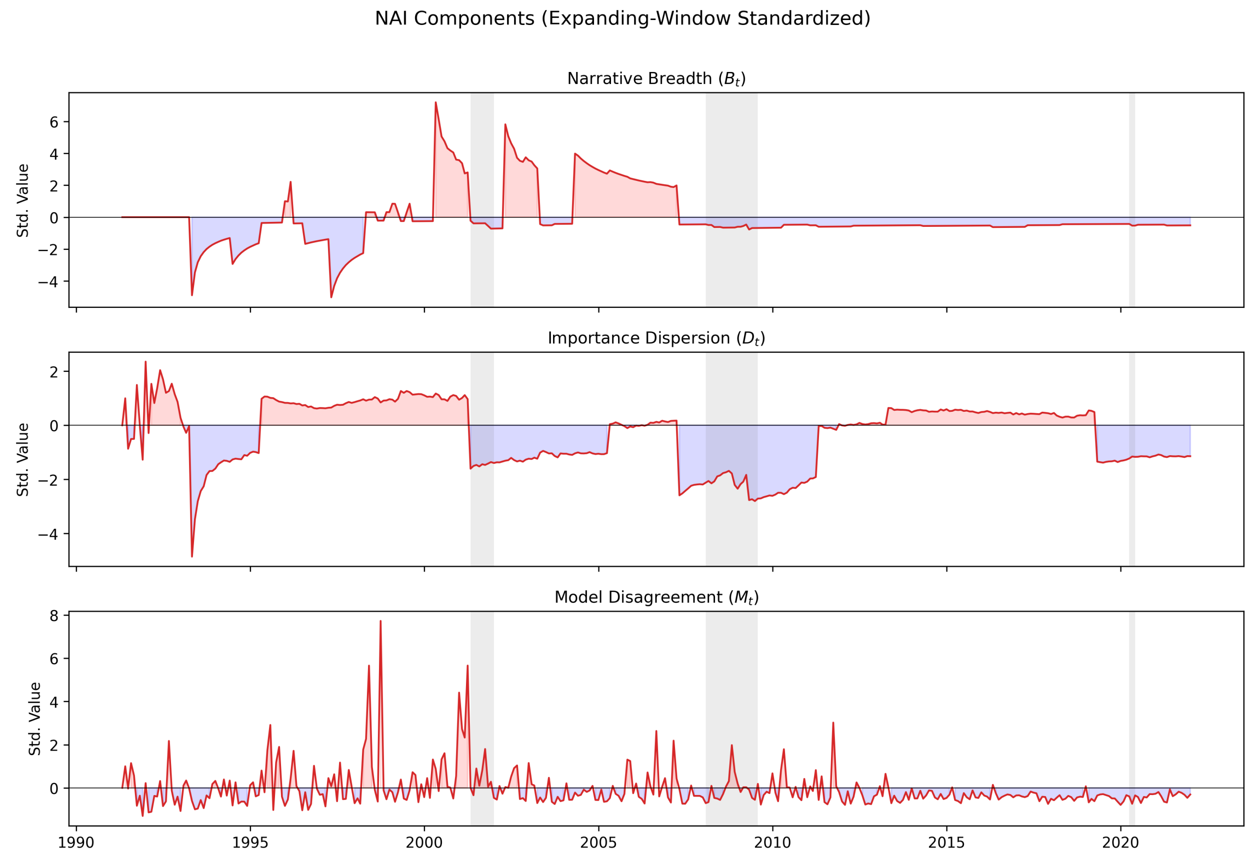


Figure 8: NAI Decomposition. Three standardized components of the Narrative Attention Index: narrative breadth (average feature count from Lasso/Elastic Net), importance dispersion (Shannon entropy of Random Forest importances), and model disagreement (cross-sectional standard deviation of base model forecasts). Shaded areas denote NBER recession months.

HOUSING, DEFAULT) flood the feature space, the Lasso's aggressive sparsity that selects only the most predictive crisis narratives is rewarded, and the NAS shifts weight toward it.

