Global News Networks and Market Returns *

Gustavo Freire†, Ali Moin‡, Alberto Quaini§, Amar Soebhag¶

June 2026

Abstract

We study the relation between local and cross-border sentiment and market returns across 14 major economies. We do so by exploiting the Global Database of Events, Language, and Tone (GDELT), a massive open-access database that catalogs real-time sentiment as reported by the world's news media. We extract sentiment across 260 themes classified by source and target country and use it to construct market-timing strategies. These strategies deliver strong out-of-sample Sharpe ratios and reveal a map of informational market integration across the globe. Preserving source, target, and theme structure and allowing for nonlinearities are fundamental for the economic gains we document.

Keywords: News Sentiment, Equity Markets, Market Integration, Big Data, Machine Learning

JEL classification: G12, G14, G15, C55

*We thank Guido Baltussen, Martijn Boons, Jefferson Duarte, Pasquale Della Corte, Oliver Hellum, Elise Gourier, Jens Jackwerth, Robin Lumsdaine, Liang Peng, Hashem Pesaran, Mirco Rubin, Olivier Scaillet, Paul Schneider, Simon Trimborn, Andrea Vedolin, Bart van Vliet, Tjeerd de Vries (discussant), Christina Dan Wang, Marvin N. Wright, and seminar participants at QFFE 2025 (Marseille), IAAE 2025 (Turin), EFA 2025 (Paris), FoFI 2026 (Lancaster), FMA Europe 2026 (Braga), Erasmus University Rotterdam, Tilburg University, and Robeco for helpful comments. The views expressed in this paper are solely those of the authors and are not necessarily shared by their institutional affiliations.

†gustavo.freire@novasbe.pt, Nova School of Business and Economics, Universidade NOVA de Lisboa, Campus de Carcavelos, 2775-405 Carcavelos, Portugal.

‡moin@ese.eur.nl. Department of Econometrics, Erasmus University Rotterdam.

§quaini@ese.eur.nl. Department of Econometrics, Erasmus University Rotterdam.

¶soebhag@ese.eur.nl (corresponding author). Robeco Quantitative Investing and Department of Finance, Erasmus University Rotterdam.

I. Introduction

Measuring sentiment in the economy is an important yet complex task. Public and private institutions invest substantial resources in developing sentiment measures that are closely monitored by market participants and used in financial decision-making.[1] In fact, investor sentiment has been shown to have strong implications for asset pricing. Theoretically, sentiment generates asset-price fluctuations through the trading of noise or liquidity traders (De Long, Shleifer, Summers, and Waldmann, 1990; Campbell, Grossman, and Wang, 1993). Empirically, numerous studies have shown that investor sentiment predicts both stock and market returns; for a recent literature review, see Zhou (2018).

A recent trend in the literature is the use of textual analysis to extract news-based sentiment measures (Shapiro, Sudhof, and Wilson, 2022). However, most studies focus on English-language news and summarize sentiment at a highly aggregated level, either across all articles or within a small set of topics. Moreover, they generally ignore the directional structure of news coverage by abstracting from who generates information and who receives it. Consequently, existing measures mask economically meaningful sentiment heterogeneity and provide limited insight into how information flows across countries and becomes reflected in asset prices.

We offer a novel approach to measuring sentiment that addresses these shortcomings and enables us to study information transmission in international financial markets. Our key innovation is to organize news along three dimensions: the source country that produces the news, the topic being discussed, and the target country to which the news refers. This structure yields directed, topic-specific sentiment measures that capture both domestic and cross-border information flows.

[1] Examples include the University of Michigan's Survey of Consumers and Consumer Sentiment Index, the AAII Investor Sentiment Survey, as well as sentiment analysis tools provided by RavenPack and Bloomberg.

For example, we can quantify sentiment in Japanese media coverage of the U.S. stock market, allowing us to identify where sentiment originates, what narrative it reflects, and which market it concerns. By preserving this directional structure, we can isolate the predictive content of individual information sources and map sentiment linkages across countries. To the best of our knowledge, this paper is the first to construct such a high-dimensional network of theme-specific sentiment connections within and across national markets.

To implement this framework, we draw on the most extensive international news corpus used in the finance literature to date: the *Global Database of Events, Language, and Tone* (GDELT).[2] GDELT covers more than 150,000 multilingual news outlets and offers two features central to our analysis. First, each article is classified into a set of thematic categories, allowing us to measure sentiment at a granular topic level. Second, GDELT identifies both the source of a news item and the entities or locations discussed, enabling us to trace the directional flow of information across countries. Considering more than 520 million articles from 14 major economies between 2015 and 2023, we construct a high-dimensional network of topic-specific sentiment linkages that captures both domestic and cross-border information transmission.

Using this dataset, we examine whether local and global sentiment can be converted into economically valuable market-timing signals for equity-index futures. We first test whether local news coverage, defined as news produced by domestic media about the domestic market, contains information about future local equity returns. We then extend the analysis to a global sentiment network that incorporates both domestic and cross-border information flows. This framework

[2]GDELT was introduced in 2013 and is supported by Google. It is a free, open-source initiative that monitors more than 150,000 news outlets in print, broadcast, and online formats from nearly every country, covering both English and translingual (non-English articles) articles in over 100 languages, and codifies them into a continuously updated record of individuals, organizations, locations, emotions, and themes refreshed every 15 minutes. Its creators describe it as the largest, highest-resolution open database of human society in existence. Further details are available at https://www.gdeltproject.org/.

allows sentiment from anywhere in the network to inform local return forecasts and enables us to identify which countries contribute most to market-timing performance, as well as whether distinct regional or global clusters of informational influence emerge.

We estimate a Random Forest model on an expanding window of historical sentiment data and use the resulting predictions to implement a real-time market-timing strategy. Following the evaluation framework of Kelly, Malamud, and Zhou (2024), we assess the economic value of these signals through portfolio performance measures, rather than focusing on statistical forecast accuracy. The results indicate that sentiment linkages contain economically valuable predictive information to improve market-timing performance across global equity markets.

Specifically, we find that market-timing strategies based on local sentiment significantly outperform buy-and-hold benchmarks across our 14-country sample. Local sentiment delivers higher Sharpe ratios in 13 of 14 markets, with average net-of-costs Sharpe ratios exceeding those of buy-and-hold by 82.5%. These gains are not driven by market exposure and are accompanied by substantially lower drawdowns. Incorporating global sentiment linkages further improves performance. Global sentiment strategies generate economically and statistically significant gains relative to both buy-and-hold and local-sentiment strategies, even after transaction costs. Consistent with this result, spanning regressions reveal positive and economically meaningful alphas of roughly 10%–30% per annum across most markets, indicating that global sentiment contains information beyond local news. An intermediate specification based solely on U.S.-sourced news performs strongly across markets, but global sentiment generally delivers superior forecasting performance.

