News Narratives and Omitted Credit-Cycle States in Corporate Bond Markets*

Yiyuan Wang

August 2026

Abstract

Standard credit-risk variables leave a broad common residual component in corporate bond markets. I ask whether topic-level narratives from more than four decades of *Wall Street Journal* business news help measure these omitted credit-cycle states. Using latent Dirichlet allocation (LDA) for historical diagnostics and a strict train-only online LDA (oLDA) design for real-time tests, I construct monthly narrative-attention measures and relate them to credit spreads and returns across rating segments. After removing the component of monthly spread changes explained by standard rates, slope, equity-market, and option-implied controls, the first principal component of the standardized A, BBB, and high-yield (HY) residuals explains about 76 percent of their variation and assigns similar weights across ratings. On the 408-month text-residual sample, full-sample text mappings are strongest in risky credit. A sparse diagnostic yields a nonempty active set only for HY, while a complementary full-topic Ridge mapping produces a monotonic diffuse-fit gradient from A to BBB to HY. Real-time forecasting evidence is smaller: a spread-based risk filter accounts for most of the HY timing improvement, while the narrative overlay adds a modest increment. The evidence therefore separates a broad omitted credit state from a rating-specific narrative mapping that is strongest in risky credit.

Keywords: textual analysis, corporate bonds, credit cycle, Wall Street Journal, topic models, out-of-sample forecasting, credit spreads

*I am grateful to Prof. Chunchi Wu for helpful comments.

1 Introduction

Corporate bond markets exhibit strong time variation in returns and credit spreads, especially in risky debt. Expected bond returns are countercyclical and concentrated in lower-grade bonds (Fama and French, 1989; Lin et al., 2018), and high-yield spreads contain information about future real activity (Gertler and Lown, 1999).[1] A central challenge is that standard observables do not fully capture the states that matter most for risky credit. Collin-Dufresne et al. (2001), henceforth CDGM, show that structural credit-risk variables explain only a limited share of monthly credit-spread changes and leave a large common residual factor. Friewald and Nagler (2019) show that over-the-counter (OTC) market frictions account for part of that common component, but their evidence also reinforces the broader point: important credit-market states are difficult to measure with conventional proxies alone.

This paper asks whether public news narratives help measure those omitted credit-cycle states. The economic premise is simple. If the missing state reflects funding conditions, intermediary balance-sheet stress, rollover risk, creditor bargaining, or shifting downside-risk perceptions, then business news should contain information about it. The prediction is also sharp: narrative information should be most useful where credit risk is most state-dependent, namely in BBB and especially high-yield (HY) bonds.

I test this idea using more than four decades of *Wall Street Journal* (WSJ) business news. I estimate topic models on the WSJ archive and aggregate document-level topic mixtures into monthly narrative-attention measures. Full-sample latent Dirichlet allocation (LDA) topics are used for construct validation and interpretation. For real-time tests, I use a strict train-only online LDA (oLDA) design that refits the topic model using only past text and aligns local topics to persistent topic IDs across refits. The object of interest is not the topic model itself, but the credit-cycle state that the topic measures help reveal.

The main evidence is residual-side. After removing the component of monthly spread changes explained by standard rates, slope, equity-market, and option-implied controls, the first principal component (PC1) of the standardized A, BBB, and HY residuals explains about 76 percent of their variation and assigns similar weights across ratings. Standard observables therefore leave a broad common residual credit state. The textual mapping is more structured. On the common 408-month text-residual sample, a sparse least absolute shrinkage and selection operator (LASSO) specification selects no topics for PC1, A, or BBB but selects twelve for HY, where adjusted R^2 is 9.55 percent

[1] A related macro-finance literature studies credit-market tightness through spreads, excess bond premia, issuance quality, and broader financial conditions. See, among others, Gilchrist and Zakrajšek (2012), Greenwood and Hanson (2013), and International Monetary Fund (2017). These objects are closely related but not identical: some measure the price of bearing credit risk, others the quantity or quality of credit supply, and still others broader downside risks to growth.

for the Main residual and 9.57 percent for the leverage-augmented residual when the same topic set is held fixed. Under this full-sample sparse diagnostic, a nonempty active set appears only for HY; the result therefore points to a more parsimonious HY representation rather than a monotonic sparse-selection gradient across ratings. Applying the same HY-selected topics without re-selection produces smaller positive fit for PC1, A, and BBB. A full-$K = 180$ Ridge mapping reveals a complementary pattern: among the rating-specific residuals, in-sample R^2 rises monotonically from A to BBB to HY. Thus the omitted residual state is broad across ratings, while the strength of its diffuse narrative mapping increases toward risky credit; under the baseline full-sample sparse diagnostic, a nonempty low-dimensional representation appears only in HY.

The out-of-sample tests serve as validation rather than as a standalone trading claim. HY return forecasts have positive out-of-sample R^2 across rolling windows, whereas aggregate bond forecasts are weak. In timing tests, a spread-based risk filter accounts for most of the Sharpe improvement, and the text overlay adds a smaller increment by refining exposure within already risk-managed states. Horse races against the FRED monthly database for macroeconomic research (FRED-MD), Goyal–Welch (GW) predictors, and Economic Policy Uncertainty (EPU) predictors show that the text signal is not dominated by standard macro information sets, although spreads remain the first-order state variable in the timing exercise.

The paper contributes to three literatures. First, it adds to corporate bond asset pricing by providing text-based evidence on the omitted-state problem highlighted by Collin-Dufresne et al. (2001) and Friewald and Nagler (2019). Second, it connects bond return predictability to public information. Relative to Lin et al. (2018), the contribution is informational rather than algorithmic: narratives help locate and interpret the low-grade credit states in which predictability concentrates. Third, it contributes to the finance text literature by shifting the corporate-bond application from tone, readability, and disclosure language toward topic-level public narratives as aggregate state variables (Tetlock, 2007; Bybee et al., 2024; Fang-Klingler, 2019; Isakin and Pu, 2023; Liu et al., 2024; Khoja, 2024).

Sections 2 and 3 describe the data and narrative measures. Section 4 develops the in-sample diagnostics and the main CDGM-style residual evidence. Sections 5 and 6 report out-of-sample validation and macro horse races. Section 7 interprets the oLDA signal and documents topic-alignment stability. Section 8 concludes.

2 Data Description

2.1 Bond Data

Bond returns and credit spreads are constructed from widely used U.S. corporate bond indices. The primary return series consist of monthly total returns from the Salomon Brothers Corporate (SBC) bond indices, including an aggregate series that we denote "ALL," together with a separate high-yield series denoted "HY." The SBC indices for ALL and the investment-grade rating buckets (AAA/AA, A, and BBB; equal-weighted) run from January 1980 through August 2025. The HY total return series is spliced across two providers: the Bloomberg U.S. Corporate High Yield Index from July 1983 to January 1989 and the FTSE Russell U.S. High-Yield Market Index from January 1989 to September 2025.

For yields, I assemble a single monthly HY yield-to-maturity series by splicing three components. From 2001 through 2021, I use the Bloomberg U.S. Corporate High Yield Index; from 2022 onward, I use the FTSE Russell U.S. High-Yield Market Index; and for 1996 through 2000, I use the ICE BofA U.S. High Yield Index yield series from Federal Reserve Economic Data (FRED). For 1986 through 1995, HY yields are backfilled using the post-2000 overlap: I estimate a linear projection of the HY yield on the contemporaneous FRED AAA and BAA corporate-bond yield series (series DAAA and DBAA) and then apply the fitted coefficients to earlier months.

To keep the out-of-sample tests strictly free of look-ahead, I do not rely on the backfilled segment when forming the risk filter. Accordingly, all timing strategies that condition on yields and spreads begin in 1996, when the HY yield series is observed rather than imputed.

2.2 The Wall Street Journal Data Set

2.2.1 Data sources, coverage, and alignment to financial series

The text input is a monthly Dow Jones-Wall Street Journal news archive. The raw archive spans June 1979 through January 2024, and the fully rebuilt and audited sample used for topic modeling also runs through January 2024, which is the last month with complete outputs.

All empirical analysis is conducted at the monthly frequency. Narrative attention is constructed as an as-of-month-t measure from articles available through month t. In the out-of-sample exercises, the return-forecast table uses a one-month lagged implementation of this train-only signal, whereas the timing strategies use the as-of-t signal directly.

2.2.2 Deduplication and article-level filtering

The WSJ rebuild targets two goals: long-sample comparability and an auditable path from raw text to the document-term matrix. The Rebuilder consolidates duplicate or split records sharing an accession or story identifier, merging headlines and bodies under a single timestamp, and applies structural filters to retain business-relevant coverage: weekday publication, front-page business sections (A/B/C), exclusion of clearly non-economic categories such as sports, and a minimum article-length requirement.

A separate challenge in long-run news archives is the prevalence of template-heavy and public-relations (PR)-like content. I therefore strip recurring boilerplate at the paragraph, line, and sentence levels, targeting items such as contact blocks, webcast notices, copyright language, and forward-looking disclaimers. Rather than excluding all such material mechanically, I assign a document weight ω_d that downweights articles identified as PR-like, template-heavy, or regulatory in style, while leaving standard WSJ reporting at unit weight. These weights enter only the aggregation of monthly topic attention and do not alter the underlying financial series.

2.2.3 Tokenization and canonicalization

Tokenization follows a Kelly-aligned stopword and regex-stopword design, supplemented with corpus-specific exclusions. I lowercase text, remove punctuation and non-letter characters, enforce a minimum token length, and discard residual contraction stumps that otherwise survive standard tokenization. I then apply a conservative canonicalization step that merges obvious morphological variants, such as plurals and common suffix forms, while preserving a surface-form map for readable display.

The goal is not aggressive stemming, but stable normalization. This improves dictionary uniformity across decades without collapsing economically distinct terms. The pipeline therefore remains auditable: each canonical token can be traced back to a representative surface form used later in topic labels and interpretability checks.

2.2.4 Dictionary construction and document-term matrix outputs

From the rebuilt token stream, I construct a fixed vocabulary using document-frequency thresholds that exclude terms that are either too rare to be informative or so common as to be nearly ubiquitous. This produces a unigram dictionary $\{1, \ldots, V\}$, and I optionally retain a bigram dictionary for collocation diagnostics and labeling. Boilerplate bigrams, such as "additional information" or "please contact," are explicitly excluded from interpretability outputs.

The final document-term representation is stored as sparse monthly count files for headline and body text, together with monthly document metadata, including document weights and audit flags.

This storage design separates raw-text processing, dictionary construction, and count generation, so later topic estimation and robustness checks can be run without re-parsing the original archive.

3 Topic Model and Estimation

This section defines the text-based state variables used in the empirical analysis. I first describe the LDA representation of the rebuilt *Wall Street Journal* corpus, then show how document-level topic mixtures are aggregated into monthly attention shares, and finally explain the strict no-look-ahead oLDA procedure used in the out-of-sample exercises.

3.1 Latent Dirichlet Allocation (LDA)

3.1.1 Bag-of-words representation

Let d index WSJ articles and $v \in \{1, \ldots, V\}$ index the fixed vocabulary. After rebuilding and tokenizing each article, I represent document d by a term-count vector $w_d \in \mathbb{Z}_{\geq 0}^{V}$, where $w_{d,v}$ is the count of term v in article d and $N_d = \sum_{v=1}^{V} w_{d,v}$ is document length. Stacking across documents yields a sparse document-term matrix $W \in \mathbb{Z}_{\geq 0}^{D \times V}$. The baseline model uses unigrams in the likelihood. Bigrams are retained only for collocation diagnostics and display-side labeling, not as primitives of estimation.[2]

3.1.2 Statistical model: topics and document-specific mixtures

LDA treats each document as a mixture of K latent topics, where each topic is a distribution over terms. Let $\Phi = [\phi_1, \ldots, \phi_K]'$ denote the topic-term distributions and let $\theta_d = (\theta_{d,1}, \ldots, \theta_{d,K})'$ denote the topic mixture for document d, with $\sum_v \phi_{k,v} = 1$. Conditional on (Φ, θ_d), term counts follow:

$$w_d \mid \theta_d, \Phi \sim \text{Multinomial}\left(N_d, \Phi' \theta_d\right), \tag{1}$$

with Dirichlet priors:

$$\theta_d \sim \text{Dirichlet}(\alpha), \qquad \phi_k \sim \text{Dirichlet}(\eta). \tag{2}$$

Here Φ captures semantic content, while θ_d captures how much attention document d allocates across narratives.

[2] When bigrams are used for labeling or coherence diagnostics, I measure collocation strength from rebuilt counts using pointwise mutual information (PMI) and normalized pointwise mutual information (NPMI): $\text{PMI}(a, b) = \log\left(\frac{p(a,b)}{p(a)p(b)}\right)$ and its normalized variant $\text{NPMI}(a, b) = \text{PMI}(a, b)/(-\log p(a, b))$.

3.1.3 Estimation and outputs

I estimate the model using collapsed Gibbs sampling (tomotopy), with K topics, Dirichlet priors α and η, and 1,200 baseline iterations. Estimation yields topic-term probabilities $\widehat{\Phi}$ and document-level topic mixtures $\widehat{\theta}_d$. I use $\widehat{\Phi}$ to construct display-side top terms and labels, but these display adjustments do not affect the estimated $\widehat{\theta}_d$ used in the empirical analysis. Following Bybee et al. (2024), I also report a rescaled topic-term matrix $\widetilde{\Phi}$ defined by:

$$\widetilde{\phi}_{k,v} \propto \frac{\phi_{k,v}}{f_v}, \qquad \sum_v \widetilde{\phi}_{k,v} = 1, \tag{3}$$

where f_v is a global baseline frequency. This rescaling improves interpretability of top-term lists without changing the topic mixtures used for forecasting.[3]

3.1.4 Monthly topic attention: from documents to a time series

The predictors entering the regressions are monthly topic-attention shares. Let m index calendar months and let $\mathcal{T}(m)$ denote the set of documents published in month m. Using estimated document-level topic shares $\widehat{\theta}_{d,k}$, I define topic k's month-m attention share as:

$$\tau_{k,m} = \frac{\sum_{d\in\mathcal{T}(m)} \omega_d N_d \widehat{\theta}_{d,k}}{\sum_{d\in\mathcal{T}(m)} \omega_d N_d}, \qquad \text{so that} \quad \sum_{k=1}^{K} \tau_{k,m} = 1. \tag{4}$$

Here ω_d is the Rebuilder's document-level quality weight and N_d is document length, so longer and less boilerplate-heavy articles receive more influence in monthly attention. The resulting full-sample series is useful for describing narrative content and for constructing stable topic labels.[4]

3.2 Online LDA (oLDA) and strict no-look-ahead construction of topic attention

A full-sample LDA is useful for interpretation, but it is not an implementable forecasting signal because the estimated topics and document mixtures implicitly use future text. For the out-of-sample exercises, I therefore construct topic attention in a strict no-look-ahead manner. Let t index calendar months and let $\mathcal{D}_t$ denote the documents published in month t. At each t, I estimate the topic model using only text dated through $t-1$, obtaining topic-term parameters Φ_{t-1}. Holding Φ_{t-1} fixed, I

[3] This transformation affects only the display layer: it suppresses high-frequency generic terms in top-term lists and labels, but leaves document-level topic mixtures unchanged.