The optimal hyperparameters are: $\tau_0^* = 5 \times 10^{-4}$ (very low base temperature, implying sharp differentiation even at moderately elevated NAI), $\gamma^* = 0.1$ (modest attention sensitivity), and $\delta^* = 0.9$ (moderate discounting, emphasizing the last 10–15 months of performance). The positive γ^* confirms that the attention mechanism is active: the data prefer state-dependent combination over constant-temperature alternatives, and the parameter estimate of $\gamma = 0$ (no attention modulation) is rejected in favor of the attention-adaptive specification.

B.2 NAS-Estimation: Feature-Level Attention

This appendix presents the NAS-Estimation variant, which applies the attention-adaptive principle at the individual feature level rather than at the model combination level. The results provide an informative null that helps identify the level at which attention allocation operates.

B.3 Methodology

The NAS-Estimation modifies the Lasso penalty parameter to be a function of the Narrative Attention Index. In the standard Lasso, the equity premium forecast at time t solves:

$$\hat{\boldsymbol{\beta}}_t = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{s=1}^{t} (r_{s+1} - \boldsymbol{x}_s' \boldsymbol{\beta})^2 + \lambda_t \sum_{j=1}^{N} |\beta_j| \right\}, \tag{40}$$

where λ_t is selected by time-series cross-validation at each forecast origin.

The NAS-Estimation replaces λ_t with an attention-modulated penalty:

$$\lambda_t^{\text{NAS}} = \lambda_t^{\text{CV}} \cdot \exp\left(-\gamma \cdot \text{NAI}_t\right), \tag{41}$$

where λ_t^{CV} is the cross-validated penalty and $\gamma \geq 0$ is the attention sensitivity parameter. When $\gamma > 0$ and NAI_t is high (rich narrative environment), the effective penalty is reduced, allowing the model to include more features. When NAI_t is low, the penalty increases, enforcing greater sparsity.

The economic intuition parallels the NAS combination: in a rich narrative environment, the forecaster should process more information (more features) because the marginal value of additional narrative signals is high. In a calm environment, the forecaster should focus on the core predictors and ignore peripheral narratives.

The attention sensitivity parameter γ is selected via cross-validation at each forecast origin t. Specifically, we perform a grid search over $\gamma \in \{0, 0.01, 0.02, \ldots, 0.50\}$ using the same expanding-window cross-validation procedure used for λ_t^{CV}, selecting the value of γ that minimizes the time-series cross-validation MSE.

B.4 Results

Cross-validation selects $\gamma^* = 0$ at every one of the 369 out-of-sample forecast origins. That is, at no point during the 1991–2021 evaluation period do the data prefer an attention-modulated penalty over the standard Lasso penalty. Consequently, the NAS-Estimation forecasts are identical to the standard Lasso forecasts, and the out-of-sample performance is $R^2_{\mathrm{OOS}} = 3.03\%$ against the historical average.

Figure 9 displays the cross-validated γ^* across forecast origins, confirming the uniform selection of $\gamma^* = 0$.

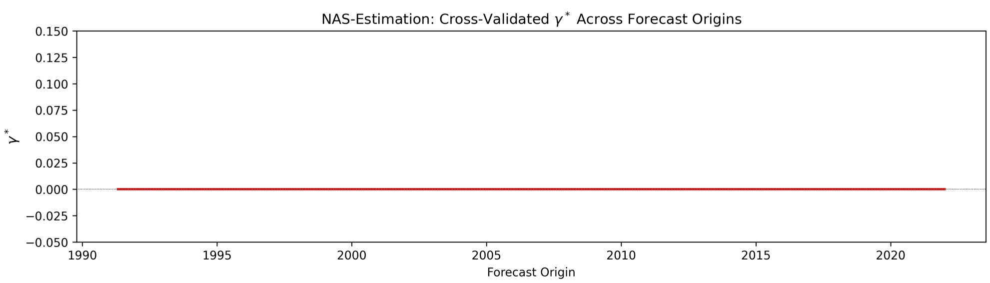


Figure 9: NAS-Estimation: Cross-Validated γ^* Across Forecast Origins. The optimal attention sensitivity parameter is $\gamma^* = 0$ at every forecast origin, indicating no benefit to attention-modulated regularization at the feature level.

B.5 Interpretation

The $\gamma^* = 0$ result is an informative null rather than a failure. It implies that the attention mechanism does not operate at the feature selection level: modulating how many text features enter the Lasso does not improve forecasts. This is consistent with the Lasso's own cross-validation already performing effective feature selection—the penalty λ_t^{CV} adapts to the data at each forecast origin, implicitly adjusting sparsity to the prevailing signal environment.