We also identify the narratives the Random Forest relies on most. Using standard gain-based variable importance measures based on how much each theme's splits reduce squared-error loss, we find that at any point in time a small subset of themes carries the bulk of this importance, and that their identity changes markedly over time. During the Covid-19 pandemic, health-related narratives become dominant predictors of future returns, while themes related to debt, financial markets, and economic conditions gain importance during periods of elevated inflation and monetary policy uncertainty. Similarly, geopolitical narratives rise in prominence around major international events such as the Russian invasion of Ukraine. These findings suggest that news sentiment predicts returns by capturing evolving narratives about economic activity, discount rates, risk premia, and macroeconomic uncertainty. Consistent with theories of time-varying investor attention and narrative-based asset pricing (Bybee, Kelly, and Su, 2023; Bybee, Kelly, Manela, and Xiu, 2024), the results indicate that the information most relevant for asset prices changes over time. Additional model comparisons show that aggregating sentiment across themes weakens performance and that linear specifications do not fully capture the signal, indicating that the economic value of news depends on preserving source, target, and theme structure while allowing for nonlinear effects.

We then summarize these cross-country predictive relationships in a global news-based sentiment network (Figure 1). Our network prominently features the United States as the central informational hub, exerting predictive influence globally, and China as a crucial secondary node. European countries such as the United Kingdom and Germany emerge as significant regional hubs, illustrating strong intra-European informational integration. By contrast, markets such as Australia and Sweden exhibit relatively limited cross-country predictive influence, underscoring their peripheral role within the global sentiment network.

Overall, these results provide compelling evidence of informational market integration by showing that international equity markets are sensitive to shared sentiment and news flows. The identified global sentiment network structure highlights how cross-border informational connections help shape equity market-timing performance across countries.

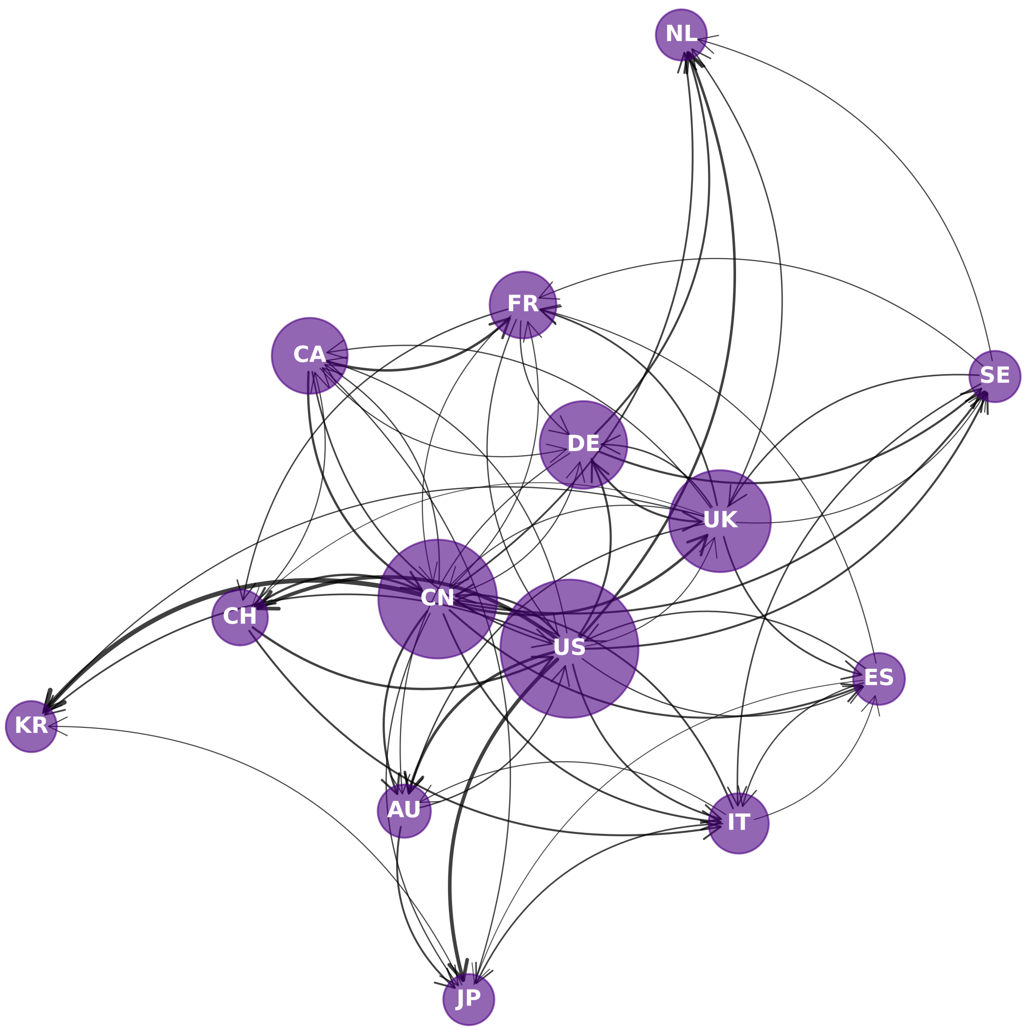
Lastly, our results remain robust across a battery of robustness checks. Our findings hold when using various alternative sentiment lexicons, indicating that the choice of dictionary does not materially influence the predictive power of news sentiment. Moreover, controlling for past returns suggests that our findings are not simply driven by a short-term momentum effect, since adding lagged returns sometimes even weakens the predictive ability of local news sentiment. Finally, to address potential translation biases, we compare native-language Hedonometer (Dodds, Harris, Kloumann, Bliss, and Danforth, 2011) lexicons with their English-translated counterparts, finding that sentiment-based predictability often strengthens after translation.

Our study contributes to several strands of the literature. First, our findings contribute to the literature on media sentiment and asset prices (see, e.g., Tetlock, 2007; Dougal, Engelberg, García, and Parsons, 2012; García, 2013; Manela and Moreira, 2017; Heston and Sinha, 2017; Calomiris and Mamaysky, 2019; Fraiberger, Lee, Puy, and Ranciere, 2021). For example, Fraiberger, Lee, Puy, and Ranciere (2021) show that international asset prices respond to foreign media coverage, indicating that sentiment crosses borders via global news channels. Our contribution is to construct granular sentiment measures that explicitly retain the source-target and topical structure of news coverage. Furthermore, we focus on market timing strategies rather than using regression-based approaches. More importantly, we are the first to model cross-country sentiment linkages as a large-scale network of information flows.

Second, our study contributes to the recent literature on narrative-based asset pricing. This literature applies textual analysis to capture media exposure towards particular themes (see, e.g., Mai and Pukthuanthong, 2021; Bhargava, Lou, Ozik, Sadka, and Whitmore, 2023; Bybee, Kelly, Manela, and Xiu, 2024; Bybee, Kelly, and Su, 2023; Lee, Lou, Ozik, and Sadka, 2025; Adämmer and Schüssler, 2020). These studies demonstrate that narratives contain predictive information for future

Figure 1: Cross-country news-based predictability network

Directed network visualization of cross-country predictive influence. Nodes represent countries, and node size encodes each country's overall importance, computed by aggregating gain-based Random Forest variable importances as both source and target in news themes. Incoming arrows indicate the top five foreign contributions to forecasting a country's equity returns, with arrow thickness proportional to the strength of that influence. The layout is force-directed to spatially reflect the connectivity structure. The out-of-sample forecasting period runs from February 2017 to December 2023. See Section IV.D for details.



asset prices, but they are typically limited to the United States. As in the previous studies, we classify news into hundreds of unique themes. We contribute by constructing attention-weighted sentiment measures that combine narrative exposure with narrative direction. Our attention-weighting of themes also relates to Da, Engelberg, and Gao (2015), who show that an aggregate measure of investor attention built from search data predicts returns. Studying attention and sentiment jointly is important because they capture distinct primitives of narrative exposure: attention measures the intensity of investor focus, whereas sentiment measures the direction of the narrative (Cookson, Lu, Mullins, and Niessner, 2024). When these two dimensions have different and potentially offsetting return implications, analyses that focus on only one can misattribute predictability to the wrong channel or mask heterogeneous effects.