[4] ω_d is a document-level quality weight from the Rebuilder. Clearly noninformational template items receive $\omega_d = 0$, and PR-like items receive $\omega_d < 1$. The weight enters attention aggregation only, not the LDA likelihood.

infer document-level topic mixtures for month-t articles, $\widehat{\theta}_{d,t} = \mathbb{E}[\theta \mid w_{d,t}, \Phi_{t-1}]$, and aggregate them into month-t topic attention:

$$nd_{k,t} = \frac{\sum_{d \in \mathcal{D}_t} \omega_d L_d \widehat{\theta}_{d,t,k}}{\sum_{d \in \mathcal{D}_t} \omega_d L_d}. \tag{5}$$

Here L_d is document length and ω_d is the Rebuilder weight. By construction, $nd_{k,t}$ depends only on information available at time t, so it is a true as-of-t predictor.

This implementation preserves the timing discipline needed for forecasting because all transformations and model estimates are formed using past text only. Relative to the sequential Hoffman-style oLDA used in Bybee et al. (2024), my procedure periodically refits the model on a rolling training window and aligns local topics to persistent global topic IDs using explicit cross-refit similarity matching. The goal is to balance adaptation and continuity so that out-of-sample narrative channels remain auditable when topic content reorganizes across news regimes.[5]

4 Narrative Diagnostics and Residual-Side Evidence

This section develops the main empirical evidence. I first use full-sample LDA topics as construct-validation diagnostics, asking whether the selected narratives vary systematically across rating segments. I then turn to the residual-side test. After standard CDGM-style controls remove observable credit-risk components from spread changes, I ask whether the remaining state is common across ratings and whether its relation to news is captured by a sparse HY-specific representation or is distributed more diffusely across the broader topic space.

4.1 Rating gradient in narrative diagnostics

Table 1 reports an in-sample return diagnostic. Month-t topic attention is used to explain month-$t+1$ excess returns under an exact-k LASSO rule ($k = 5$) with a post-selection ordinary least squares (OLS) refit. Standard errors are heteroskedasticity- and autocorrelation-consistent (HAC), following Newey and West (1987). Because variable selection and coefficient estimation use the same full sample, the reported post-selection p-values are descriptive rather than selection-adjusted inference. This exercise checks whether the text measures select economically plausible narratives across rating buckets.

The selected topics display a rating gradient. In safer segments, the active set emphasizes

[5] Bybee et al. (2024) use a sequential online topic estimator in the spirit of Hoffman et al. (2010). My approach instead periodically refits the model and aligns local topics to persistent global topic IDs so that out-of-sample narrative channels remain easier to track over time.

broad macro-financing narratives such as sovereign stress, industrial supply chains, and liability management. Moving down the rating spectrum, the selected narratives tilt toward more state-dependent downside risks. BBB adds downturn and geopolitical narratives, while HY selects downturn, natural-disaster, and policy topics. The return R^2 does not rise monotonically because HY returns are noisier, but the composition of selected narratives shifts toward the states in which risky credit should be most exposed. The ALL panel is reported in Appendix Table A1.

Table 2 repeats the exercise for yield spreads. Fit is higher for spreads than for returns, with R^2 ranging from 0.56 for HY to 0.70 for ALL. Downturn, yield-curve, and balance-sheet topics emerge as spread-state variables, again with stronger economic relevance in lower-grade credit. Supporting same-month counterparts are reported in Appendix Table A2 (Panels A and B).

4.2 Visualizing selected narratives and their semantic content

Figure 1 complements Tables 1 and 2 by plotting monthly topic shares alongside each narrative's key terms, showing both when and what each narrative represents.

The displayed topics have economically distinct content. Economic Downturn and Yield Curve track recession and discount-rate conditions; European Fiscal & Political Crisis / Sovereign Debt captures sovereign stress and policy uncertainty; Liability Management: Tender & Exchange Offers identifies refinancing and creditor-action episodes; and Natural Disasters & Severe Incidents captures episodic downside shocks. Figure 1 therefore supports interpreting the topic shares as narrative state variables rather than as undifferentiated news volume.

4.3 Supporting benchmark diagnostic: topics versus FRED-MD

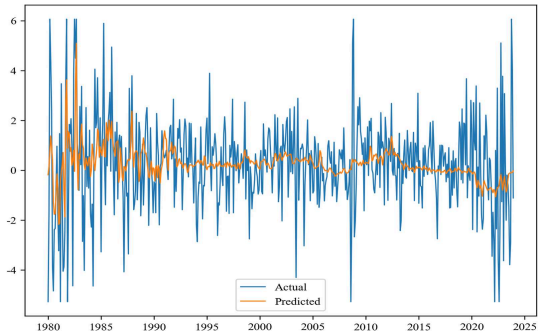
Before turning to the residual evidence, I check whether the narrative measures contain information not spanned by a broad macro predictor pool. Following Bybee et al. (2024), Table 3 compares a FRED-MD benchmark[6] with a model that adds orthogonalized topic measures under the same five-regressor sparsity cap.

[6]The FRED-MD database (McCracken and Ng, 2016) includes a broad set of macro series, some of which are asset prices. Following the concern that such variables can be mechanically linked to asset returns, I exclude a small subset of credit-yield–type series from the return benchmark pool. This exclusion reduces the risk that "incremental" fit is driven by variables that are close to the outcome by construction rather than by macroeconomic state information.

Table 1. Five-Topic In-Sample Predictive Diagnostics — Returns

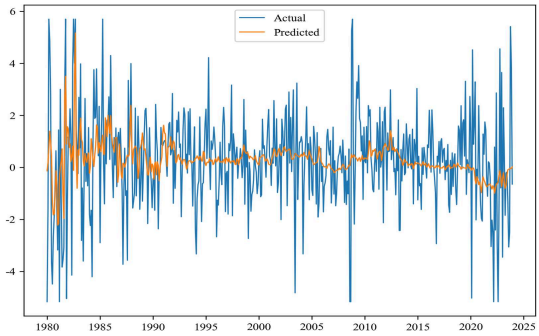
AAA & AA — Return

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.43	0.00
Industrial Machinery & EV Supply Chain	0.34	0.00
Liability Management: Tender & Exchange Offers	0.12	0.00
Consumer Staples	-0.41	0.00
Prospectus & New Issuance	-0.13	0.04
R^2	0.11	



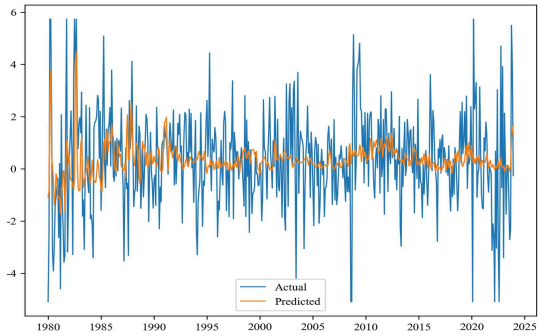
A — Return

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.43	0.00
Industrial Machinery & EV Supply Chain	0.34	0.00
Consumer Staples	-0.44	0.00
Liability Management: Tender & Exchange Offers	0.12	0.00
Prospectus & New Issuance	-0.13	0.04
R^2	0.11	



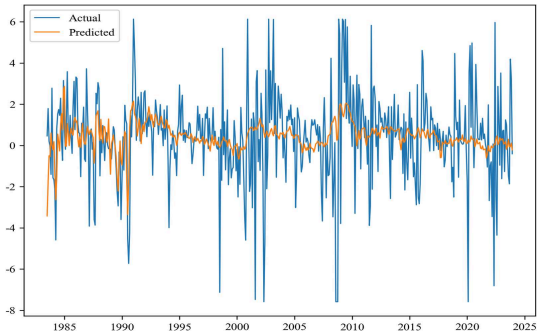
BBB — Return

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.43	0.00
Industrial Machinery & EV Supply Chain	0.36	0.00
Economic Downturn	-0.24	0.00
Appropriation	0.62	0.00
Middle East	0.65	0.02
R^2	0.10	



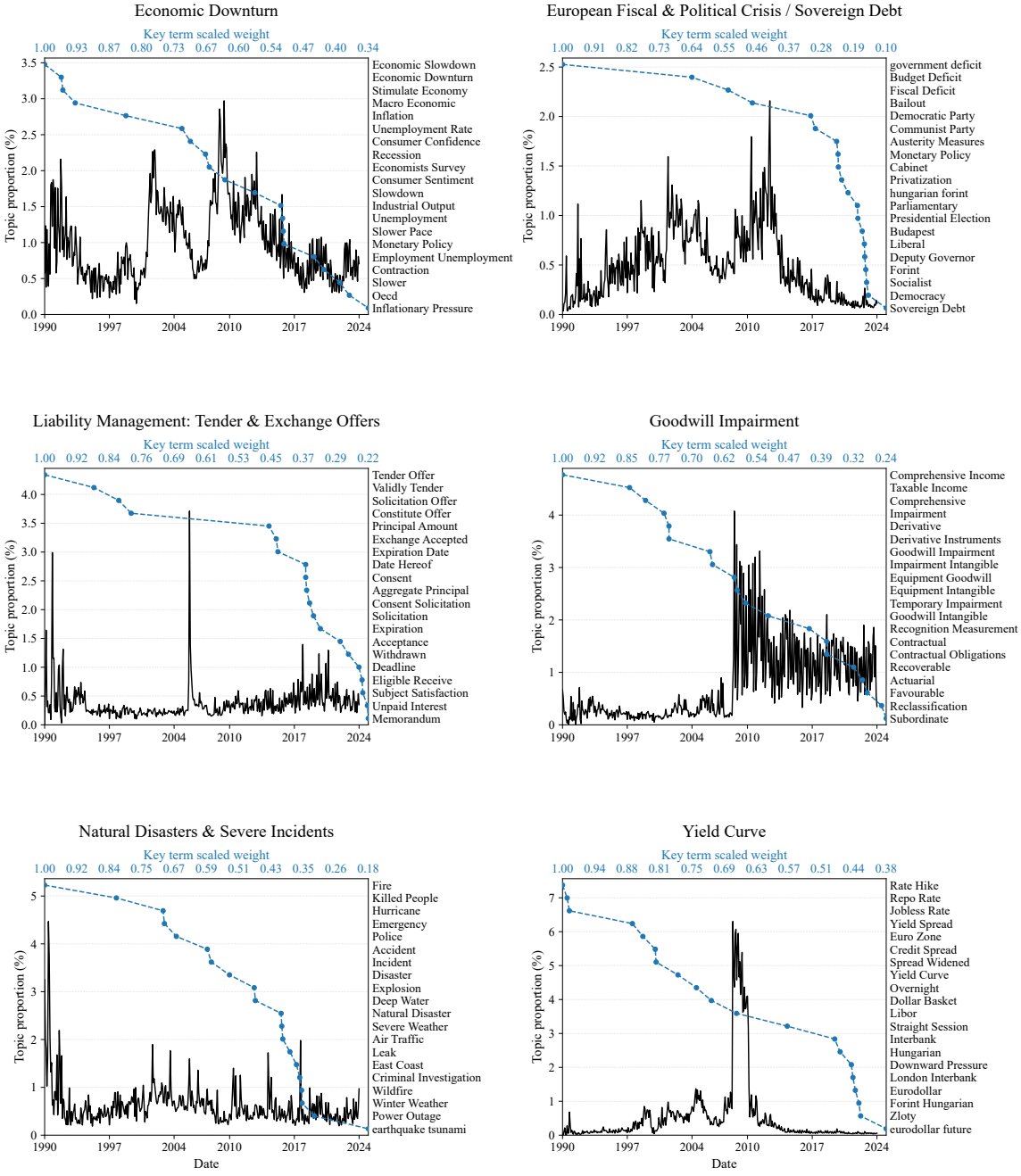
High Yield — Return

Topic	Coeff.	p-val.
Economic Downturn	-0.54	0.00
Natural Disasters & Severe Incidents	-0.39	0.00
Strategic Initiatives & Operational Efficiency	-0.20	0.02
Government Policy & Regulation	-1.17	0.12
European Fiscal & Political Crisis / Sovereign Debt	-0.27	0.30
R^2	0.09	



Notes: For each rating bucket, month-t topic attention shares are used to explain month-t + 1 excess returns in a full-sample predictive validation exercise. For each rating bucket, I select exactly five topics and refit the selected set by ordinary least squares. Reported p-values are conventional post-selection values and are provided as descriptive diagnostics. EV denotes electric vehicle.

Figure 1. Topic Attention for Selected WSJ Topics

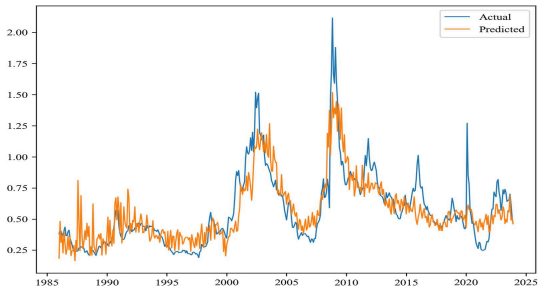


Notes: Each panel plots the monthly share of a selected LDA topic in the *Wall Street Journal* corpus (black line). The terms shown on the right are the topic's displayed key terms, ordered by within-topic weight, and the dashed blue curve reports their relative weights scaled by the largest displayed term weight within that topic, so the horizontal scale runs from 0 to 1. The topics shown are representative topics selected in Tables 1 and 2 and are used to visualize both the time variation and semantic content of the static full-sample LDA narratives. The figure is descriptive and is shown for interpretation rather than as the train-only out-of-sample forecasting signal used later in the paper.