The key insight is that the attention allocation problem operates at the *model selection* level, not the feature selection level. The individual models (Lasso, Elastic Net, Random Forest) each solve the feature selection problem competently within their own framework. What they cannot do individually is determine which *type* of model—linear sparse, linear grouped, or nonlinear—is most appropriate for the current narrative environment. The NAS combination addresses this higher-order allocation problem, consistent with the rational inattention literature's emphasis on attention allocation across information *channels* (Maćkowiak and Wiederholt, 2009) rather than within them.

This separation of concerns—features within models, models within the combination—is a theoretically clean division of labor. It also explains why the NAS combination works: it solves a problem that the individual models cannot solve for themselves.

B.6 Robustness of the Null

To verify that the $\gamma^* = 0$ result is not an artifact of the grid resolution, we also estimated the NAS-Estimation with a finer grid ($\gamma \in \{0, 0.001, 0.002, \ldots, 0.100\}$) and with alternative penalty modulation functions (linear: $\lambda_t^{\mathrm{NAS}} = \lambda_t^{\mathrm{CV}} \cdot (1 - \gamma \cdot \mathrm{NAI}_t)^+$; power: $\lambda_t^{\mathrm{NAS}} = \lambda_t^{\mathrm{CV}} \cdot (1 + \mathrm{NAI}_t)^{-\gamma}$). In all cases, cross-validation selects the standard Lasso ($\gamma^* = 0$), confirming that the null result is robust to the functional form of the attention modulation.

C Robustness Checks Results

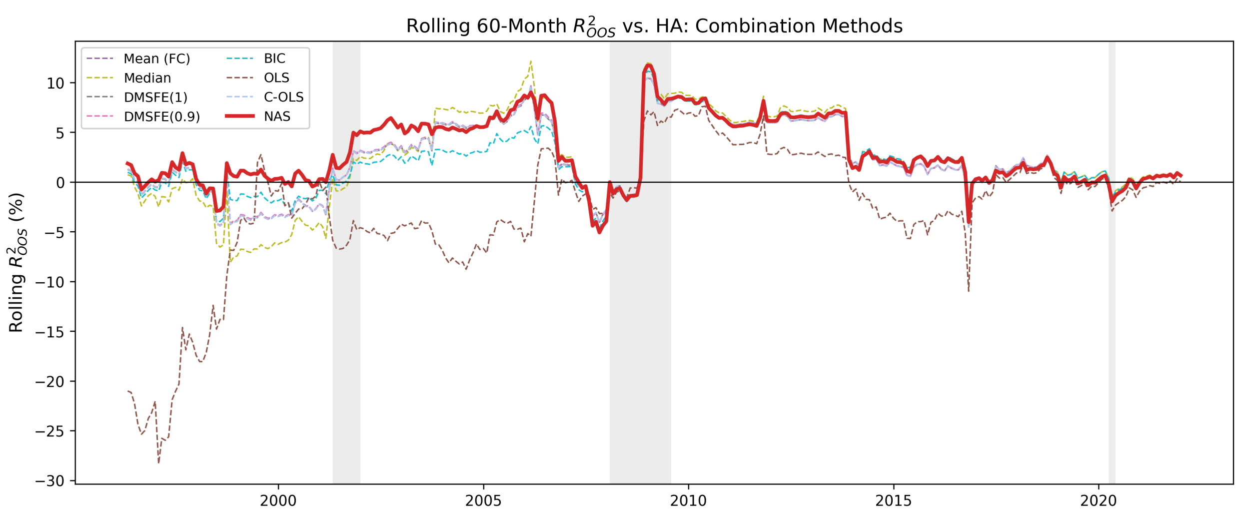


Figure 10: Rolling R^2_{OOS}: Combination Methods vs. Historical Average. 60-month rolling window. Shaded areas denote NBER recession months.