Third, we contribute to the market integration literature. Numerous studies have documented varying degrees of equity market integration (see, e.g., De Jong and De Roon, 2005; Karolyi and Stulz, 2003; Bekaert, Hodrick, and Zhang, 2009; Lewis, 2011; Carrieri, Chaieb, and Errunza, 2013; Hau, 2011). Our spanning tests reveal that global sentiment measures provide economically distinct and incremental market timing ability beyond information contained in local sentiment measures. In addition, the network representation underscores the global prominence of the U.S. and existence of regional informational hubs. These findings strongly support the existence of informational market integration through sensitivity to shared foreign news and sentiment, consistent with prior studies emphasizing the increasing prominence of common information in shaping international return dynamics (Pukthuanthong and Roll, 2009; Rapach, Strauss, and Zhou, 2013; Hellum, Pedersen, and Rønn-Nielsen, 2025; Choi, Jiang, and Zhang, 2025), and reinforcing the idea that equity markets operate as parts of an integrated informational network. Our results also resonate with the trade network centrality literature (Richmond, 2019), which highlights the importance of trading linkages, whereas we highlight news linkages.

The paper is structured as follows. Section II describes the GDELT database and the construction of our set of sentiment indices. Section III presents our machine learning methodology to construct sentiment-based return forecasts. Section IV contains our main empirical analysis. Section V considers a battery of robustness tests. Section VI concludes.

II. Data

This section introduces our data and provides descriptive statistics. Section II.A presents the GDELT database, while Section II.B describes the sample construction procedure. Section II.C details the methodology used to construct our high-dimensional set of sentiment measures. Finally, Section II.D briefly describes equity returns data.

A. The GDELT Database

To develop our high-dimensional set of sentiment measures across themes and countries, we exploit an open-source, global news database: the Global Database of Events, Language, and Tone (GDELT), which was introduced in 2013 and is supported by Google. GDELT provides two main datasets: the GDELT Event Database and the Global Knowledge Graph Database (GKG). GKG files are versioned and refreshed every 15 minutes. We use the GKG to leverage two important features of this database.

First, GKG collects global news from over 150,000 news outlets offering both English and translingual articles (covering over 100 different languages). This is important as it allows us to significantly broaden the coverage of global news in different languages, a feature that is often overlooked in the prior literature. In Figure 2, we plot the proportion of English-based and translingual news for each month in our sample. We observe that translingual news outlets produce a significant portion (55% on average) of the total global news, which underlines the importance of including translingual news in sentiment measures.

Second, GKG labels each publicly available global news article with individuals, organizations, locations, emotions, and themes. For instance, GKG tracks the “Election” (“Theme”) in the U.S. (“Target”), which can be covered by U.S. media or other international outlets (“Source”). This feature is crucial as it allows us to construct sentiment measures between and within countries for a wide range of themes. Furthermore, it enables us to construct sentiment measures of one source country about another target country for a specific theme, which we dub *sentiment linkages*. To our knowledge, modeling sentiment linkages is novel to the literature.

Although GDELT contains a wealth of global news information, it does not include the full text of news articles and instead provides historical links. Some of these links may be inaccessible for several reasons like broken URLs, paywalls, or changes in website structures. GDELT overcomes these challenges by offering a comprehensive set of sentiment metrics. This approach eliminates the need for individual researchers to deploy computationally intensive NLP techniques or large language models, making it easier to construct high-dimensional sentiment measures efficiently.

B. Sample Construction

We collect GKG data from February 2015 through December 2023 for the following 14 countries: United States (US), China (CN), Germany (DE), Japan (JP), United Kingdom (UK), France (FR), Italy (IT), Canada (CA), Australia (AU), South Korea (KR), Spain (ES), Netherlands (NL), Switzerland (CH), and Sweden (SE). These countries collectively account for approximately 72% of global GDP (World Bank, 2023) and provide us with a representative sample of major economies.

GKG classifies articles into themes based on the content of the article. Our study focuses on 260 major themes provided by GKG since its inception.[3] These major themes exclude entertainment and lifestyle-related content (such as gossip, leisure, and sports), and focus mostly on socioeconomic news events. Articles in GDELT may be tagged with multiple themes. Furthermore, GDELT updates its news database every 15 minutes and all news items are recorded in Coordinated Universal Time (UTC). In total, our sample covers over 520 million web articles.

To provide insights into the composition of our sample, we show the article distribution by source country in Figure 3(A). We find that the U.S. plays a major role in the production of news articles, with a relative share of 49%, followed by China (10.9%), U.K. (6.9%) and Germany (6.0%). We also show the article distribution by theme in Figure 3(B).[4] There is no single dominating theme in our sample, stressing the diversity in themes that occur in global news. For instance, the "Leader" theme is tagged in 5.9% of all articles, "General Health" in 4.8%, and "Election" in 1.6%. Articles with the "Stock Market" theme occur 1.4% of the time.

These illustrative numbers are over the full sample. However, themes are time-varying and dynamic. For example, some themes are short-lived, trending, and eventually fade away. In Figure 4, we show the top 20 themes per month in our sample, with the U.S. as the target country. We find that some topics structurally obtain high levels of attention, such as science ("Science") and crime ("Soc_GeneralCrime"). Other topics fluctuate over time and are recurring, such as the general discussion of holidays ("Gen_Holiday"). In addition, some topics are short-lived and

[3]A brief description of each theme can be found at data.gdeltproject.org. This link has also been archived via the Wayback Machine on 2025-01-21 at web.archive.org.

[4]For ease of exposition, we only show the top 30 themes over the full sample.

capture significant media attention, such as impeachment ("Impeachment", heavily covered from September 2019 till February 2020), shortage ("Shortage", January 2022), "Sanctions" (March 2022), and the Covid pandemic ("Health Pandemic").

As mentioned in Section II.A, GDELT provides entity linkages within articles. This feature allows us to construct cross-country sentiment linkages. In Figure 5, we show the number of articles that each source country (s) publishes about each target country (k) as a fraction of the total number of news articles published by s.[5] We observe three notable patterns. First, most of the produced news by a source country is about itself, i.e., national news. For example, 85% of the news articles published by the U.S. are targeted at the U.S. itself. Second, non-U.S. countries have a considerable share of news about the U.S., highlighting the importance of the U.S. in terms of attention. For example, 26% of the Japanese news in our sample mentions the United States. Third, countries that are geographically close tend to have stronger connections in terms of news targeting. For instance, 20% and 27% of the South Korean news is about China and Japan, respectively. Germany mentions France and the U.K. in 10% and 8% of all German articles, which is relatively higher compared to other countries for Germany. We explicitly take these features into consideration when we measure sentiment, as explained in the next section.