Table 2. Five-Topic In-Sample Predictive Diagnostics — Yield Spreads

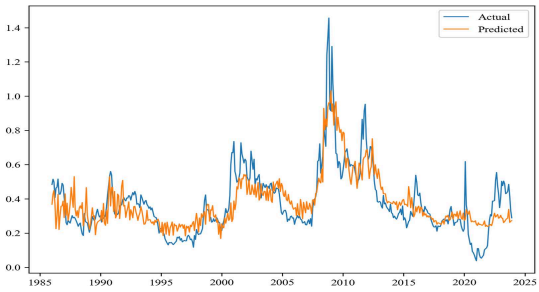
All Bonds — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	0.24	0.00
Goodwill Impairment	0.12	0.00
Severe Weather & Regional Macro	0.12	0.00
EBIT & Goodwill	0.06	0.01
Yield Curve	0.07	0.01
R^2	0.70	



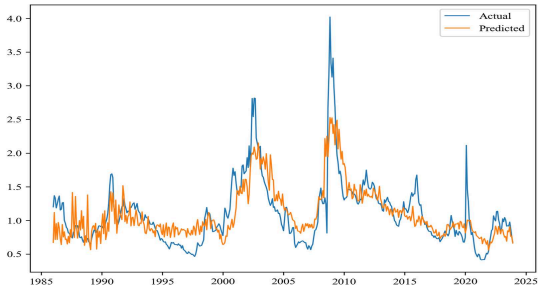
A — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	0.12	0.00
Revenue — Macro	0.04	0.00
Yield Curve	0.08	0.00
Goodwill Impairment	0.04	0.00
Eurozone Policy Stress	0.04	0.01
R^2	0.63	



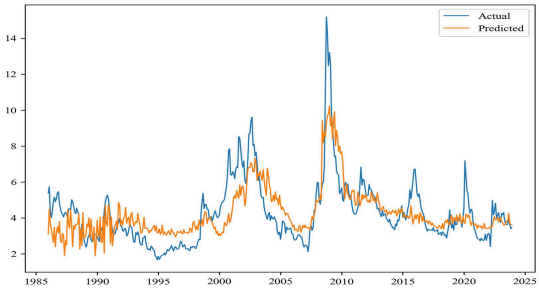
BBB — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	0.39	0.00
Goodwill Impairment	0.13	0.00
Severe Weather & Regional Macro	0.19	0.00
Strategic Initiatives & Operational Efficiency	-0.11	0.00
Yield Curve	0.12	0.04
R^2	0.62	



High Yield — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	1.13	0.00
Goodwill Impairment	0.39	0.00
Yield Curve	0.68	0.01
Severe Weather & Regional Macro	0.54	0.01
Environmental Protection	-0.77	0.01
R^2	0.56	



Notes: See Table 1. The specification here uses yield spreads in a full-sample predictive validation exercise. EBIT denotes earnings before interest and taxes.

Table 3. Supporting Return Diagnostic: LDA Topics versus FRED-MD

ALL Return				HY Return			
Topic	Coeff.	t	p-val.	Topic	Coeff.	t	p-val.
Panel A: Full (LDA Topics + FRED-MD)							
S&P 500 (FRED-MD)	-0.102	-4.83	1.4e-06	Economic Downturn	-0.307	-3.01	0.003
Liability Management: Tender & Exchange Offers	0.126	4.57	4.9e-06	Government Policy & Regulation	-1.64	-2.98	0.003
CES1021000001 (FRED-MD)	-0.244	-3.80	1.5e-04	Natural Disasters & Severe Incidents	-0.255	-2.91	0.004
Industrial Machinery & EV Supply Chain	0.245	3.20	0.001	BUSINVx (FRED-MD)	-0.287	-1.57	0.117
European Fiscal & Political Crisis / Sovereign Debt	0.285	2.61	0.009	CP3Mx (FRED-MD)	-0.00813	-1.29	0.196
Panel B: Benchmark (FRED-MD only)							
CES1021000001 (FRED-MD)	-0.198	-3.57	3.6e-04	T10YFFM (FRED-MD)	0.00208	2.86	0.004
S&P 500 (FRED-MD)	-0.0781	-3.47	5.1e-04	VIXCLSx (FRED-MD)	0.000330	2.32	0.020
AMDMUOx (FRED-MD)	-0.249	-3.23	0.001	BUSINVx (FRED-MD)	-0.372	-1.80	0.072
PCEPI (FRED-MD)	-0.943	-2.68	0.007	TB3MS (FRED-MD)	-0.00537	-0.45	0.653
COMPAPFFx (FRED-MD)	0.00447	1.51	0.131	CP3Mx (FRED-MD)	-0.00529	-0.41	0.682
Full R^2	0.13			Full R^2	0.10		
Benchmark R^2	0.11			Benchmark R^2	0.07		

Notes: Five-regressor active-set regressions after LASSO selection. Predictors at month t forecast next-month returns ($t + 1$). Panel A uses LDA topics + FRED-MD; Panel B uses FRED-MD only. This is an in-sample predictive diagnostic with fixed full-sample LDA topics. Reported coefficient t-statistics and p-values are conventional descriptive post-selection statistics. The out-of-sample tests in Sections 5 and 6 replace the fixed full-sample LDA measures with train-only oLDA measures. EV denotes electric vehicle.

For the aggregate series, the FRED-MD benchmark already explains a meaningful share of return variation, and adding narratives raises R^2 only modestly, from 0.11 to 0.13. The incremental role is larger in HY: R^2 rises from 0.07 to 0.10, and the selected active set includes Economic Downturn, Government Policy & Regulation, and Natural Disasters & Severe Incidents. Because this exercise uses full-sample LDA measures and same-sample variable selection, I treat it as supporting evidence rather than as the paper's main result. The spread-side counterpart is reported in Appendix Table A3.

4.4 Main evidence: CDGM-style residual credit-cycle states

Tables 4 and 5 provide the paper's central residual-side evidence. Following Collin-Dufresne et al. (2001), I remove from monthly spread changes the component explained by standard rates, slope, equity-market, and option-implied controls. I then ask two questions. First, how much common variation remains across ratings after these controls? Second, is the relation between WSJ narratives and that residual variation sparse and concentrated in risky credit, or distributed more diffusely across the broader topic space?

Table 4 shows that the Main controls explain 15.6 percent of monthly spread-change variation for A, 27.9 percent for BBB, 28.2 percent for ALL, and 38.3 percent for HY. Adding aggregate leverage changes adjusted R^2 by about one-tenth of a percentage point or less in every rating bucket, so the residual component is not sensitive to this baseline extension.[7] Before extracting the residual principal component, I standardize the A, BBB, and HY residuals to zero mean and unit variance so that the common component is not mechanically driven by the higher volatility of lower-grade spreads. PC1 explains 76.26 percent of standardized residual variation under the Main specification and 76.21 percent under +Leverage. Its weights remain similar across A, BBB, and HY. Standard credit-risk controls therefore leave a broad common residual state rather than a factor mechanically dominated by HY.

Table 5 separates three different questions about how news maps into that residual state. All three panels use the same 408 months from 1990-02 through 2024-01. Panel A asks whether each residual state admits a sparse topic representation under the baseline LASSO specification. Panel B holds the twelve HY-selected topics fixed and applies exactly the same information set to PC1, A, BBB, and HY. Panel C removes the sparsity restriction and uses Ridge to map the complete $K = 180$ topic space. These exercises are historical in-sample mappings.

Panel A provides a full-sample sparse representation of the residual–text relation. Under the global normalization used for this historical static-LDA diagnostic, no topic survives selection for the standardized residual PC1, A, or BBB residuals. HY is the only residual series with a nonempty active set: twelve topics are selected, with adjusted $R^2 = 0.0955$ under the Main residual and 0.0957 under +Leverage when the Main-selected set is held fixed. The corresponding R^2 values are 0.1221 and 0.1223. I interpret this result as a parsimonious full-sample summary rather than as evidence that the exact HY active set is invariant to alternative regularization or preprocessing schemes. The contrast with Table 4 remains informative: the residual covariance contains a broad common state, while a low-dimensional representation under the baseline sparse diagnostic appears only in HY.

Panel B asks a different question: whether the HY-selected information set is relevant only for HY. Holding the same twelve topics fixed without re-selection yields Main adjusted R^2 of 6.23

[7] Spread series begin earlier, but the baseline sample starts in 1990 once option-implied controls enter.

percent for PC1, 3.99 percent for A, 1.62 percent for BBB, and 9.55 percent for HY. The +Leverage results are nearly unchanged. These values are not monotonic across ratings. Rather, the same narratives that form a sparse HY representation also map some variation in the broader residual states. HY therefore differs in the sparsity of its representation, not because the selected narratives are economically irrelevant elsewhere.

Panel C shows why the distinction matters. When all 180 topic measures enter jointly under Ridge regularization, Main in-sample R^2 is 11.08 percent for PC1, 1.92 percent for A, 7.55 percent for BBB, and 8.90 percent for HY. Among the rating-specific residuals, the diffuse full-topic fit therefore rises monotonically from A to BBB to HY. Using the same Ridge penalty with the leverage-augmented residuals leaves this ordering essentially unchanged. This complements Panel A: diffuse narrative fit is weak in A and stronger in BBB and HY, while under the baseline full-sample sparse diagnostic only HY yields a nonempty active set. The PC1 result further shows that the broad common residual state contains substantial narrative information in the full topic space even though no small active set is selected for PC1. Because these are global in-sample Ridge fits using static full-sample LDA measures, I interpret the pattern as descriptive evidence about the distribution of narrative information.

Table 4. CDGM-style Baseline Spread-Change Models and Residual Common Factor

Panel A. Baseline spread-change model fit

Rating bucket	N	Adj. R^2 (Main)	Adj. R^2 (+Leverage)	Δ Adj. R^2
ALL	428	0.2815	0.2813	-0.0002
A	428	0.1558	0.1548	-0.0010
BBB	428	0.2793	0.2796	0.0003
HY	429	0.3832	0.3841	0.0009

Panel B. Principal component analysis of standardized residuals

Series	PC1 weight (Main)	PC1 weight (+Leverage)	Notes
A residual	0.5871	0.5873	PC1 explained variance ratio (Main) = 0.7626
BBB residual	0.6018	0.6018	PC1 explained variance ratio (+Leverage) = 0.7621
HY residual	0.5414	0.5412	PC1 sample: 1990-02 to 2025-09; N = 428

Notes: Panel A reports CDGM-style spread-change regressions using standard rates, slope, equity-market, and option-implied controls, with and without an aggregate leverage term. The ALL, A, and BBB regressions use 428 months from 1990-02 through 2025-09; HY has one additional available month through 2025-10. In Panel B, the A, BBB, and HY residuals are standardized to zero mean and unit variance before principal component analysis. The reported entries are first-component eigenvector weights. PC1 explains about 76 percent of standardized residual variation under both residual definitions and assigns similar weights across the three rating segments, indicating a broad common residual credit state rather than an exclusively HY factor. The PC1 uses the joint A–BBB–HY residual sample from 1990-02 through 2025-09.

Table 5. WSJ Narratives and CDGM-Style Residual Variation

Panel A. Sparse LASSO mapping on the 408-month sample

Dependent variable	N	Adj. R^2 Main	Adj. R^2 +Leverage	R^2 Main	R^2 +Leverage	Topics selected
Standardized residual PC1	408	0.0000	0.0000	0.0000	0.0000	0
A residual	408	0.0000	0.0000	0.0000	0.0000	0
BBB residual	408	0.0000	0.0000	0.0000	0.0000	0
HY residual	408	0.0955	0.0957	0.1221	0.1223	12

Notes: Panel A reports full-sample sparse text mappings from 1990-02 through 2024-01. Because the static-LDA exercise is used for historical construct validation, topic predictors and the dependent variable are standardized once over the full 408-month analysis sample. The LASSO penalty is selected using five-fold time-series cross-validation as a regularization rule, with a 12-topic cap and no fallback rule. Topics are selected under the Main residual specification and then held fixed when the corresponding leverage-augmented residual is used. No topic survives selection for the standardized residual PC1, A, or BBB residuals; twelve topics are selected for HY. Post-LASSO OLS is estimated in the original dependent-variable units with HAC standard errors.

Panel B. Fixed HY-selected information-set mapping

Dependent variable	N	Adj. R^2 Main	Adj. R^2 +Leverage	R^2 Main	R^2 +Leverage	Topics fixed
Standardized residual PC1	408	0.0623	0.0612	0.0899	0.0888	12
A residual	408	0.0399	0.0382	0.0682	0.0666	12
BBB residual	408	0.0162	0.0157	0.0452	0.0447	12
HY residual	408	0.0955	0.0957	0.1221	0.1223	12

Notes: Panel B holds the narrative information set fixed across residual states. The twelve topics selected once from the HY Main specification in Panel A are applied without re-selection to the standardized residual PC1 and to the A, BBB, and HY residuals over the identical 408-month sample. The +Leverage regressions use the same twelve topics. Differences across rows therefore measure descriptive fit for a common HY-selected information set rather than differences generated by variable selection. No additional LASSO selection or fallback rule is used. The comparison is descriptive and does not imply a causal or monotonic cross-rating relation.

Panel C. Full-topic Ridge mapping (descriptive)

Dependent variable	N	K	Effective df	R^2 Main	R^2 +Leverage	ΔR^2
Standardized residual PC1	408	180	31.77	0.1108	0.1087	-0.0020
A residual	408	180	7.51	0.0192	0.0184	-0.0008
BBB residual	408	180	27.20	0.0755	0.0738	-0.0017
HY residual	408	180	19.58	0.0890	0.0873	-0.0018

Notes: Panel C reports a global in-sample Ridge mapping using all $K = 180$ static-LDA topic measures over the same 408-month sample. All topics enter jointly; there is no variable selection or fallback rule. Topic predictors are standardized over the analysis sample. For each residual state, the Ridge penalty is chosen under the Main specification by five-fold time-series cross-validation and is then held fixed for the +Leverage residual. Effective df summarizes shrinkage-adjusted model flexibility at the Main tuning parameter. The raw penalty is omitted because its magnitude is scale dependent. The reported R^2 values are descriptive global-fit statistics; no Ridge coefficient p-values are reported, and the panel is not a real-time forecasting test.

5 Out-of-Sample Validation

The residual results are based on historical narrative measures and spread-change decompositions. I next ask whether the same information survives in real time. These tests are intentionally narrower than the residual exercise: they examine whether train-only narrative measures retain predictive content without look-ahead, not whether text alone generates a dominant trading strategy.

5.1 Return forecasts

Table 6 reports one-step-ahead return forecasts. Each month, I form predictors using only information available at that time, estimate the forecasting model on a rolling training window, and evaluate the next forecast month. Train-only narrative innovations enter the return-forecasting regression with a one-month lag, so the implemented forecast uses the previous month's narrative signal to predict next-month excess return.