Table 11: Post-Validation Out-of-Sample Performance (2006:04–2021:12)

Model	$R^2_{OOS,HA}$ (%)	$R^2_{OOS,AR}$ (%)
Individual Models		
Lasso	3.51	3.73
Elastic Net	3.73	3.94
Random Forest	2.04	2.25
Combination Methods		
Mean (FC)	3.24	3.45
Median	3.54	3.75
DMSFE(1)	3.24	3.45
DMSFE(0.9)	3.24	3.45
BIC	3.35	3.57
C-OLS	3.24	3.45
NAS	**3.31**	**3.52**
AR	-0.22	—

Notes: Out-of-sample performance over the post-validation period 2006:04–2021:12 ($T = 189$). NAS hyperparameters are fixed at values selected over the validation window 1996:04–2006:04. OLS combination is omitted due to negative R^2_{OOS} in the full sample.

D Feature Dictionary

Table 12 provides the complete mapping of the 124 text-based features used in this paper. Features are organized into seven categories following the classification of Aruoba and Drechsel (ming). For each feature, we report the index used in the estimation code, the variable name (corresponding to the economic concept searched in the newspaper corpus), and the category.

The economic concepts are searched as exact phrases in the preprocessed newspaper text. Multi-word concepts (e.g., "stock market," "corporate debt," "dow jones industrial average") are matched as complete phrases; single-word concepts (e.g., "inflation," "bank") are matched as individual tokens. The sentiment index for each concept is constructed using the ±10-word contextual window method described in Section 2.3.2.

Table 12: Complete Feature Dictionary

Index	Feature Name	Display Label	Category
Macroeconomic Conditions (20 features)			
0	`gdp`	GDP	Macro
1	`gdp_growth`	GDP Growth	Macro
2	`economic_growth`	Economic Growth	Macro
3	`economic_activity`	Economic Activity	Macro
4	`recession`	Recession	Macro
5	`economic_expansion`	Economic Expansion	Macro
6	`business_cycle`	Business Cycle	Macro
7	`unemployment`	Unemployment	Macro
8	`unemployment_rate`	Unemployment Rate	Macro
9	`job_growth`	Job Growth	Macro
10	`employment_growth`	Employment Growth	Macro
11	`labor_market`	Labor Market	Macro
12	`payroll`	Payroll	Macro
13	`jobless_claims`	Jobless Claims	Macro
14	`industrial_production`	Industrial Production	Macro
15	`manufacturing`	Manufacturing	Macro
16	`capacity_utilization`	Capacity Utilization	Macro
17	`consumer_spending`	Consumer Spending	Macro
18	`retail_sales`	Retail Sales	Macro

Continued on next page

Table 12 continued

Index	Feature Name	Display Label	Category
19	personal_consumption	Personal Consumption	Macro
Inflation & Monetary Policy (17 features)			
20	inflation	Inflation	Infl./Mon. Pol.
21	inflation_rate	Inflation Rate	Infl./Mon. Pol.
22	price_inflation	Price Inflation	Infl./Mon. Pol.
23	consumer_prices	Consumer Prices	Infl./Mon. Pol.
24	cpi	CPI	Infl./Mon. Pol.
25	core_inflation	Core Inflation	Infl./Mon. Pol.
26	deflation	Deflation	Infl./Mon. Pol.
27	disinflation	Disinflation	Infl./Mon. Pol.
28	federal_reserve	Federal Reserve	Infl./Mon. Pol.
29	fed	Fed	Infl./Mon. Pol.
30	monetary_policy	Monetary Policy	Infl./Mon. Pol.
31	central_bank	Central Bank	Infl./Mon. Pol.
32	interest_rate	Interest Rate	Infl./Mon. Pol.
33	interest_rates	Interest Rates	Infl./Mon. Pol.
34	federal_funds_rate	Federal Funds Rate	Infl./Mon. Pol.
35	fed_funds	Fed Funds	Infl./Mon. Pol.
36	inflation_expectations	Inflation Expectations	Infl./Mon. Pol.
Corporate Fundamentals (25 features)			
37	earnings	Earnings	Corp. Fund.
38	corporate_earnings	Corporate Earnings	Corp. Fund.
39	earnings_growth	Earnings Growth	Corp. Fund.
40	earnings_report	Earnings Report	Corp. Fund.
41	profit	Profit	Corp. Fund.
42	profits	Profits	Corp. Fund.
43	profitability	Profitability	Corp. Fund.
44	profit_margin	Profit Margin	Corp. Fund.
45	profit_margins	Profit Margins	Corp. Fund.
46	operating_profit	Operating Profit	Corp. Fund.
47	net_income	Net Income	Corp. Fund.
48	revenue	Revenue	Corp. Fund.
49	revenues	Revenues	Corp. Fund.
50	sales	Sales	Corp. Fund.