C. Sentiment Measures

To measure sentiment for each article in our sample, we rely on measures provided by GDELT. Specifically, we use the SentiWordNet 3.0 (SWN) lexicon-based framework (Baccianella, Esuli, and Sebastiani, 2010). SWN is able to take the context of words in sentences and texts into account. Each "synset" (synonym set) of WordNet groups together words that are synonyms and share the

[5] Note that the rows do not sum to 1.00 since an article published by country s can have multiple target countries.

same meaning in different contexts. Synsets are the fundamental building blocks of WordNet and serve as a way to organize the lexicon based on semantic relationships rather than just lexical ones. Each synset is associated with three sentiment scores (summing to one): positivity, negativity, and objectivity. We utilize both the positive and negative scores of an article. These scores are derived based on the context in which words appear, making SentiWordNet effective for nuanced sentiment analysis across diverse themes and contexts.

A well-known alternative is the financial lexicon-based approach of Loughran and McDonald (2011) (LM). We choose not to use this approach because our dataset contains 260 themes, the majority of which contain non-financial texts, where the LM method would be less effective in accurately capturing sentiment. However, in Section V.A, we conduct a robustness test across different lexicons for the national news of the United States and obtain similar results.

We now describe the construction of our sentiment measures. Let t index days, p index source countries, q index target countries, and $\lambda = 1, \ldots, 260$ index themes. For each such index, we define $S_{t,p,q,\lambda} \in \mathbb{R}$ as the average daily sentiment score, and $a_{t,p,q,\lambda} \in \mathbb{N}_0$ as the total number of articles on day t regarding theme λ from source country p about target country q. To avoid cumbersome triple indices (p, q, λ) when describing sentiment linkages, we introduce a single index ℓ to denote each unique combination of $(p_\ell, q_\ell, \lambda_\ell)$. Formally, let $\mathcal{L} = \{(p, q, \lambda)\}$ be the set of all such triplets and let $\ell = (p_\ell, q_\ell, \lambda_\ell) \in \mathcal{L}$ refer to a specific link. For instance, rather than writing $S_{t,p,q,\lambda}$ for the average daily sentiment on day t from source country p about target country q, we write $S_{t,\ell}$.

We aggregate article-level sentiment scores into a daily measure, the attention-weighted sentiment (AWS) change, for each source-target-theme combination as follows.[6] First, we compute the one-day change in sentiment for each linkage ℓ. Second, we compute the total daily news

[6]To avoid staleness and trends in levels, we consider changes in sentiment.

percentage for each theme λ within each source-target pair (p, q). Lastly, we multiply the average sentiment change by its attention-based weights:

$$\text{AWS}_{t,\ell} \coloneqq \underbrace{\frac{a_{t,\ell}}{\sum_{\lambda'=1}^{260} a_{t,p_\ell,q_\ell,\lambda'}}}_{\text{attention weight}} \times \underbrace{\left(S_{t,\ell} - S_{t-1,\ell}\right)}_{\text{one-day change in sentiment}} .$$

Recent studies stress the role of narratives in asset pricing (e.g., Bybee, Kelly, Manela, and Xiu, 2024). Our approach takes both the direction and intensity of sentiment into account, thereby also accounting for time-varying narratives. Given that our sample has 14 countries and 260 themes, each with a positive and negative sentiment dimension, we obtain a high-dimensional set of 101,920 AWS sentiment linkages.

Figure 6 illustrates the net AWS measure (i.e., positive minus negative AWS sentiment) for the "Stock Market" (panel A) and "Health Pandemic" (panel B) themes for the United States. The upper plot shows that the sign of sentiment, i.e., the extent to which news are perceived as positive or negative about the stock market, varies over time. While the net AWS for the stock market was mostly positive until 2019, it turned significantly negative in 2020 due to the Covid-19 pandemic, eventually bouncing back to positive values towards the end of the sample in 2023. On the other hand, articles on pandemics received very little attention before 2020 due to the absence of major international outbreaks. In 2020, articles on the Covid-19 pandemic significantly increased and played a major role in the negative narrative, reverting back over time as the pandemic subsided. This example illustrates how our AWS measure accounts for rising and trending themes in the news.

D. Market Return Data

Our daily market data consists of an excess return series $\{r_1, \ldots, r_T\} \subset \mathbb{R}$ constructed from futures contract prices for each of the 14 equity indexes from February 2015 through December 2023, obtained from Bloomberg. For each country, we select the most liquid, accessible equity-index futures contract. For China, we use Hang Seng Index (HSI) futures rather than a mainland (A-share) index because the HSI is a highly liquid, freely accessible, market-cap-weighted index that avoids the foreign-access restrictions, daily price limits, and capital controls affecting onshore A-share futures over our sample. Because the HSI is dominated by mainland-Chinese constituents, it serves as our proxy for the Chinese equity market. Table A1, in the appendix, details the sample universe, market closing times, timezones, index name, and the annualized Sharpe ratio of each instrument.

Returns are close-to-close on daily settlement prices: (r_{t+1}), the realized return on day $t+1$, runs from the day t settlement to the day $t+1$ settlement. As previously noted, GDELT records news events in Coordinated Universal Time (UTC), while market return data are timestamped in local time. To prevent any forward-looking bias in our forecasting process, we convert each news event's UTC timestamp to the local timezone of the corresponding market. If a news item occurs after the local market's closing time, it is reassigned to the following trading day. The forecast $\hat{r}_{t+1}$, and hence the position w_{t+1}, is therefore formed at the day t settlement using only news published up to that settlement. This adjustment ensures that our analysis only incorporates information that would have been available to market participants at the relevant time, and that news arriving after the day t settlement enters only later forecasts, never the position held over the day $t+1$ return.

III. Machine Learning Setup

Our decomposition of aggregate media sentiment by country, theme, and source-target structure allows us to study this time variation across a wide range of equity markets. Given the high dimensionality of our set of sentiment predictors, we rely on machine learning techniques. This section describes and motivates the machine learning setup to predict international equity excess returns using our high-dimensional set of sentiment series.

A. The Prediction Problem and Random Forests

To frame our setting as a general prediction problem, let the collection of AWS sentiment linkages observed by the end of day t be summarized by a state vector $x_t \in \mathbb{R}^k$, where $k = PQ\Lambda D$ is the product of the numbers of source countries (P), target countries (Q), themes (Λ), and sentiment polarity components (D). We assume x_t is measurable at time t and that investors use this information to form expectations about equity returns. The general return prediction problem involves identifying a functional form $g(\cdot)$ for $\mathbb{E}(r_{t+1} \mid x_t)$ that maps the predictor set x_t such that maximum out-of-sample predictive power relative to r_{t+1} is obtained. We describe this mapping between the excess market return and the set of sentiment measures as

$$r_{t+1} = \mathbb{E}(r_{t+1} \mid x_t) + \epsilon_{t+1} = g(x_t) + \epsilon_{t+1},$$

where ϵ_{t+1} is the residual with $\mathbb{E}(\epsilon_{t+1} \mid x_t) = 0$.