Real-time predictive fit is concentrated in HY but remains modest in magnitude. Relative to the rolling-mean benchmark, HY out-of-sample R^2 is positive across the 60-, 84-, and 120-month training windows—0.025, 0.017, and 0.014—with hit rates between 58.1 and 62.6 percent. Aggregate bond forecasts produce negligible or negative gains. The cross-market pattern is more informative than the absolute magnitude: positive real-time fit appears in HY, consistent with the stronger historical text–residual mapping for risky credit in Table 5. Appendix Table A4 reports the contemporaneous as-of-t implementation.

Table 6. Out-of-Sample (OOS) Return Forecast Performance

Series	Train Win (m)	OOS Months	OOS Start	OOS End	OOS R^2 vs RollMean	OOS R^2 vs ExpMean	Utility Gain (%, ann.) vs RollMean	Utility Gain (%, ann.) vs ExpMean	Hit Rate
HY	60	353	1994-07	2023-11	0.025	-0.000	2.1	2.8	62.6%
ALL	60	353	1994-07	2023-11	-0.047	-0.063	-0.1	-0.1	59.5%
HY	84	329	1996-07	2023-11	0.017	0.007	3.2	3.0	58.1%
ALL	84	329	1996-07	2023-11	-0.039	-0.048	-0.3	-0.1	58.4%
HY	120	293	1999-07	2023-11	0.014	0.005	1.7	2.3	59.7%
ALL	120	293	1999-07	2023-11	-0.029	-0.029	0.9	1.1	59.7%

Notes: This table reports the paper's main lagged out-of-sample (OOS) return-forecast specification for the ALL and HY bond return series. Each row uses a rolling training window of length Train Win (m) to forecast next-month excess return. Train-only narrative innovations are constructed recursively and then entered into the forecasting regression with a one-month lag, so the implemented forecast uses the previous month's narrative signal to predict the next forecast month's excess return. Forecasts are produced using an exact-k LASSO ($k = 5$) with post-selection OLS, together with a small train-only shrinkage toward the rolling-mean baseline, where SSE denotes the sum of squared forecast errors. OOS R^2 is computed against the corresponding rolling-mean and expanding-mean benchmarks as $1 - \mathrm{SSE}_{model}/\mathrm{SSE}_{baseline}$. Utility gain is the annualized certainty-equivalent return gain of the risk-managed strategy based on the model forecast relative to the corresponding baseline strategy, following Campbell and Thompson (2008), using $w_t = \mathrm{clip}[f_t/(\gamma\hat{\sigma}_t^2), 0, w_{cap}]$ with $\gamma = 5$, $w_{cap} = 1.5$, and $\hat{\sigma}_t^2$ estimated from the trailing 36 months of realized returns using training information only. Hit Rate is the fraction of months where $\mathrm{sign}(f_t)$ matches $\mathrm{sign}(y_{t+1})$.

5.2 Timing strategies

Table 7 asks whether the train-only signal has economic content inside a risk-managed implementation. Forecasts are formed out of sample using a rolling 84-month training window. The text signal enters as a bounded exposure overlay rather than as a sign-flipping trade. I also include a spread-based risk filter that gradually reduces exposure when credit conditions deteriorate.[8] The design separates three objects: buy-and-hold, the spread-based filter, and the incremental text overlay.

In HY, the spread-based filter accounts for most of the performance gain: the Sharpe ratio rises from 0.495 for buy-and-hold to 0.783 for RiskFilter-only, while beta to buy-and-hold falls to 0.453. Adding the text overlay raises the Sharpe ratio further to 0.823. The combined strategy has annualized alpha of 2.2% ($t = 3.62$) relative to buy-and-hold, but only 0.3% ($t = 1.79$) relative to RiskFilter-only. The post-2008 sample produces the same ordering: the combined strategy reaches a Sharpe ratio of 0.885, with alpha of 2.4% ($t = 2.78$) relative to buy-and-hold and 0.5% ($t = 1.72$) relative to RiskFilter-only. The text signal therefore refines exposure within credit states already identified by observable spreads rather than replacing the spread signal.

The result is robust to moderate transaction costs, nearby risk-filter thresholds, and alternative topic counts, as reported in Appendix Tables A8, A9, and A10. Appendix Table A5 reports the detrending grid. For ALL bonds, the same architecture is much weaker and the spread filter reduces performance, consistent with duration and policy-rate exposure being more important than default-sensitive credit states in the broad investment-grade market. This cross-market contrast is also consistent with Table 5, where the historical text–residual evidence is strongest in HY and the diffuse full-topic fit increases toward lower credit quality.

[8] The risk filter thresholds are not optimized on realized returns. They are set to reflect a conservative prior: exposure is scaled down when credit spreads are unusually high relative to their expanding historical distribution, with delevering beginning around a spread z-score of 1 and reaching its minimum by a z-score of 2. Exposure is also scaled down when monthly spread widening is positive and reaches its minimum when the widening is about 50 basis points. This is an implementation choice rather than a fitted parameter in the forecasting model.

Table 7. Out-of-Sample Performance of Timing Strategies

					Relative to Buy-and-hold			Relative to RiskFilter-only		
Strategy	Series	Sample	Months	Sharpe	$\text{Alpha}_{\text{ann}}$	$t(\alpha)$	Beta	$\text{Alpha}_{\text{ann}}$	$t(\alpha)$	Beta
Buy-and-hold (HY)	HY	FULL	325	0.495	—	—	1.000	—	—	—
Text Overlay only (HY)	HY	FULL	325	0.588	0.009	2.35	0.859	—	—	—
RiskFilter-only (HY)	HY	FULL	325	0.783	0.019	3.13	0.453	—	—	—
Text Overlay + RiskFilter (HY)	HY	FULL	325	0.823	0.022	3.62	0.420	0.003	1.79	0.980
Buy-and-hold (HY) post-2008	HY	POST2008	191	0.583	—	—	1.000	—	—	—
Text Overlay only (HY) post-2008	HY	POST2008	191	0.731	0.015	3.21	0.820	—	—	—
RiskFilter-only (HY) post-2008	HY	POST2008	191	0.832	0.019	2.16	0.466	—	—	—
Text Overlay + RiskFilter (HY) post-2008	HY	POST2008	191	0.885	0.024	2.78	0.419	0.005	1.72	0.954
Buy-and-hold (ALL)	ALL	FULL	325	0.533	—	—	1.000	—	—	—
Text Overlay only (ALL)	ALL	FULL	325	0.566	0.003	1.53	0.910	—	—	—
RiskFilter-only (ALL)	ALL	FULL	325	0.316	-0.009	-2.27	0.808	—	—	—
Text Overlay + RiskFilter (ALL)	ALL	FULL	325	0.362	-0.006	-1.85	0.751	0.003	1.47	0.937
Buy-and-hold (ALL) post-2008	ALL	POST2008	191	0.515	—	—	1.000	—	—	—
Text Overlay only (ALL) post-2008	ALL	POST2008	191	0.521	0.002	0.56	0.879	—	—	—
RiskFilter-only (ALL) post-2008	ALL	POST2008	191	0.327	-0.010	-1.55	0.837	—	—	—
Text Overlay + RiskFilter (ALL) post-2008	ALL	POST2008	191	0.376	-0.005	-1.12	0.762	0.003	1.19	0.919

Notes: This table reports out-of-sample performance for timing strategies built from narrative-based forecasts and an external risk filter. Buy-and-hold holds the broad bond market series each month. Text Overlay applies forecast-based exposure scaling to the strategy return. RiskFilter applies the spread-based risk filter without using the text forecast. Text Overlay + RiskFilter combines the two. $\text{Alpha}_{\text{ann}}$ is annualized alpha relative to the benchmark shown in the corresponding column; $t(\alpha)$ is the associated t-statistic; beta is the slope coefficient with respect to that benchmark. "FULL" uses the full out-of-sample window, while "POST2008" restricts the sample to the post-2008 subsample. All statistics are computed strictly out-of-sample using information available at time t. The risk filter is constructed using spread data available from 1996 onward; to avoid contamination from pre-1996 backfilled spreads, the evaluation window is restricted to months from 1996 onward.

6 Out-of-Sample Horse Race: Text versus Macro Information Sets

The final validation exercise asks whether the text signal is spanned by standard macro information sets. Table 8 (Panels A and B) compares timing strategies built from Text, FRED-MD, Goyal–Welch (GW), Economic Policy Uncertainty (EPU), Macro-core, and Text+Macro predictors under the same implementation.

In HY, Text raises the Sharpe ratio from 0.494 for buy-and-hold to 0.587 and reduces maximum drawdown from −36.1% to −29.6%. Macro-core and GW are competitive, with Sharpe ratios of 0.579 and 0.567. Text is therefore not dominated by the standard macro information sets, although the failure of Text+Macro to improve on the separate models suggests substantial overlap. For ALL, text adds little and EPU is more competitive, consistent with the greater importance of rates and policy uncertainty in broad investment-grade returns.

Across investment-grade ratings, text performs best in BBB: the Sharpe ratio rises from 0.537 to 0.628 and maximum drawdown falls from −21.4% to −14.4%. A-rated bonds are intermediate, whereas the narrative signal is weak in AAA/AA and EPU performs better. The same ordering holds in the post-2008 subsample reported in Appendix Table A6 (Panels A and B). This cross-rating ordering matches the residual evidence: narrative information is most useful in BBB and HY, where returns are more sensitive to credit-cycle states.

Table 8. Horse Race — Text vs Macro (Panel A: Full Sample)

	Metric	Buy-and-hold	Text	FRED-MD	GW	EPU	Macro-core	Text+Macro
HY	Sharpe ratio	0.494	0.587	0.534	0.567	0.430	0.579	0.527
	Expected return	4.536	4.817	4.648	5.263	3.902	5.264	4.730
	t-stat	2.187	2.407	2.210	2.226	1.932	2.376	2.181
	Information ratio	—	0.108	0.040	0.260	-0.183	0.263	0.079
	Alpha (ann., %)	—	0.921	0.550	0.889	-0.257	0.984	0.459
	Alpha t-stat	—	2.351	1.106	0.957	-0.314	1.497	1.259
	Max loss	-36.146	-29.629	-29.034	-30.366	-39.524	-29.975	-35.961
	Skew	-0.973	-0.683	-1.134	-0.346	-3.041	-0.864	-1.062
	Months	325	325	325	325	325	325	325
ALL	Sharpe ratio	0.532	0.566	0.497	0.489	0.583	0.501	0.497
	Expected return	3.191	3.197	3.254	3.049	3.501	3.210	3.031
	t-stat	2.494	2.635	2.316	2.311	2.870	2.292	2.235
	Information ratio	—	0.004	0.035	-0.088	0.174	0.009	-0.086
	Alpha (ann., %)	—	0.294	-0.096	-0.156	0.447	-0.037	-0.059
	Alpha t-stat	—	1.531	-0.305	-0.382	1.477	-0.117	-0.212
	Max loss	-22.434	-20.983	-21.179	-22.179	-18.411	-21.209	-23.328
	Skew	-0.541	-0.560	-0.722	-1.178	-0.896	-0.917	-1.147
	Months	325	325	325	325	325	325	325

Notes: This table reports out-of-sample timing performance in a horse race between text and macro information sets. Each column corresponds to a strategy built from a different information set: Buy-and-hold, Text, FRED-MD, GW, EPU, Macro-core, and Text+Macro. Rows report annualized Sharpe ratio, expected return, the Newey–West t-statistic of the mean, information ratio and alpha relative to Buy-and-hold (where applicable), maximum drawdown ("Max loss"), skewness, and the number of months. Panel A reports results for HY and the broad ALL aggregate; Panel B (Table 8 continued) reports results separately by investment-grade rating. All quantities are computed strictly out-of-sample using information available at time t.

Table 8. Horse Race — Text vs Macro by Rating (Panel B: Full Sample)

	Metric	Buy-and-hold	Text	FRED-MD	GW	EPU	Macro-core	Text+Macro
BBB	Sharpe ratio	0.537	0.628	0.519	0.452	0.493	0.532	0.540
	Expected return	3.466	3.714	3.639	3.245	3.379	3.723	3.504
	t-stat	2.440	2.950	2.226	2.156	2.436	2.288	2.368
	Information ratio	—	0.123	0.078	-0.114	-0.040	0.116	0.018
	Alpha (ann., %)	—	0.696	0.066	-0.476	-0.106	0.158	0.202
	Alpha *t*-stat	—	1.828	0.176	-0.881	-0.183	0.420	0.533
	Max loss	-21.443	-14.399	-20.943	-23.401	-21.364	-20.892	-19.151
	Skew	-1.039	-0.666	-0.892	-2.153	-2.848	-0.855	-1.277
	Months	325	325	325	325	325	325	325
A	Sharpe ratio	0.477	0.519	0.452	0.468	0.554	0.457	0.474
	Expected return	2.868	3.030	2.935	2.853	3.253	2.929	3.038
	t-stat	2.265	2.392	2.221	2.174	2.752	2.152	2.123
	Information ratio	—	0.102	0.034	-0.009	0.216	0.030	0.088
	Alpha (ann., %)	—	0.343	-0.018	0.059	0.577	0.029	0.124
	Alpha *t*-stat	—	1.973	-0.041	0.149	2.093	0.078	0.391
	Max loss	-23.430	-23.177	-22.323	-22.557	-19.799	-21.441	-24.140
	Skew	-0.607	-0.728	-1.604	-1.171	-0.082	-1.787	-1.581
	Months	325	325	325	325	325	325	325
AAA/AA	Sharpe ratio	0.440	0.465	0.475	0.447	0.545	0.505	0.457
	Expected return	2.463	2.542	2.668	2.666	2.904	2.746	2.613
	t-stat	2.126	2.006	2.473	2.181	2.701	2.417	2.157
	Information ratio	—	0.051	0.120	0.132	0.270	0.161	0.088
	Alpha (ann., %)	—	0.230	0.308	0.130	0.659	0.472	0.213
	Alpha *t*-stat	—	1.037	0.820	0.550	2.484	1.299	0.735
	Max loss	-26.581	-29.017	-21.724	-27.485	-21.779	-21.389	-24.844
	Skew	-0.214	-0.558	-0.410	0.362	0.433	-0.705	-0.627
	Months	325	325	325	325	325	325	325

Notes: This table reports out-of-sample timing performance in a horse race between text and macro information sets, separately by credit rating: AAA/AA, A, and BBB. Each column corresponds to a strategy built from a different information set: Buy-and-hold, Text, FRED-MD, GW, EPU, Macro-core, and Text+Macro. Rows report annualized Sharpe ratio, expected return, the Newey–West *t*-statistic of the mean, information ratio and alpha relative to Buy-and-hold (where applicable), maximum drawdown ("Max loss"), skewness, and the number of months. Panel B decomposes the investment-grade segment into AAA/AA, A, and BBB. The sample is the full out-of-sample window. All quantities are computed strictly out-of-sample using information available at time t.