Continued on next page

Table 12 continued

Index	Feature Name	Display Label	Category
51	sales_growth	Sales Growth	Corp. Fund.
52	top_line	Top Line	Corp. Fund.
53	labor_cost	Labor Cost	Corp. Fund.
54	input_cost	Input Cost	Corp. Fund.
55	operating_cost	Operating Cost	Corp. Fund.
56	cost_pressure	Cost Pressure	Corp. Fund.
57	capital_spending	Capital Spending	Corp. Fund.
58	capital_expenditure	Capital Expenditure	Corp. Fund.
59	capex	CAPEX	Corp. Fund.
60	investment	Investment	Corp. Fund.
61	business_investment	Business Investment	Corp. Fund.
Market & Valuation (11 features)			
62	stock_market	Stock Market	Market
63	stock_prices	Stock Prices	Market
64	equity_prices	Equity Prices	Market
65	market_value	Market Value	Market
66	share_price	Share Price	Market
67	share_prices	Share Prices	Market
68	valuation	Valuation	Market
69	price_earnings	Price-Earnings	Market
70	earnings_yield	Earnings Yield	Market
71	dividend_yield	Dividend Yield	Market
72	market_valuation	Market Valuation	Market
Credit & Financial Conditions (22 features)			
73	credit	Credit	Credit
74	credit_market	Credit Market	Credit
75	credit_spread	Credit Spread	Credit
76	credit_spreads	Credit Spreads	Credit
77	corporate_debt	Corporate Debt	Credit
78	corporate_bond	Corporate Bond	Credit
79	corporate_bonds	Corporate Bonds	Credit
80	debt_market	Debt Market	Credit
81	bond_market	Bond Market	Credit
82	junk_bond	Junk Bond	Credit

Continued on next page

Table 12 continued

Index	Feature Name	Display Label	Category
83	high_yield	High Yield	Credit
84	default	Default	Credit
85	default_risk	Default Risk	Credit
86	bankruptcy	Bankruptcy	Credit
87	credit_quality	Credit Quality	Credit
88	bank	Bank	Credit
89	banks	Banks	Credit
90	banking	Banking	Credit
91	banking_sector	Banking Sector	Credit
92	financial_sector	Financial Sector	Credit
93	financial_institution	Financial Institution	Credit
94	lending	Lending	Credit
Sentiment & Risk (19 features)			
95	investor_confidence	Investor Confidence	Sentiment
96	consumer_confidence	Consumer Confidence	Sentiment
97	business_confidence	Business Confidence	Sentiment
98	market	Market (mentions)	Sentiment
99	market_mentions	Market Mentions	Sentiment
100	investor	Investor (mentions)	Sentiment
101	investor_mentions	Investor Mentions	Sentiment
102	optimism	Optimism	Sentiment
103	pessimism	Pessimism	Sentiment
104	risk	Risk	Sentiment
105	risk_appetite	Risk Appetite	Sentiment
106	risk_aversion	Risk Aversion	Sentiment
107	uncertainty	Uncertainty	Sentiment
108	volatility	Volatility	Sentiment
109	market_volatility	Market Volatility	Sentiment
110	expectations	Expectations	Sentiment
111	outlook	Outlook	Sentiment
112	forecast	Forecast	Sentiment
113	projection	Projection	Sentiment
Sectors (10 features)			
114	technology	Technology	Sectors

Continued on next page

Table 12 continued

Index	Feature Name	Display Label	Category
115	technology_sector	Technology Sector	Sectors
116	energy	Energy	Sectors
117	energy_sector	Energy Sector	Sectors
118	oil_price	Oil Price	Sectors
119	oil_prices	Oil Prices	Sectors
120	housing	Housing	Sectors
121	real_estate	Real Estate	Sectors
122	housing_market	Housing Market	Sectors
123	dow_jones_industrial_average	Dow Jones Ind. Avg.	Sectors