There are a variety of methods to estimate $g(\cdot)$ and produce out-of-sample predictions using this estimate. Traditionally, the simple linear regression model has been the mainstay of return predictability. Despite its popularity, linear regression is infeasible in our setting, given the high dimensionality of our set of predictors. A popular alternative strategy, especially in the context of a large set of predictors, is to employ regularized linear regression methods such as Ridge regression (Hoerl and Kennard, 1970), Lasso regression (Tibshirani, 1996) or Elastic Net (Zou and Hastie, 2005). While these regularized linear methods are natural benchmarks in high dimensions, they remain linear in the predictors and therefore impose a restrictive functional form on return predictability. In our setting, the mapping from theme-level attention and sentiment to subsequent returns may be nonlinear and state-dependent. For example, effects may exhibit thresholds, saturation, or interaction patterns across narratives, so a purely linear specification risks missing economically relevant predictability.[7]

For this reason, and because our primary objective is out-of-sample forecasting performance rather than structural interpretation of a small parametric model, we estimate $g(\cdot)$ using a Random Forest (Breiman, 2001). Random Forests are nonparametric ensemble methods that approximate complex functions by averaging predictions across many decision trees, which allows them to capture nonlinearities and high-order interactions in a data-adaptive way while remaining robust to overfitting through aggregation. This flexibility is particularly valuable in our application: the sentiment measures summarize evolving narratives and investor attention across a large set of themes, and there is little reason to expect their predictive content for equity returns to be well described by a stable linear index. We view the Random Forest as a pragmatic choice that improves predictive accuracy at the cost of model simplicity; we therefore complement it with additional analyses that recover the time-varying importance of themes and provide economic interpretation

[7] One can enrich linear models by manually adding nonlinear transformations and interactions, but this approach requires ex ante specification of the relevant nonlinearities and becomes quickly infeasible in large predictor sets.

of the drivers of predictability.

To fix notation, we first describe the construction and estimation of a single regression tree, which is the basic building block of the forest. A decision tree is grown by recursively partitioning the predictor space into increasingly homogeneous subsets. Each subset corresponds to a node of the tree (a terminal node, or leaf, is a subset on which the tree issues a constant prediction). In regression settings, each split is chosen to maximize the reduction in squared-error loss, equivalently to reduce the dispersion of returns within the child nodes created by the split.

Let $D_m \subseteq \{(r_{t+1}, x_t)\}_{t=0}^{T-1}$ denote the set of observations reaching node m. A candidate split is defined by a feature index $j \in \{1, \ldots, k\}$ and a threshold $\tau \in \mathbb{R}$, which partitions D_m into

$$D_m^L(j,\tau) = \{(r_{t+1}, x_t) \in D_m : x_{t,j} \leq \tau\}, \qquad D_m^R(j,\tau) = \{(r_{t+1}, x_t) \in D_m : x_{t,j} > \tau\}.$$

For any subset $D \subseteq D_m$, let

$$\bar{r}_D = \frac{1}{|D|} \sum_{(r_{t+1}, x_t) \in D} r_{t+1}$$

denote the sample mean of the response variable in D. We define the within-node variance as

$$\mathrm{Var}(D) = \frac{1}{|D|} \sum_{(r_{t+1}, x_t) \in D} (r_{t+1} - \bar{r}_D)^2,$$

which is the mean squared deviation of the response from its within-node mean and is proportional to the within-node sum of squared errors.

The optimal split at node m is obtained by minimizing the weighted within-node variance:

$$(j^*, \tau^*) \coloneqq \arg\min_{j,\tau} \left[\frac{|D_m^L(j,\tau)|}{|D_m|} \mathrm{Var}\Big(D_m^L(j,\tau)\Big) + \frac{|D_m^R(j,\tau)|}{|D_m|} \mathrm{Var}\Big(D_m^R(j,\tau)\Big) \right].$$

This splitting procedure is applied recursively, yielding a binary tree structure.

Tree growth is terminated when a predefined stopping criterion is met. Common stopping criteria include: a maximum tree depth is reached; the number of observations in a node falls below a minimum threshold; the reduction in loss achieved by further splitting is smaller than a specified tolerance (e.g., a minimum split gain); or the node variance is zero, implying perfect homogeneity. These criteria control tree complexity and mitigate overfitting.

Once a node satisfies a stopping criterion, it becomes a terminal (leaf) node. In a regression tree, the prediction associated with a leaf node l is the empirical mean of the response variable among observations in that node:

$$\hat{r}_l = \bar{r}_{D_l}.$$

Equivalently, $\hat{r}_l$ is the constant that minimizes the squared-error loss within the leaf:

$$\hat{r}_l = \arg\min_{c \in \mathbb{R}} \sum_{(r_{t+1}, x_t) \in D_l} (r_{t+1} - c)^2.$$

Predictions for new observations are obtained by routing an input x through the learned split rules and returning the value $\hat{r}_l$ of the reached leaf l.

A Random Forest aggregates multiple decision trees trained on randomized versions of the data (Breiman, 2001). Formally, let $\mathcal{D}_1, \ldots, \mathcal{D}_B$ denote B subsamples of fixed size drawn without replacement from the full dataset $\mathcal{D}$. For each subsample $\mathcal{D}_b$, a decision tree $\mathcal{T}_b$ is trained following the procedure above. The Random Forest prediction for a new observation x is obtained by averaging the predictions of the individual trees:

$$\hat{g}(x) = \frac{1}{B} \sum_{b=1}^{B} \mathcal{T}_b(x).$$

In contrast to bootstrap aggregating (Breiman, 1996), we employ subsample aggregating (subagging), whereby each tree is trained on a subsample drawn without replacement from the original dataset (Bühlmann and Yu, 2002). In addition, at each split only a randomly selected subset of predictors is considered, which further decorrelates trees and improves generalization (Breiman, 2001). Subagging is particularly well motivated when predictive relationships are diffuse and the signal-to-noise ratio is low, as is common in economic applications. Shen and Xiu (2025) emphasize that in weak-signal environments, predictive methods can easily overfit: estimation error from fitting flexible models may dominate the small gains from capturing many tiny effects. Averaging many trees trained on different *without-replacement* subsamples reduces this estimation variance by stabilizing the hard split decisions of individual trees (Bühlmann and Yu, 2002). Recent theoretical work further shows that, for a fixed tree size, subagging improves upon a single regression tree and that the gains are larger when trees are highly adaptive (i.e., when variance is otherwise large) (Revelas, Boldea, and Werker, 2024). Practically, subagging also avoids duplicate observations within a tree, a feature of bootstrap samples, which can yield more stable split selection in finite samples (Bühlmann and Yu, 2002).

Alternatively, we could also train Neural Networks to model complex nonlinear relationships in our data. However, this demands meticulous feature engineering and tuning, making them overly sensitive to pre-processing and prone to overfitting.[8]

[8] Grinsztajn, Oyallon, and Varoquaux (2022) argue that tree-based models often outperform Neural Networks on tabular datasets due to their inherent robustness and lower sensitivity to pre-processing.

B. Model Estimation and Evaluation

To investigate the effects of news-based sentiment on asset prices, we predict daily market returns from a given country using information from a high-dimensional set of sentiment measures. We proceed in three stages. First, we fit Random Forests on the initial 18 months of data (from February 2015 through July 2016) to establish our base models. Second, we use the following 6 months (August 2016–January 2017) as a validation sample to select the optimal hyperparameters. We implement a Random Forest with subagging and choose hyperparameters by minimizing validation MSE. Tuned parameters include: (i) the predictor-subset fraction per tree, (ii) the subagging fraction of observations per tree, (iii) the number of trees, (iv) the maximum number of leaves per tree, and (v) the minimum number of observations per leaf. Finally, having locked in those hyperparameters, we retrain the model in an expanding-window fashion, adding 20 trading days at a time (roughly one calendar month), so that each new fit incorporates all data from February 2015 up to the end of the previous month.