7 Interpreting the oLDA Signal

This section interprets the train-only text signal. The goal is not to claim a fully identified mechanism, but to verify that the narratives entering the out-of-sample forecasts correspond to plausible credit-cycle channels. Because the refit-and-align pipeline preserves topic identity across refits, I decompose forecasts into topic-level contributions and examine which narratives repeatedly enter the HY and BBB timing signals.

7.1 Stable "important topics" decomposition

Table 9 (Panels A and B) decomposes each post-2008 forecast into topic-level contributions,

$$\text{Contribution}_{k,t} = \hat{\beta}_{k,t} \times z(nd_{k,t}),$$

where $\hat{\beta}_{k,t}$ is the post-selection OLS coefficient on topic k in month t, and $z(nd_{k,t})$ is the train-only standardized topic loading. I aggregate monthly contributions into topic-level importance measures.[9]

To reduce sensitivity to single episodes, Table 9 (Panels A and B) reports topics that are important under two rankings: winsorized total absolute contribution (W99) and a trim-one alternative (TRIM1).[10] The resulting topics are best interpreted as recurring narrative channels rather than as causal mechanisms. HY and BBB share a core of discount-rate, intermediary, and financing narratives, but differ in the channels that appear at the margin.

7.1.1 Shared anchor topics across HY and BBB

The three shared topics span discount-rate news (T0041, Yield Curve & Fed Policy), intermediary stress (T0039, Bank Loan Losses & Nonperforming Assets), and financing conditions (T0126, Convertible Debentures & Notes). T0041 loads more heavily in BBB, consistent with greater duration sensitivity; T0039 rises in stress episodes; and T0126 appears in bursts around refinancing events. Together, these anchors tie the timing signal to economically interpretable credit-cycle channels rather than a generic news factor. In the topic labels below, PDMR denotes a "person discharging managerial responsibilities," the standard regulatory terminology.

[9] The standardized topic loading is the topic attention measure, or its train-only innovation in detrended robustness specifications, centered and scaled by training-window moments.

[10] W99 denotes the winsorized total absolute contribution ranking, in which monthly absolute contributions are winsorized before aggregation to limit the influence of extreme single-month spikes. TRIM1 denotes a trim-one ranking that recomputes topic importance after removing the single largest monthly absolute contribution for each topic. The overlap between the two rankings is used to identify a stable common set of important topics.

Table 9. Important oLDA Topics for HY Timing (Panel A: Post-2008, Stable Common Set)

Topic ID	Label	Freq	Sign	Importance (W99)	Dominance Ratio
T0112	Greater China & Hong Kong	29.1%	+	24.819	0.070
T0148	European PDMR & Auction Notices	24.0%	–	21.727	0.071
T0126	Convertible Debentures & Notes	37.4%	–	20.512	0.045
T0041	Yield Curve & Fed Policy	25.7%	+	12.326	0.093
T0089	Biopharma Clinical Trials	11.7%	–	11.704	0.075
T0058	Oil Refining & Petroleum Markets	48.0%	+	9.201	0.063
T0034	Electric Utilities & Power Generation	12.3%	+	7.754	0.128
T0147	Bond Tender Offers & Consents	26.8%	–	7.227	0.046
T0081	Indenture Covenants & Note Redemption	23.5%	+	6.799	0.052
T0039	Bank Loan Losses & Nonperforming Assets	29.1%	–	6.410	0.033

Notes: Panel A reports the post-2008 HY oLDA topic display table constructed from the intersection of the winsorized (W99) and trim-1 ranking lists, emphasizing stability across importance definitions. Topic ID and label identify each topic. Freq is the proportion of post-2008 months in which the topic is selected into the month-specific post-OLS decomposition. Sign is the sign of the topic's average contribution (beta × standardized topic loading) across selected months. Importance is the topic's total absolute contribution, winsorized at the 1st and 99th percentiles, used as the primary display importance measure. Dominance Ratio equals $\max_t|\text{contribution}_t| / \sum_t|\text{contribution}_t|$ within topic, so lower values indicate the topic is less driven by one or two episodic months. Top terms are omitted here by design and are reported in Figure 2.

Table 9. Important oLDA Topics for BBB Timing (Panel B: Post-2008, Stable Common Set)

Topic ID	Label	Freq	Sign	Importance (W99)	Dominance Ratio
T0148	European PDMR & Auction Notices	25.1%	–	34.330	0.056
T0126	Convertible Debentures & Notes	34.6%	–	17.944	0.083
T0041	Yield Curve & Fed Policy	39.1%	+	14.964	0.070
T0093	Oilfield Services & Drilling Activity	12.8%	+	13.366	0.382
T0082	Securities Class Actions & Litigation	32.4%	+	11.443	0.089
T0039	Bank Loan Losses & Nonperforming Assets	38.0%	–	8.818	0.031
T0118	Canadian Banking	9.5%	+	7.648	0.105
T0161	Trading halts / market microstructure stress	20.1%	–	7.332	0.096
T0141	Food Retail & Inflation	14.5%	+	7.247	0.108

Notes: Panel B reports the post-2008 BBB oLDA topic display table constructed from the intersection of the winsorized (W99) and trim-1 ranking lists, emphasizing stability across importance definitions. Topic ID and label identify each topic. Freq is the proportion of post-2008 months in which the topic is selected into the month-specific post-OLS decomposition. Sign is the sign of the topic's average contribution (beta × standardized topic loading) across selected months. Importance is the topic's total absolute contribution, winsorized at the 1st and 99th percentiles, used as the primary display importance measure. Dominance Ratio equals $\max_t|\text{contribution}_t| / \sum_t|\text{contribution}_t|$ within topic, so lower values indicate the topic is less driven by one or two episodic months. Top terms are omitted here by design and are reported in Figure 2.

7.2 Visualizing the Narrative Channels Behind Timing

Figure 2 (Panels A and B) plots the time-series evolution and display terms for the selected narratives.[11] The shared anchors identified in the preceding subsection are visible in both rating segments: Yield Curve & Fed Policy (T0041), Bank Loan Losses & Nonperforming Assets (T0039), and Convertible Debentures & Notes (T0126). The remaining panels highlight the channels that differ between HY and BBB.

The cross-rating differences are also consistent with the residual evidence. In HY, the decomposition highlights narratives tied to distress resolution and debt-contract frictions, including Bond Tender Offers and Consents (T0147) and Indenture Covenants and Note Redemption (T0081). These topics contain liability-management and creditor-rights language, such as tender offers, consent solicitations, indentures, covenants, redemptions, and subordination. They are natural proxies for rollover risk, renegotiation risk, and creditor bargaining, which should matter most in high yield (Gilson et al., 1990).

BBB tilts toward market-friction and uncertainty channels. Trading Halts and Market Microstructure Stress (T0161) points to liquidity disruptions, consistent with evidence that corporate bond liquidity is priced and becomes especially salient during stress episodes (Bao et al., 2011). Securities Class Actions and Litigation (T0082) is an uncertainty topic, while Food Retail and Inflation (T0141) links BBB to inflation-sensitive macro pricing. These channels fit BBB's position at the edge of investment grade: less default-intensive than HY, but more credit-sensitive than safer bonds.

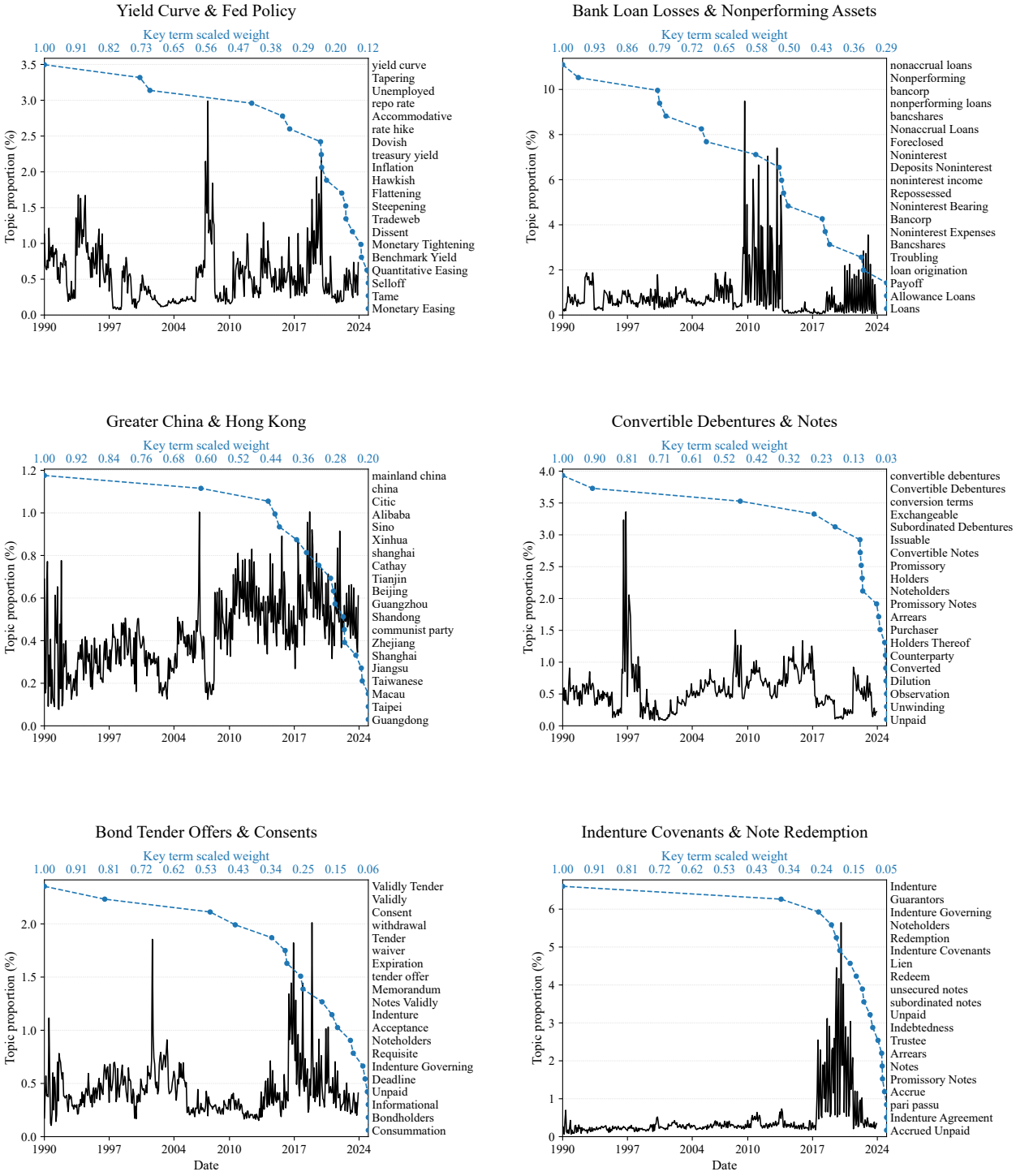
7.3 Topic Alignment Stability

A natural concern is whether topic identities remain stable enough to support this interpretation. Figure 3 reports system-level diagnostics for the refit-and-align procedure. Across adjacent refits, aligned same-ID topics have high average cosine similarity, positive diagonal-minus-best-off-diagonal gaps, frequent row-wise best matches, and near-zero permutation p-values for the observed diagonal mean. These diagnostics do not imply that every topic is unchanged at every refit, but they indicate that the persistent topic IDs retain enough semantic continuity for the out-of-sample decomposition to be interpretable.

Table 10 focuses on the specific Figure 2 narratives. Most satisfy a conservative stability screen across all four diagnostics. A post-2008 version in Appendix Table A7 leads to the same qualitative conclusion, although continuity is weaker for a subset of topics. These results support interpreting Figure 2 as tracing consistent narrative channels across refits.

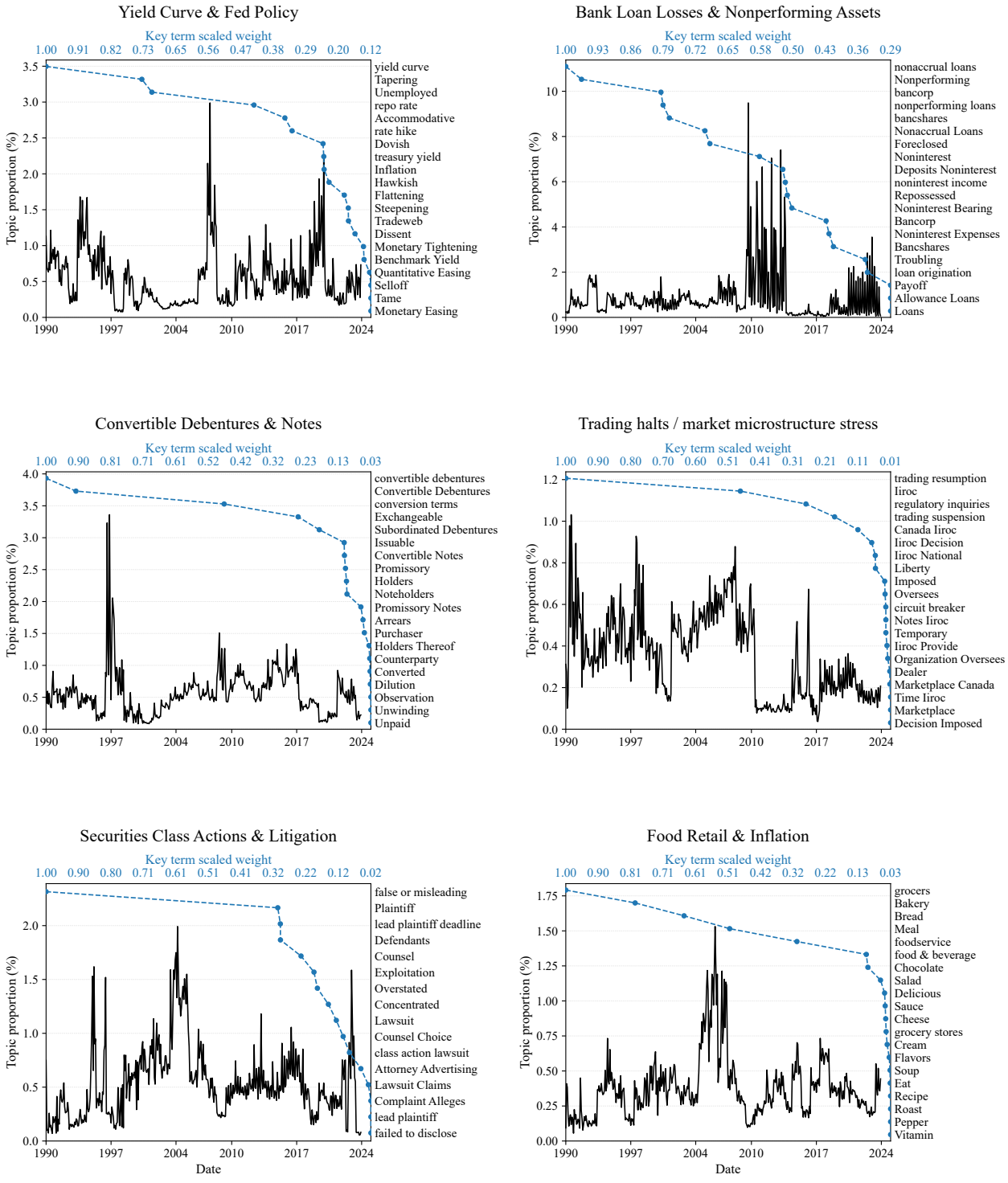
[11] The term display is produced by an "audit-polishing" step: a display-only cleaning of key-term lists, including the removal of boilerplate fragments, the demotion of proper-name dominance, and the merging of redundant variants. This step does not alter model estimation, predictor construction, or any out-of-sample forecast.

Figure 2. Topic Attention for Selected oLDA Topics (Panel A: HY Timing Channels)



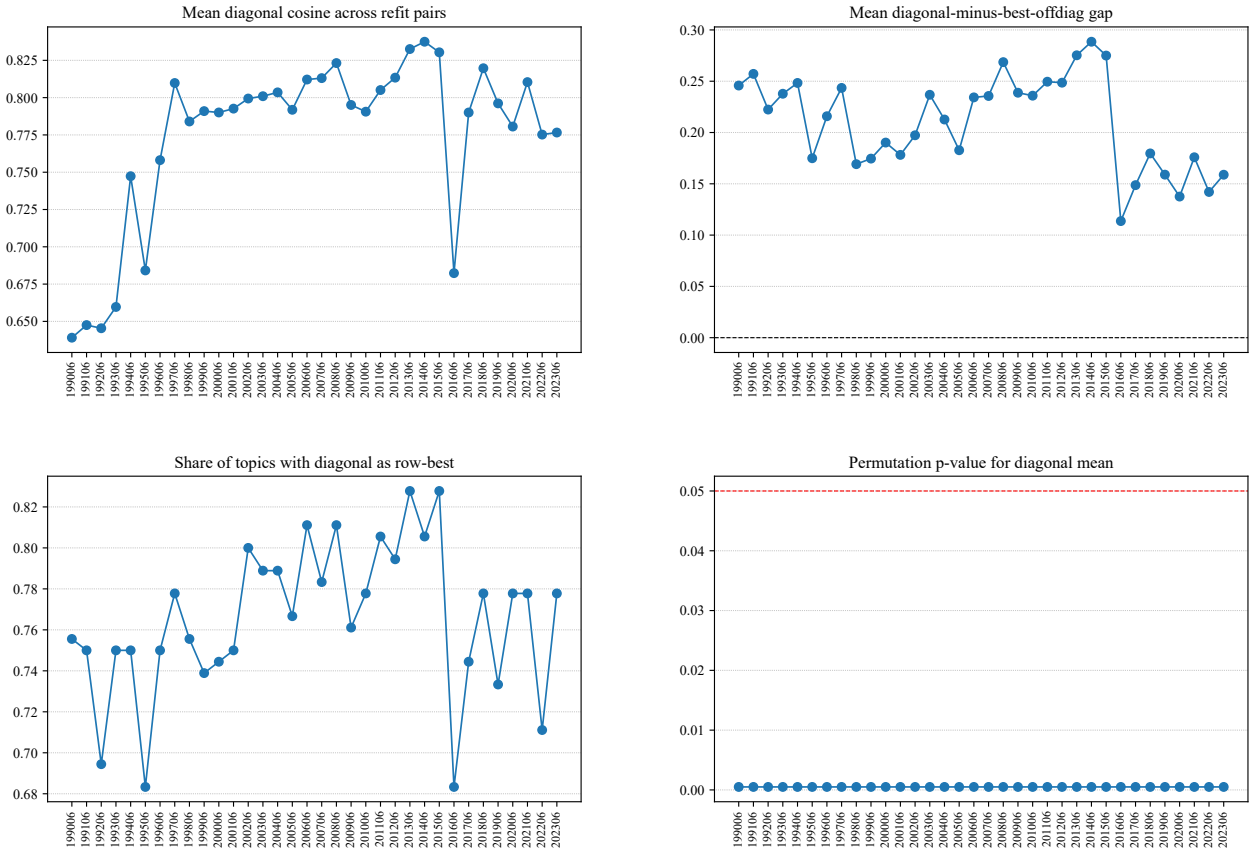
Notes: Each panel plots the monthly share of a selected oLDA topic from the *Wall Street Journal* corpus (black line). The terms shown on the right are the topic's displayed key terms after display-side cleaning and merging, ordered by within-topic weight, and the dashed blue curve reports their relative weights scaled by the largest displayed term weight within that topic, so the horizontal scale runs from 0 to 1. Panel A displays the selected HY timing-channel topics from the strict train-only oLDA pipeline. The figure traces the narratives that enter the out-of-sample HY timing signal, not the static full-sample LDA of Figure 1. The displayed terms are used for interpretation only and do not alter the underlying topic-attention series.

Figure 2. Topic Attention for Selected oLDA Topics (Panel B: BBB Timing Channels)



Notes: Panel B displays the selected BBB timing-channel topics from the strict train-only oLDA pipeline. The figure traces the narratives that enter the out-of-sample BBB timing signal, not the static full-sample LDA of Figure 1. The displayed terms are used for interpretation only and do not alter the underlying topic-attention series.

Figure 3. Alignment Audit — Overall Refit-to-Refit Diagnostics



Note: This figure reports overall alignment diagnostics across adjacent refit pairs in the oLDA refit-and-align pipeline. The upper-left panel plots the mean diagonal cosine similarity between aligned same-ID topics across adjacent refits. The upper-right panel plots the mean diagonal-minus-best-off-diagonal gap, where larger positive values indicate the aligned same-ID topic is more distinct than the best alternative topic in the same refit snapshot. The lower-left panel reports the share of topics for which the diagonal match is the row-wise best match. The lower-right panel reports the permutation p-value for the observed diagonal mean relative to random pair matching. Together, these panels provide a compact system-level diagnostic of out-of-sample topic continuity.

Table 10. Figure 2 Topic Alignment Stability Screen: Full Sample

Topic ID	Figure 2 Topic	*N*	Mean Diag. Cosine	Mean Gap	Share Row-Best	Mean Top20 Overlap	Score	Soft Pass
T0041	Yield Curve & Fed Policy	34	0.721	0.133	0.676	0.535	4/4	Yes
T0039	Bank Loan Losses & Nonperforming Assets	34	0.820	0.093	0.794	0.631	4/4	Yes
T0112	Greater China & Hong Kong	34	0.787	0.359	0.882	0.579	4/4	Yes
T0126	Convertible Debentures & Notes	34	0.798	0.230	0.794	0.628	4/4	Yes
T0147	Bond Tender Offers & Consents	34	0.902	0.366	0.941	0.759	4/4	Yes
T0081	Indenture Covenants & Note Redemption	34	0.738	0.255	0.794	0.575	4/4	Yes
T0161	Trading halts / market microstructure stress	34	0.693	0.294	0.735	0.562	4/4	Yes
T0082	Securities Class Actions & Litigation	34	0.706	0.198	0.706	0.631	4/4	Yes
T0141	Food Retail & Inflation	34	0.510	0.074	0.559	0.431	2/4	No

Notes: This table reports a full-sample stability screen for the Figure 2 narrative-channel topics under the refit-and-align out-of-sample pipeline. Each row corresponds to one topic displayed in Figure 2, using the same topic IDs and display titles. *N* is the number of adjacent refit pairs used in the screen over the full displayed sample. Mean Diag. Cosine is the average cosine similarity between the topic's aligned term-distribution vector in adjacent refit snapshots. Mean Gap equals the average diagonal cosine minus the best off-diagonal cosine within the same row, so larger positive values indicate that the aligned same-ID topic is more distinct from nearby alternatives. Share Row-Best is the proportion of refit pairs in which the diagonal match is the row-wise best match. Mean Top20 Overlap is the average overlap rate of the topic's top-20 terms across adjacent refits. Score counts how many of the four soft-screen thresholds are satisfied: mean diagonal cosine ≥ 0.65, mean gap ≥ 0.05, share row-best ≥ 0.60, and mean top-20 overlap ≥ 0.30. Soft Pass indicates that a topic satisfies at least three of the four thresholds. The table is intended as a compact tractability screen for the Figure 2 narratives rather than a claim that every displayed topic is perfectly unchanged at every refit.

8 Conclusion

This paper shows that standard credit-risk observables leave a broad common residual state in corporate bond markets and that public news narratives help characterize its content. After CDGM-style controls, the first principal component of standardized A, BBB, and HY spread-change residuals explains about 76 percent of their variation and assigns similar weights across ratings. The textual mapping of that state varies across ratings. In the full-sample sparse diagnostic, a nonempty active set appears only for HY, while the full-topic Ridge specification reveals a monotonic diffuse-fit gradient from A to BBB to HY. Taken together, the results distinguish a broad common residual state from its rating-specific textual representation: narrative information becomes stronger toward risky credit, although the exact low-dimensional representation is treated as descriptive rather than structurally identified.

The real-time evidence is consistent with this interpretation but smaller in magnitude. HY return forecasts produce positive out-of-sample R^2, whereas aggregate bond forecasts are weak. In HY timing, observable spreads account for most of the improvement, while the narrative overlay adds a modest incremental adjustment. Text also remains competitive with FRED-MD, Goyal–Welch, and EPU information sets, particularly in BBB and HY. These results support treating narratives as supplementary measures of credit-cycle states rather than as substitutes for standard credit-market observables.

The contribution is therefore informational rather than algorithmic: topic-level public narratives help characterize and interpret credit-market variation that conventional observables do not fully capture. The results also show that this information need not appear as the same low-dimensional representation across rating segments. Extending the narrative-state measures to bond-level pricing would permit sharper tests involving issuer characteristics, liquidity, and contractual features.

References

Bao, J., Pan, J., and Wang, J. (2011). The illiquidity of corporate bonds. *The Journal of Finance*, 66(3):911–946.

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

Collin-Dufresne, P., Goldstein, R. S., and Martin, J. S. (2001). The determinants of credit spread changes. *The Journal of Finance*, 56(6):2177–2207.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Fang-Klingler, J. (2019). Impact of readability on corporate bond market. *Journal of Risk and Financial Management*, 12(4):184.

Friewald, N. and Nagler, F. (2019). Over-the-counter market frictions and yield spread changes. *The Journal of Finance*, 74(6):3217–3257.

Gertler, M. and Lown, C. S. (1999). The information in the high-yield bond spread for the business cycle: Evidence and some implications. *Oxford Review of Economic Policy*, 15(3):132–150.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720.

Gilson, S. C., John, K., and Lang, L. H. P. (1990). Troubled debt restructurings: An empirical study of private reorganization of firms in default. *Journal of Financial Economics*, 27(2):315–353.

Greenwood, R. and Hanson, S. G. (2013). Issuer quality and corporate bond returns. *Review of Financial Studies*, 26(6):1483–1525.

Hoffman, M. D., Bach, F. R., and Blei, D. M. (2010). Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems*, volume 23, pages 856–864.

International Monetary Fund (2017). Is growth at risk? Global Financial Stability Report, October, Chapter 3.

Isakin, M. and Pu, X. (2023). Dispersion in news sentiment and corporate bond returns. *International Review of Financial Analysis*, 89:102761.

Khoja, M. (2024). Large language models, information processing, and post-earnings-announcement drift in corporate bonds. Working paper.

Lin, H., Wu, C., and Zhou, G. (2018). Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. *Management Science*, 64(9):4218–4238.

Liu, M., Wang, J., and Wu, C. (2024). Media coverage and corporate bond momentum. Working paper.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

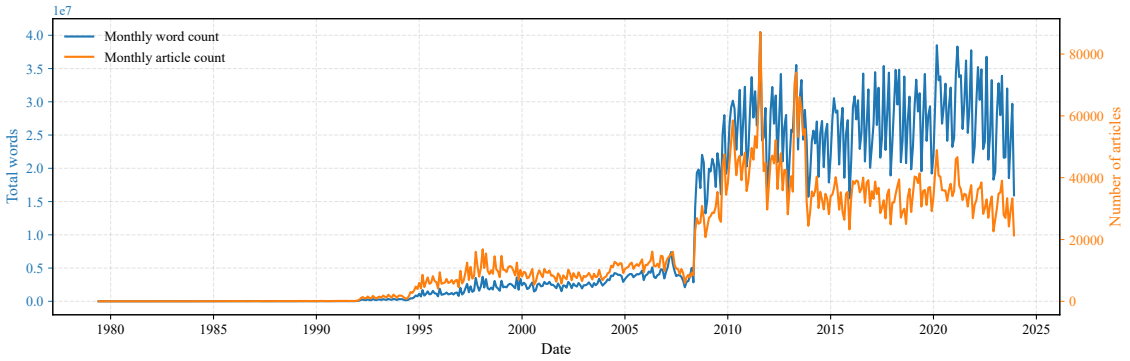
Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Appendix:

Figure A1: Monthly Word and Article Counts – Full-article Filter

Appendix Figure A1 reports the monthly word count and monthly article count in the rebuilt Wall Street Journal archive after applying the full-article filter that retains articles with at least 100 tokens. The figure shows that the usable text sample is sparse in the early years, expands gradually through the 1990s and 2000s, and becomes much denser after the late 2000s. This pattern is descriptive, but it is useful for interpreting the later results because it shows that the scale of the underlying news archive varies substantially across subperiods.

Figure A1. Monthly Word and Article Counts — Full-article Filter



Notes: Monthly word count and article count in the rebuilt *Wall Street Journal* archive after applying the full-article filter that retains articles with at least 100 tokens.

Table A1: All Bonds – Return

Table A1 reports the ALL-bond counterpart to the main return diagnostics and compares the paper's forward-looking specification, in which month-t narrative attention is related to month $t + 1$ excess returns, with a same-month benchmark. This comparison is included as a supporting appendix exercise to show that the broad narrative mapping for ALL bonds is not driven entirely by the timing convention.