To evaluate our machine learning-based predictive performance, we assess economic significance via a market-timing strategy.[9] To this end, we consider the following real-time trading strategy. Let $\widehat{r}_{t+1}$ denote the out-of-sample forecast of the one-day-ahead market return formed at the end of day t. We take a unit long/short position for day $t+1$ given by

$$w_{t+1} \coloneqq \operatorname{sign}(\widehat{r}_{t+1}) = \begin{cases} +1, & \widehat{r}_{t+1} > 0, \\ -1, & \widehat{r}_{t+1} < 0, \end{cases}$$

[9] Although the out-of-sample R^2 (as in Campbell and Thompson, 2008) is commonly used in the literature, it is a purely statistical measure of forecast accuracy and need not correspond to economic gains. A market timer can still generate significant economic profits even when the out-of-sample R^2 is negative (see, e.g., Kelly, Malamud, and Zhou, 2024; Leitch and Tanner, 1991; Cenesizoglu and Timmermann, 2012; Rapach and Zhou, 2013).

and define the corresponding strategy return as

$$r_{t+1}^{\text{strat}} := w_{t+1}\, r_{t+1},$$

where r_{t+1} is the realized market return on day $t+1$. The restriction $w_{t+1} \in \{-1, +1\}$ ensures that strategy returns are not inflated by leverage. Moreover, if the signal does not change sign between two consecutive days, i.e., $w_{t+1} = w_t$, we maintain the existing position and do not trade, which limits turnover as trading only occurs when $w_{t+1} \neq w_t$. For the buy-and-hold benchmark, we simply set the portfolio weight to be constant, i.e., $w_{t+1}^{\text{BH}} \equiv 1$ for all t, so that $r_{t+1}^{\text{BH}} = r_{t+1}$.

We then compute annualized Sharpe ratios (SR), information ratios (IR), maximum drawdowns ($\text{DD}_{\max}$), and (net) alphas for these real-time investment strategies. When reporting Sharpe ratios, we annualize them by multiplying by $\sqrt{252}$. We compute Sharpe-ratio t-statistics using the autocorrelation-robust standard errors of Lo (2002), to adjust the i.i.d. standard error for serial dependence in strategy returns. We measure the information ratio as the annualized mean active return relative to the buy-and-hold benchmark divided by the annualized volatility of active returns, i.e., using $r_t^{\text{strat}} - r_t^{\text{BH}}$. We quantify downside risk by reporting the maximum drawdown $\text{DD}_{\max}$, defined as the largest peak-to-trough decline of the cumulative return process over the out-of-sample period. To obtain the alpha of our trading strategy, we regress the percentage returns of our trading strategy on the buy-and-hold benchmark percentage returns; when reporting alphas, we annualize them by multiplying by 252. Standard errors for the alphas and betas are Newey and West (1987) HAC standard errors with 7 lags. Our target variables are excess returns obtained from futures instruments. These are known to be highly liquid, with low transaction costs. We therefore account for trading costs in our analysis of net alphas by assuming a fixed transaction cost of 2 basis points (bps) per trade direction for equity futures.

IV. Empirical Analysis

In this section, we study market timing strategies based on predictions constructed from local and global sentiment measures. In addition, we shed light on which theme-specific sentiment measures drive the local market timing ability, and on the relative importance of theme granularity and nonlinearities in the predictive model. Lastly, we show that U.S.-centered sentiment carries an outsized influence in domestic and global contexts.

A. Local News and Market Returns

We first examine the predictive ability of local sentiment measures for local equity market returns, where we define local news as coverage produced by a country's media outlets that predominantly concerns that same country. Such coverage constitutes the majority of a nation's news output: Figure 5 shows that between 65% and 87% of articles originating from a given country explicitly reference it. Whereas previous studies have documented such a local effect for the United States (see, e.g., Baker and Wurgler, 2007; García, 2013), we test it across 14 countries, predicting one-day-ahead returns with the Random Forest of Section III using 260 theme-specific sentiment measures per market and evaluating the resulting out-of-sample forecasts through a real-time trading strategy.

Table 1 presents the out-of-sample Sharpe ratios, $CAPM$ alphas and other performance statistics for our local news sentiment strategy. We document that local news sentiment delivers strong market-timing performance across all countries. For example, in the U.S., a local news sentiment strategy achieves an out-of-sample Sharpe ratio of 1.34 (t-statistic = 3.59), more than double the buy-and-hold Sharpe ratio of 0.62. Furthermore, we find sizable CAPM alphas (versus the buy-and-hold benchmark) of 20.41% per annum.

Given that our strategy is rebalanced on a daily basis, transaction costs may substantially diminish the documented profitability. However, even after accounting for transaction costs, the net Sharpe ratio (1.11 for the U.S.) and net alpha (15.98%) remain significantly sizable. In 13 out of the 14 countries in our sample, we find that the net Sharpe ratio is larger than that of holding the underlying instrument. Overall, the average net Sharpe ratio is 82.5% larger than the average benchmark Sharpe ratio. Furthermore, local news sentiment strategies exhibit low betas relative to their underlying markets, ranging from -0.20 (France) to 0.41 (U.S.).[10] Trading on local news sentiment also reduces left-tail risk—measured by maximum drawdown—by approximately one third compared to the average buy-and-hold strategy, indicating improved downside protection.

Taken together, our results suggest that local news about a country itself delivers significant economic value for timing future equity returns. Furthermore, we show that the market-timing value of news sentiment is pervasive and virtually present in all countries in our sample. This is consistent with and extends prior literature showing that news sentiment predicts asset prices. From a theoretical perspective, our results align with theories of investor sentiment.

B. Interpreting Local News Sentiment

So far, our results indicate that local news sentiment about a country's own equity market delivers significant economic value for timing next-day returns. Specifically, our approach employs hundreds of theme-specific sentiment measures within a Random Forest framework. As is often the case with

[10]In unreported results, we also compare our strategies against time series momentum (TSMOM) strategies of Moskowitz, Ooi, and Pedersen (2012) and find that our strategies are not spanned by TSMOM and considerably outperform TSMOM strategies.

machine learning methods, it can be difficult to explain why they work (Nagel, 2021), particularly when using a large set of predictors. In this section, we shed light on which theme-specific sentiment measures drive the observed return predictability.

To measure the importance of theme-specific sentiment, we employ a gain-based approach within our Random Forest framework. This metric reflects the improvement in predictive accuracy attributable to each feature. Specifically, when a feature is used to split a decision tree in the Random Forest, it helps reduce the prediction error. The gain metric quantifies the extent to which including that feature improves the model's accuracy at each split. We aggregate these gains across all trees and splits where the feature is used, resulting in an overall measure of its contribution to the model's predictions.

To facilitate comparisons over time, we normalize the raw importance values for each feature so that the sum of importance scores across all features at any given time equals 100%. This normalization allows for direct comparison of each feature's relative significance over time. Additionally, to enable meaningful comparisons across themes, we aggregate the variable importance of the positive and negative sentiment variants within each news-based theme. By summing the importance scores of both sentiment measures for a given theme, we obtain a unified importance score. This approach highlights which themes exert the most significant influence on equity returns over time.

Figure 7 presents the five most important themes in each quarter for the U.S. equity market. Several patterns emerge. First, predictive importance is highly concentrated. At any point in time, only a small subset of themes accounts for a substantial fraction of the model's forecasting ability. Second, the identity of these themes changes considerably over time. The predictive content of news sentiment therefore appears neither constant nor broadly dispersed across all topics. Instead, return predictability is episodically concentrated in a limited set of narratives that vary with the economic environment.

This time variation aligns closely with major shifts in macroeconomic and geopolitical conditions. During the Covid-19 pandemic, health-related themes become dominant predictors of future returns. Rather than simply reflecting public concern, these themes likely summarize information relevant for expectations about economic activity, corporate cash flows, and aggregate uncertainty. Similarly, themes related to debt, financial markets, and broader economic conditions become more important during periods characterized by elevated concerns about inflation, monetary policy, and financing conditions. Their predictive importance is consistent with the view that news sentiment helps capture variation in expected discount rates and risk premia.

A similar pattern emerges around major geopolitical events. For example, propaganda-related themes become highly important around the Russian invasion of Ukraine. We do not interpret this result as evidence that propaganda itself drives stock returns. Instead, such themes likely proxy for broader geopolitical uncertainty and the arrival of information concerning energy markets, international trade, inflationary pressures, and global growth prospects. In this sense, the importance of a theme should be interpreted as reflecting its ability to summarize economically relevant information rather than its direct causal effect on prices.

Figure A2 provides a complementary perspective on the concentration of predictive importance. Specifically, it plots the share of total Random Forest gain importance attributable to the five most important themes, averaged across countries. Although the top five themes account for a substantial fraction of total importance—typically between 50% and 75%—they do not fully explain the model's predictive performance. A considerable share of importance remains distributed across a broader set of themes, indicating that return predictability is driven by both a small number of dominant narratives and a wider collection of lower-importance signals. This finding motivates our high-dimensional approach, as restricting attention to only the most prominent themes would overlook a meaningful component of the information embedded in the news.

More broadly, our analysis suggests that the predictive content of news sentiment rotates across themes as economic and geopolitical conditions evolve. This finding is consistent with theories emphasizing time-varying investor attention and changing sources of macroeconomic uncertainty. Rather than relying on a fixed set of predictors, the Random Forest dynamically reallocates forecasting weight toward those themes that are most informative in the prevailing environment. The results therefore highlight the value of a flexible, data-driven approach for identifying which narratives are most relevant for asset pricing at different points in time.

C. Global News and Market Returns

The evidence reported in Section IV.A indicates that local news sentiment about the domestic market effectively predicts one-day ahead local equity market returns. Beyond local news about the domestic market, global news coverage represents a potentially powerful source of incremental information, capturing cross-border sentiment linkages, i.e., news from country X reporting on developments in country Y. Specifically, we construct a high-dimensional set of theme-specific sentiment measures that accounts for both incoming and outgoing news from a given country, allowing us to explore the predictive role of global sentiment measures for individual equity markets.

Our local predictor set consists of 260 themes, each separately assessed with positive and negative sentiment scores, totaling 520 local measures per market. To incorporate global informational linkages, we expand our predictor space to include sentiment flows from each market to the other 13 countries and vice versa, yielding a high-dimensional dataset of 14,040 sentiment metrics per country. We employ this extensive predictor set within our Random Forest model to generate one-day-ahead return forecasts for each national equity market. These out-of-sample forecasts form the basis of our real-time trading strategies.

Table 2 summarizes the performance statistics of these global sentiment-based strategies. The empirical results clearly demonstrate that global sentiment measures provide economically substantial and statistically robust market-timing performance. The strategies consistently yield Sharpe ratios that significantly outperform the buy-and-hold benchmarks across all markets, even after accounting for realistic transaction costs. Importantly, strategy alphas remain economically meaningful post-costs, underscoring the persistent profitability derived from global sentiment information. Moreover, the estimated market betas are generally modest (predominantly below 0.2), indicating that strategy returns reflect distinct informational content embedded in global sentiment flows rather than simple market exposure. The strategies also feature lower maximum drawdowns relative to passive market positions, highlighting their resilience to tail events.

To assess whether the local news sentiment captures the incremental informational content embedded in global sentiment, we conduct spanning regressions. Specifically, for each market, we regress the returns from the global sentiment-based strategy on the returns from the local sentiment-based strategy. A statistically significant and economically large intercept (alpha) from these regressions indicates that global sentiment delivers orthogonal predictive content not encapsulated by local signals alone. Consequently, a significant spanning alpha suggests potential improvements in risk-adjusted portfolio performance when investors integrate global sentiment signals with existing local strategies.

Figure 8 presents the results of these spanning tests, depicting annualized gross spanning alphas and their associated t-statistics. Remarkably, we find significant and economically large alphas across all markets, except Canada, with annualized magnitudes ranging from roughly 10% to 30%. These findings strongly suggest that international sentiment linkages convey incremental, economically valuable information beyond local sentiment, underscoring their potential for enhancing portfolio performance.

Collectively, these results provide compelling evidence of market integration from an informational perspective. The observed significance of global sentiment measures in timing market returns across nearly all countries illustrates that equity markets worldwide are sensitive to common information flows and sentiment. Such informational integration aligns with the theoretical predictions and empirical findings documented in the broader literature on international financial market linkages, further reinforcing the pivotal role of shared sentiment and information in driving global asset price dynamics.

D. Global News Sentiment Network

Thus far, our analysis does not explicitly identify which countries drive the return predictability observed across international markets. To shed light on these directional influences, we construct a news-based predictability network employing a force-directed visualization framework (Fruchterman and Reingold, 1991). Specifically, we estimate Random Forest models separately for each country's equity returns using the comprehensive global predictor set and extract gain-based variable importance metrics. These metrics quantify the predictive power of each sentiment linkage, defined by source and target countries, for forecasting returns. To isolate cross-country sentiment effects, we remove local sentiment measures (within-country predictions) and aggregate the resulting importance scores by country pairs. Subsequently, we standardize these scores within each predictive model, ensuring comparability across countries.

Figure 1 and Figure A1 present our resulting global sentiment network visualization. Several noteworthy patterns emerge from this figure. First, the United States occupies a central and dominant position within the network, exerting substantial predictive influence on virtually all other countries, consistent with its documented role as a leading driver of global financial markets (Rapach, Strauss, and Zhou, 2013). China also emerges as a key country, ranking second in overall importance, suggesting significant bilateral informational flows between the U.S. and China influencing global market predictability.

Second, our visualization reveals pronounced regional clustering, especially within Europe, where the United Kingdom and Germany serve as primary hubs, closely connected with France, Italy, Spain, and Switzerland. This European cluster highlights strong regional informational integration. Conversely, certain markets, including Australia and Sweden, exert a limited cross-border predictive influence.

Third, we identify a distinct Asian cluster featuring China, South Korea, and Japan. While China significantly interacts with the U.S. and other key global players, Japan and South Korea exhibit relatively moderate predictive linkages primarily involving the U.S. Lastly, countries such as Australia and Canada display fewer and weaker connections within the global sentiment network, indicating their peripheral role relative to the U.S. and the major European economies.

Taken together, these results illustrate substantial heterogeneity in cross-country predictive influences, underscoring both the central role of major global players and significant regional integration patterns in international news-driven equity return predictability.