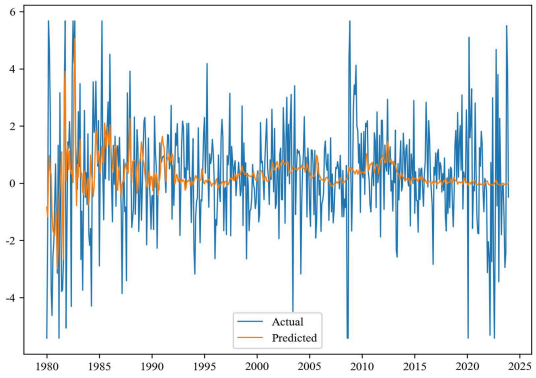
Table A1. All Bonds — Return ($t + 1$ vs t)

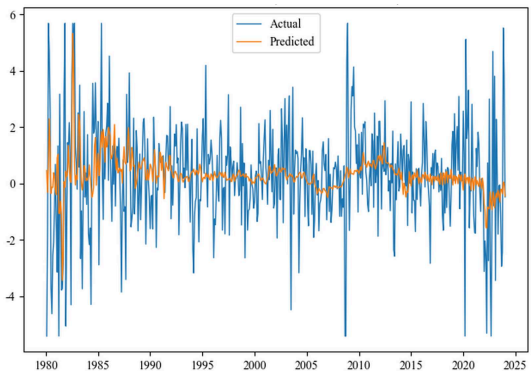
All Bonds — Return (t+1)

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.43	0.00
Industrial Machinery & EV Supply Chain	0.32	0.00
Liability Management: Tender & Exchange Offers	0.14	0.00
Consumer Staples	-0.37	0.01
Registration Qualification	-0.09	0.06
R^2	0.11	

All Bonds — Return (t)

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.40	0.00
Biopharma, Gene Therapy & Specialty Pharmaceuticals	-0.16	0.00
Industrial Machinery & EV Supply Chain	0.33	0.00
Distributable Cash Flow & Unitholder Payouts	0.50	0.00
Russia-Ukraine War & European Geopoliti	-0.31	0.04
R^2	0.11	





Notes: See Table 1. The specification here uses ALL-bond return mapping under the main t + 1 specification and a same-month benchmark.

Table A2: Same-Month Narrative Diagnostics

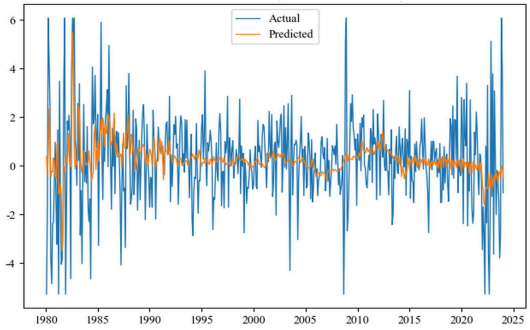
Because the business-cycle diagnostics in Bybee et al. (2024) are contemporaneous, Appendix Table A2 reports same-month versions of the return and spread diagnostics as supporting benchmarks.

Panel A relates month-t static-LDA topic attention to month-t excess returns. Panel B relates month-t topic attention to month-t yield spreads. These specifications assess whether the narrative mappings remain broadly similar under a same-month timing convention.

Table A2. Same-Month Narrative Diagnostics — Panel A Returns

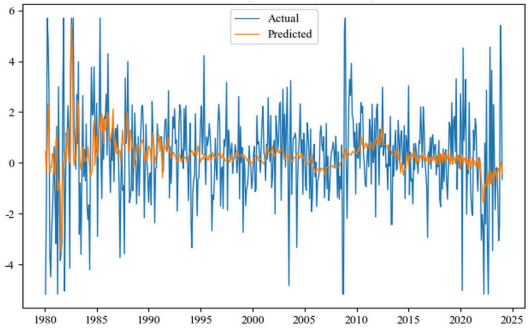
AAA & AA — Return

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.41	0.00
Biopharma, Gene Therapy & Specialty Pharmaceuticals	-0.15	0.00
Industrial Machinery & EV Supply Chain	0.35	0.00
Distributable Cash Flow & Unitholder Payouts	0.53	0.00
Russia-Ukraine War & European Geopoliti	-0.30	0.05
R^2	0.12	



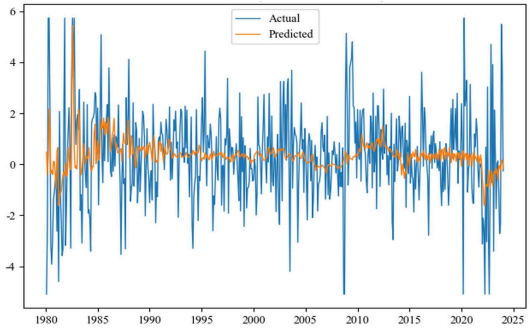
A — Return

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.41	0.00
Biopharma, Gene Therapy & Specialty Pharmaceuticals	-0.16	0.00
Industrial Machinery & EV Supply Chain	0.33	0.00
Distributable Cash Flow & Unitholder Payouts	0.52	0.00
Russia-Ukraine War & European Geopoliti	-0.30	0.04
R^2	0.11	



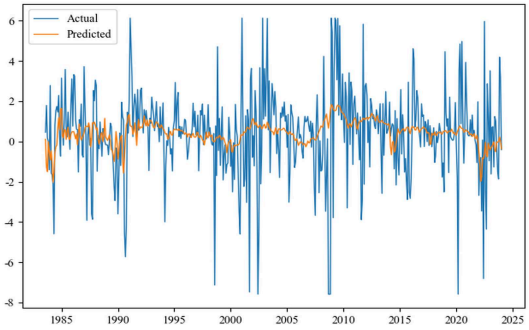
BBB — Return

Topic	Coeff.	p-val.
European Fiscal & Political Crisis / Sovereign Debt	0.43	0.00
Distributable Cash Flow & Unitholder Payouts	0.56	0.00
Industrial Machinery & EV Supply Chain	0.24	0.00
Economic Downturn	-0.09	0.01
Russia-Ukraine War & European Geopoliti	-0.36	0.02
R^2	0.09	



High Yield — Return

Topic	Coeff.	p-val.
Natural Disasters & Severe Incidents	-0.29	0.00
Russia-Ukraine War & European Geopoliti	-0.42	0.00
Telecom Networks & Mobile Devices	-0.17	0.03
Economic Downturn	-0.30	0.03
Operational Efficiency & Strategic Initiatives	-0.11	0.16
R^2	0.07	

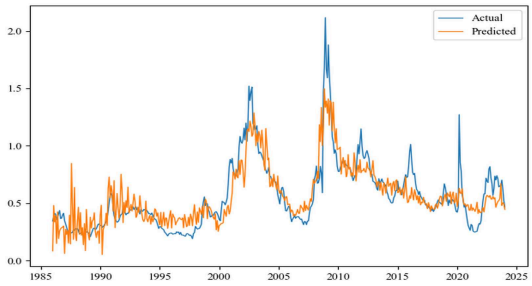


Notes: See Table 1. The specification here uses month t excess returns in a full-sample predictive validation exercise.

Table A2. Same-Month Narrative Diagnostics — Panel B Yield Spreads

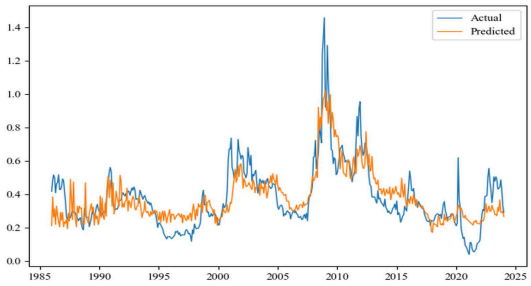
All Bonds — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	0.26	0.00
Goodwill Impairment	0.12	0.00
Severe Weather & Regional Macro	0.13	0.00
Potash	-0.07	0.00
Yield Curve	0.06	0.04
R^2	0.72	



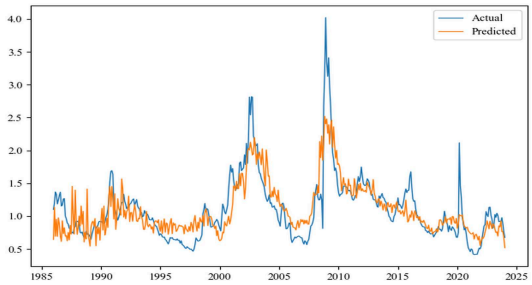
A — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	0.14	0.00
Yield Curve	0.06	0.00
Goodwill Impairment	0.04	0.00
Cannabis	-0.10	0.00
Eurozone Policy Stress	0.04	0.05
R^2	0.64	



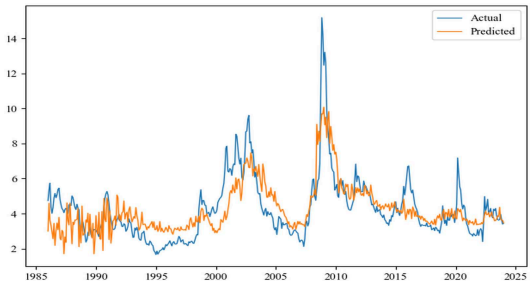
BBB — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	0.43	0.00
Goodwill Impairment	0.14	0.00
Severe Weather & Regional Macro	0.19	0.00
Operational Efficiency & Strategic Initiatives	-0.11	0.00
Yield Curve	0.10	0.09
R^2	0.64	



High Yield — Yield Spread

Topic	Coeff.	p-val.
Economic Downturn	1.31	0.00
Goodwill Impairment	0.45	0.00
Severe Weather & Regional Macro	0.54	0.00
Environmental Protection	-0.83	0.01
Yield Curve	0.59	0.01
R^2	0.58	



Notes: See Table 2. The specification here uses month t yield spreads in a full-sample predictive validation exercise.

Table A3. Predicting Bond Yield Spreads (LDA Topics + FRED-MD)

ALL Spread				HY Spread			
Topic	Coeff.	t	p-val.	Topic	Coeff.	t	p-val.
Panel A: Full (LDA Topics + FRED-MD)							
Economic Downturn	0.263	7.59	3.1e-14	VIXCLSx	0.00112	7.28	3.4e-13
EBIT & Goodwill	0.126	5.18	2.2e-07	Economic Downturn	0.977	4.60	4.2e-06
Goodwill Impairment	0.0936	5.04	4.7e-07	BUSINVx	-0.530	-3.88	1.1e-04
VIXCLSx	0.000102	3.53	4.2e-04	Goodwill Impairment	0.343	3.53	4.1e-04
Yield Curve	0.0337	1.52	0.128	Yield Curve	0.300	2.16	0.031
Panel B: Benchmark (FRED-MD only)							
VIXCLSx	0.000151	4.14	3.4e-05	VIXCLSx	0.00123	6.31	2.8e-10
BUSINVx	-0.116	-3.54	4.0e-04	BUSINVx	-0.702	-3.80	1.4e-04
T10YFFM	0.000641	2.90	0.004	CES0600000007	-0.00568	-2.75	0.006
HOUSTNE	-0.00178	-2.49	0.013	PERMITMW	-0.0103	-2.00	0.045
HOUSTW	-0.000597	-0.73	0.467	HOUSTW	-0.00257	-0.68	0.498
Full R^2	0.69			Full R^2	0.73		
Benchmark R^2	0.51			Benchmark R^2	0.65		

Notes: Five-regressor active-set regressions after LASSO selection. Predictors at month t forecast next-month spreads ($t + 1$). Panel A uses LDA topics + FRED-MD; Panel B uses FRED-MD only. This is an in-sample predictive diagnostic with fixed full-sample LDA topics. The out-of-sample tests in Sections 5 and 6 replace the fixed full-sample LDA measures with train-only oLDA measures. EBIT here denotes Earnings Before Interest and Taxes.

Table A4. OOS Return Forecast Performance — Contemporaneous As-of-t Narrative Implementation

Series	Train Win (m)	OOS Months	OOS Start	OOS End	OOS R^2 vs RollMean	OOS R^2 vs ExpMean	Utility Gain (%, ann.) vs RollMean	Utility Gain (%, ann.) vs ExpMean	Hit Rate
HY	60	354	1994-06	2023-11	0.017	-0.008	1.4	2.1	61.6%
ALL	60	354	1994-06	2023-11	-0.03	-0.046	0.2	0.2	59.9%
HY	84	330	1996-06	2023-11	-0.039	-0.049	2.6	2.4	58.5%
ALL	84	330	1996-06	2023-11	-0.015	-0.024	0.1	0.2	58.8%
HY	120	294	1999-06	2023-11	-0.027	-0.036	1.3	1.8	62.2%
ALL	120	294	1999-06	2023-11	-0.011	-0.01	0.1	0.3	58.2%

Notes: This table reports the contemporaneous as-of-t counterpart to the main OOS return-forecast specification in Table 6. Train-only narrative innovations constructed through month t are used directly to forecast the next forecast month's excess return, without the additional one-month lag used in the main specification. All other forecasting, shrinkage, and evaluation choices are unchanged.

Table A5. Robustness to Detrending Choices (HY)

Strategy	DetrendMode	Sample	Months	Sharpe	$\text{Alpha}_{\text{ann}}$ vs BH	Alpha t vs BH	$\text{Alpha}_{\text{ann}}$ vs RFonly	Alpha t vs RFonly
RiskFilter-only (HY)	none	FULL	325	0.783	0.019	3.13	—	—
Text Overlay + RiskFilter (HY)	none	FULL	325	0.823	0.022	3.62	0.003	1.79
Text Overlay only (HY)	none	FULL	325	0.588	0.009	2.35	—	—
RiskFilter-only (HY)	rolling	FULL	325	0.783	0.019	3.13	—	—
Text Overlay + RiskFilter (HY)	rolling	FULL	325	0.768	0.021	3.16	0.001	0.30
Text Overlay only (HY)	rolling	FULL	325	0.548	0.006	1.62	—	—
RiskFilter-only (HY)	expanding	FULL	325	0.783	0.019	3.13	—	—
Text Overlay + RiskFilter (HY)	expanding	FULL	325	0.795	0.022	3.52	0.002	1.09
Text Overlay only (HY)	expanding	FULL	325	0.560	0.007	1.87	—	—
RiskFilter-only (HY) post-2008	none	POST2008	191	0.832	0.019	2.16	—	—
Text Overlay + RiskFilter (HY) post-2008	none	POST2008	191	0.885	0.024	2.78	0.005	1.72
Text Overlay only (HY) post-2008	none	POST2008	191	0.731	0.015	3.21	—	—
RiskFilter-only (HY) post-2008	rolling	POST2008	191	0.832	0.019	2.16	—	—
Text Overlay + RiskFilter (HY) post-2008	rolling	POST2008	191	0.848	0.025	2.51	0.003	1.01
Text Overlay only (HY) post-2008	rolling	POST2008	191	0.692	0.013	2.28	—	—
RiskFilter-only (HY) post-2008	expanding	POST2008	191	0.832	0.019	2.16	—	—
Text Overlay + RiskFilter (HY) post-2008	expanding	POST2008	191	0.855	0.024	2.67	0.003	1.08
Text Overlay only (HY) post-2008	expanding	POST2008	191	0.704	0.013	2.83	—	—

Notes: This table reports robustness of the HY timing results to alternative train-only detrending choices. DetrendMode indicates how the train-only narrative-based forecast signal is detrended using only information in the training set: "none" applies no detrending; "rolling" detrends within each rolling training window; "expanding" detrends using an expanding-window scheme. Sample = FULL uses the full OOS window; POST2008 restricts to the post-2008 subsample. Sharpe is annualized. $\text{Alpha}_{\text{ann}}$ is 12 times the monthly OLS intercept from a regression of the strategy return on the indicated benchmark return, and Alpha t is the corresponding t-statistic for that intercept. The benchmarks are Buy-and-hold (BH) and RiskFilter-only (RFonly). The risk filter uses spreads available from 1996 onward.

Table A6. Horse Race — Text vs Macro (Panel A: Post-2008)

Series	Metric	Buy-and-hold	Text	FRED-MD	GW	EPU	Macro-core	Text+Macro
HY	Sharpe ratio	0.581	0.729	0.641	0.743	0.462	0.704	0.629
	Expected return	5.837	6.276	5.916	7.462	5.021	6.907	6.098
	t-stat	2.005	2.352	2.106	2.179	1.647	2.267	2.045
	Information ratio	—	0.141	0.025	0.482	-0.228	0.332	0.092
	Alpha (ann., %)	—	1.492	0.820	1.957	-0.939	1.502	0.698
	Alpha t-stat	—	3.209	1.228	1.353	-1.068	1.427	1.500
	Max loss	-33.375	-26.589	-25.419	-25.431	-36.712	-26.109	-31.327
	Skew	-1.100	-0.649	-1.129	-0.122	-3.099	-0.903	-1.113
	Months	191	191	191	191	191	191	191
ALL	Sharpe ratio	0.514	0.519	0.527	0.451	0.517	0.513	0.502
	Expected return	3.541	3.271	3.896	3.184	3.605	3.720	3.396
	t-stat	1.802	1.772	1.781	1.572	1.981	1.699	1.611
	Information ratio	—	-0.140	0.175	-0.193	0.031	0.079	-0.066
	Alpha (ann., %)	—	0.159	0.241	-0.315	0.176	0.182	0.097
	Alpha t-stat	—	0.556	0.598	-0.498	0.410	0.410	0.210
	Max loss	-22.434	-20.983	-21.179	-22.179	-18.411	-21.209	-23.328
	Skew	-0.593	-0.645	-0.719	-1.262	-1.054	-0.967	-1.270
	Months	191	191	191	191	191	191	191

Notes: This table reports out-of-sample timing performance in a horse race between text and macro information sets. Each column corresponds to a strategy built from a different information set: Buy-and-hold, Text, FRED-MD, GW, EPU, Macro-core, and Text+Macro. Rows report annualized Sharpe ratio, expected return, the Newey–West t-statistic of the mean, information ratio and alpha relative to Buy-and-hold (where applicable), maximum drawdown ("Max loss"), skewness, and the number of months. Panel A reports results for HY and the broad ALL aggregate; Panel B (Table A6 continued) reports results separately by investment-grade rating. Both panels restrict the sample to the post-2008 subsample. All quantities are computed strictly out-of-sample using information available at time t.

Table A6. Horse Race — Text vs Macro by Rating (Panel B: Post-2008)

Series	Metric	Buy-and-hold	Text	FRED-MD	GW	EPU	Macro-core	Text+Macro
BBB	Sharpe ratio	0.543	0.662	0.604	0.435	0.427	0.591	0.589
	Expected return	3.991	4.313	4.831	3.603	3.473	4.698	4.300
	t-stat	1.837	2.293	1.909	1.529	1.660	1.864	1.896
	Information ratio	—	0.127	0.317	-0.172	-0.208	0.271	0.122
	Alpha (ann., %)	—	0.990	0.736	-0.738	-0.730	0.626	0.574
	Alpha t-stat	—	1.748	1.481	-0.885	-0.824	1.182	0.966
	Max loss	-21.443	-14.399	-20.943	-22.585	-21.181	-20.892	-19.151
	Skew	-1.241	-0.865	-1.009	-2.375	-3.024	-0.978	-1.518
	Months	191	191	191	191	191	191	191
A	Sharpe ratio	0.421	0.442	0.424	0.389	0.462	0.429	0.422
	Expected return	2.905	2.901	3.062	2.655	3.120	3.084	3.054
	t-stat	1.510	1.515	1.563	1.353	1.800	1.497	1.394
	Information ratio	—	-0.002	0.070	-0.127	0.103	0.075	0.064
	Alpha (ann., %)	—	0.253	0.169	-0.099	0.410	0.231	0.170
	Alpha t-stat	—	0.947	0.272	-0.154	1.036	0.414	0.332
	Max loss	-23.430	-23.177	-20.640	-22.557	-19.799	-21.441	-24.140
	Skew	-0.610	-0.773	-1.791	-1.229	-0.096	-1.956	-1.723
	Months	191	191	191	191	191	191	191
AAA/AA	Sharpe ratio	0.365	0.350	0.429	0.368	0.450	0.444	0.402
	Expected return	2.362	2.152	2.709	2.535	2.745	2.711	2.573
	t-stat	1.340	1.118	1.707	1.385	1.731	1.581	1.425
	Information ratio	—	-0.112	0.168	0.095	0.197	0.163	0.101
	Alpha (ann., %)	—	0.001	0.522	0.112	0.620	0.606	0.358
	Alpha t-stat	—	0.002	0.952	0.332	1.664	1.039	0.765
	Max loss	-26.581	-29.017	-21.724	-27.485	-21.779	-21.389	-24.844
	Skew	-0.150	-0.538	-0.381	0.516	0.522	-0.756	-0.610
	Months	191	191	191	191	191	191	191

Notes: This table reports out-of-sample timing performance in a horse race between text and macro information sets, separately by credit rating: AAA/AA, A, and BBB. Each column corresponds to a strategy built from a different information set: Buy-and-hold, Text, FRED-MD, GW, EPU, Macro-core, and Text+Macro. Rows report annualized Sharpe ratio, expected return, the Newey–West t-statistic of the mean, information ratio and alpha relative to Buy-and-hold (where applicable), maximum drawdown ("Max loss"), skewness, and the number of months. Panel B decomposes the investment-grade segment into AAA/AA, A, and BBB. The sample is restricted to the post-2008 subsample. All quantities are computed strictly out-of-sample using information available at time t.

Table A7. Figure 2 Post-2008 Topic Alignment Stability Screen

Topic ID	Figure 2 Topic	*N*	Mean Diag. Cosine	Mean Gap	Share Row-Best	Mean Top20 Overlap	Score	Soft Pass
T0041	Yield Curve & Fed Policy	16	0.634	0.012	0.562	0.441	1/4	No
T0039	Bank Loan Losses & Nonperforming Assets	16	0.853	-0.001	0.812	0.725	3/4	Yes
T0112	Greater China & Hong Kong	16	0.956	0.551	1.000	0.653	4/4	Yes
T0126	Convertible Debentures & Notes	16	0.783	0.224	0.812	0.594	4/4	Yes
T0147	Bond Tender Offers & Consents	16	0.877	0.334	0.938	0.756	4/4	Yes
T0081	Indenture Covenants & Note Redemption	16	0.741	0.147	0.750	0.588	4/4	Yes
T0161	Trading halts / market microstructure stress	16	0.575	0.190	0.562	0.472	2/4	No
T0082	Securities Class Actions & Litigation	16	0.921	0.371	0.875	0.834	4/4	Yes
T0141	Food Retail & Inflation	16	0.618	0.060	0.562	0.425	2/4	No

Notes: Table A7 reports a post-2008 stability screen for the Figure 2 narrative-channel topics under the refit-and-align OOS pipeline. Each row corresponds to one topic displayed in Figure 2, using the same topic IDs and display titles. *N* is the number of post-2008 refit pairs used in the screen. Mean Diag. Cosine is the average cosine similarity between the topic's aligned term-distribution vector in adjacent refit snapshots. Mean Gap equals the average diagonal cosine minus the best off-diagonal cosine within the same row, so larger positive values indicate that the aligned same-ID topic is more distinct from nearby alternatives. Share Row-Best is the proportion of post-2008 refits in which the diagonal match is the row-wise best match. Mean Top20 Overlap is the average overlap rate of the topic's top-20 terms across adjacent refits. Score counts how many of the four soft-screen thresholds are satisfied: mean diagonal cosine ≥ 0.65, mean gap ≥ 0.05, share row-best ≥ 0.60, and mean top-20 overlap ≥ 0.30. Soft Pass indicates that a topic satisfies at least three of the four thresholds. The table is intended as a compact tractability screen for the Figure 2 narratives rather than as a claim that every displayed topic is perfectly unchanged at every refit.

Table A8. Transaction-Cost Sensitivity of HY Timing Result

Sample	Months	Gross alpha vs RFonly	Gross t vs RFonly	Incremental ann. turnover	Break-even bp (one-way)	Net alpha @0bps	Net alpha @10bps	Net alpha @20bps	Net alpha @40bps
FULL	325	0.003	1.79	0.780	41.04	0.003	0.002	0.002	0.000
POST2008	191	0.005	1.72	1.066	42.21	0.005	0.003	0.002	0.000

Notes: This table examines whether the incremental alpha of Text Overlay + RiskFilter relative to RiskFilter-only survives plausible transaction costs in the HY market. Gross alpha is the annualized OLS intercept from a regression of the Text Overlay + RiskFilter strategy return on the RiskFilter-only (RFonly) benchmark return, and Gross t is the corresponding t-statistic for that intercept. Incremental annual turnover captures the additional trading generated by the text overlay. The net alpha columns report the remaining annualized alpha after applying one-way transaction costs of 0, 10, 20, and 40 basis points. Break-even bp (one-way) denotes the break-even transaction cost, in basis points, at which the gross incremental alpha is fully eliminated.

Table A9. Risk-Filter Threshold Grid Robustness (HY)

Sample	Spread z threshold	Spread change threshold (bp)	Baseline	RF-only Sharpe	Text+RF Sharpe	Alpha_ann vs BH	Alpha t vs BH	Alpha_ann vs RFonly	Alpha t vs RFonly
FULL	1.50	25		0.774	0.821	0.021	3.54	0.003	1.98
FULL	1.50	50		0.773	0.813	0.022	3.58	0.003	1.77
FULL	1.50	75		0.746	0.788	0.021	3.45	0.003	1.72
FULL	2.00	25		0.781	0.827	0.022	3.54	0.003	2.00
FULL	2.00	50	*	0.783	0.823	0.022	3.62	0.003	1.79
FULL	2.00	75		0.755	0.796	0.022	3.48	0.003	1.73
FULL	2.50	25		0.769	0.812	0.021	3.32	0.003	1.87
FULL	2.50	50		0.774	0.811	0.022	3.45	0.003	1.68
FULL	2.50	75		0.748	0.786	0.021	3.32	0.003	1.64
POST2008	1.50	25		0.811	0.877	0.022	2.57	0.005	2.02
POST2008	1.50	50		0.811	0.865	0.023	2.68	0.004	1.71
POST2008	1.50	75		0.808	0.862	0.024	2.77	0.005	1.63
POST2008	2.00	25		0.833	0.898	0.024	2.64	0.005	2.03
POST2008	2.00	50	*	0.832	0.885	0.024	2.78	0.005	1.72
POST2008	2.00	75		0.829	0.882	0.025	2.88	0.005	1.64
POST2008	2.50	25		0.846	0.908	0.024	2.67	0.005	2.00
POST2008	2.50	50		0.845	0.896	0.025	2.81	0.004	1.69
POST2008	2.50	75		0.842	0.893	0.026	2.92	0.005	1.61

Notes: This table reports the Sharpe ratio and annualized alpha of the RF-only and Text + RF strategies over a coarse 3×3 grid of spread-level and spread-change thresholds around the baseline HY risk-filter specification, RF denotes the RiskFilter. The baseline parameter pair is marked with *. Alpha_ann vs BH and Alpha_ann vs RF-only denote annualized alpha relative to Buy-and-Hold (BH) and RiskFilter-only (RFonly), respectively; Alpha t reports the corresponding t-statistic for each alpha estimate. The qualitative conclusion is stable across the displayed region: the risk filter improves HY timing, and the text overlay provides positive incremental alpha on top of the filter at every grid point. Alpha relative to Buy-and-Hold remains positive throughout; the corresponding t-statistics exceed 3 in the full sample and are approximately 2.6–2.9 in the post-2008 sample.

Table A10. Sensitivity to the Number of oLDA Topics: HY Timing

						Relative to Buy-and-hold		Relative to RiskFilter-only	
K	Sample	Months	Buy-and-hold Sharpe	RiskFilter-only Sharpe	Text + RiskFilter Sharpe	α_{ann}	$t(\alpha)$	α_{ann}	$t(\alpha)$
150	FULL	325	0.495	0.783	0.801	0.022	3.22	0.002	1.05
180	**FULL**	**325**	**0.495**	**0.783**	**0.823**	**0.022**	**3.62**	**0.003**	**1.79**
210	FULL	325	0.495	0.783	0.797	0.021	3.41	0.002	1.12
150	POST2008	191	0.583	0.832	0.828	0.022	2.13	0.001	0.41
180	**POST2008**	**191**	**0.583**	**0.832**	**0.885**	**0.024**	**2.78**	**0.005**	**1.72**
210	POST2008	191	0.583	0.832	0.853	0.023	2.44	0.002	1.08

Notes: This table repeats the high-yield timing exercise using alternative numbers of oLDA topics. The baseline specification uses $K = 180$, shown in bold. All forecasting, text-overlay, and risk-filter settings are kept fixed. Buy-and-hold holds HY each month. RiskFilter-only applies the spread-based risk filter without the text forecast. Text + RiskFilter combines the text overlay with the risk filter. α_{ann} is annualized alpha relative to the benchmark shown in the column heading, so 0.022 corresponds to about 2.2 percent per year. FULL uses the full out-of-sample window, and POST2008 restricts the sample to months from 2008 onward.