E. Theme Granularity and Nonlinearities

Our baseline specification exploits heterogeneous theme-level news sentiment within a nonlinear Random Forest. Two design choices distinguish this approach: allowing for nonlinear interactions across themes, and preserving the granular theme-level structure of the sentiment data rather than collapsing it into an aggregate index. We now show that both are essential to the economic value we document. To isolate the contribution of each, we compare the baseline local and global Random Forest specification with alternative models based on aggregated sentiment measures and elastic-net regularization, which respectively remove theme-level heterogeneity and impose linearity.

Specifically, we consider two dimensions of variation. First, we aggregate sentiment across themes to construct a single local and global sentiment index ('Agg'), thereby removing cross-theme heterogeneity. Second, we estimate Elastic Net (ELN) models that impose linear regularization while retaining the same information set. We evaluate local and global versions of the aggregated specification, but estimate the Elastic Net for the local case only, since the global predictor set is too high-dimensional for a stable linear fit. We report out-of-sample annualized Sharpe ratios.

Figure 9 shows that thematic aggregation substantially reduces predictive performance. Strategies based on aggregate sentiment consistently generate lower Sharpe ratios than models that exploit heterogeneous theme-level information. Elastic Net specifications improve upon the aggregated benchmarks, suggesting that regularization helps extract predictive signals from the news data. However, ELN models remain dominated by the RF specifications in most countries. The superior performance of the RF models implies that the predictive content of news sentiment is not fully captured by linear combinations of sentiment measures and instead reflects nonlinear interactions across themes.

Overall, these results show that both design choices are essential to the economic value of news sentiment. First, preserving the granularity of our 260 themes is critical: collapsing sentiment into an aggregate index substantially degrades performance. The theme-level heterogeneity our framework retains therefore carries genuine market-timing information that aggregate or single-topic measures discard. Second, accounting for nonlinearities matters: the Random Forest dominates the Elastic Net specifications in most markets, indicating that the signal is not well captured by a linear combination of sentiment measures and instead reflects nonlinear interactions across themes.

F. U.S.-Sourced News and Market Returns

Sections IV.A and IV.C establish that both domestic (local) sentiment and cross-border (global) sentiment flows contain predictive information for next-day equity market returns. Given the central role of U.S. media in global information production and dissemination, a natural intermediate question is whether sentiment extracted specifically from U.S.-sourced coverage about foreign countries already contains substantial predictive content for those countries' equity returns.

To answer this question, we restrict the *source* of news to U.S. media outlets and, for each market, construct a high-dimensional set of theme-specific sentiment measures from U.S.-originating articles that discuss the market country. We then estimate the same Random Forest forecasting model as in Section III and implement the corresponding trading strategy. Table 3 reports the resulting out-of-sample performance statistics across the 14 target markets.

Table 3 shows that U.S.-sourced sentiment delivers strong and economically meaningful market-timing performance for virtually all international equity markets. In every country, the gross Sharpe ratio of the U.S.-news strategy exceeds the buy-and-hold benchmark. This dominance largely survives transaction costs. Net Sharpe ratios remain above the benchmark in all countries but France, where the net ratio (0.54) is marginally below buy-and-hold (0.57). Several markets show strong risk-adjusted performance. The Netherlands achieves the highest Sharpe ratios in the table (SR_{gross} = 1.67, SR_{net} = 1.42), followed by Sweden and Spain (SR_{net} = 1.13 and 1.11, respectively), while China also exhibits a high net Sharpe ratio of 1.08. In parallel, U.S.-news strategies generate sizable CAPM alphas: gross alphas range from roughly 13% to 29% per annum across countries and remain economically large after costs (e.g., China 23.75%, Spain 19.54%, the Netherlands 20.56%).

U.S.-sourced sentiment strategies also meaningfully reduce downside risk relative to passive equity exposure. Maximum drawdowns are uniformly less severe than the buy-and-hold benchmark across markets, with especially large improvements in countries such as Canada and Switzerland. Hence, U.S.-based coverage appears to provide not only return predictability but also meaningful downside protection.

Comparing Tables 1 and 3 further highlights heterogeneity in which source of news is most informative for a given market. China, Japan, the U.K., Canada, Korea, Spain, the Netherlands, and Sweden all improve when switching from their own local media coverage to U.S.-sourced coverage, suggesting that U.S. outlets may either (i) aggregate and filter country-specific information efficiently, or (ii) emphasize globally salient developments that are most relevant for pricing. Comparing Tables 2 and 3, China, Canada, Australia, Spain, the Netherlands, and Sweden also improve when moving from global sentiment measures to U.S.-sourced sentiment, implying that for a nontrivial subset of markets, U.S. news may be a particularly informative single node in the broader cross-border information network. Nevertheless, the global strategies still dominate in most countries, consistent with the view that incorporating the full set of bilateral sentiment linkages improves the signal-to-noise ratio beyond what any single source country can provide.

Finally, the U.S. itself provides a useful benchmark for interpreting these patterns. By construction, U.S.-sourced news performance in Table 3 coincides with U.S. local news performance in Table 1. However, Table 2 shows that U.S. market-timing performance weakens when we expand from U.S. local sentiment to the full global predictor set. This decline is consistent with a signal-to-noise interpretation: when the U.S. local signal is already strong, adding foreign-sourced coverage may dilute predictive content rather than enrich it.

Overall, Table 3 establishes that U.S.-sourced news contains substantial economic value for timing international equity returns and often improves upon purely local sentiment signals. At the same time, the fact that global sentiment remains superior for most markets motivates our broader network-based approach in Section IV.C, where we exploit the entire system of cross-border sentiment flows rather than relying on any single information hub.

V. Robustness

We examine the robustness of results from Section IV by verifying whether our main results persist under various alternative specifications and potential sources of bias. We consider three checks: (i) alternative sentiment lexicons; (ii) controlling for lagged returns; and (iii) potential translation biases in non-English news sentiment.

A. Alternative Sentiment Measures

Our primary results rely on sentiment measures derived from the SentiWordNet 3.0 (SWN) lexicon-based framework. To assess the robustness of our findings to the choice of sentiment lexicon, we replicate the U.S. local effect Random Forest trading strategy using four alternative lexicons: (i) the financial lexicon of Loughran and McDonald (2011) (LM), (ii) the Subjectivity Lexicon (Wilson, Wiebe, and Hoffmann, 2005), (iii) ML-SENTICON (Cruz, Troyano, Pontes, and Ortega, 2014), and (iv) Lexicoder (Young and Soroka, 2012). The results, summarized in Table 4, indicate that all four lexicons yield broadly similar performance relative to our SWN-based strategy, consistently outperforming the buy-and-hold Sharpe ratio of S&P 500 futures (0.62).

Among the alternative lexicons, the LM lexicon produced the lowest Sharpe ratio (1.15). A possible explanation is that the LM lexicon is specifically designed for financial text and may not be well-suited for analyzing general news articles. Since a significant portion of the news themes in our dataset extends beyond strictly financial topics, the LM lexicon may fail to capture sentiment nuances effectively.

Overall, these results suggest that our findings are not sensitive to the choice of sentiment lexicon. One reason for this robustness is that we estimate daily sentiment by averaging across a large number of news articles. This aggregation reduces the influence of individual lexicon-specific biases and minimizes potential measurement errors. Consequently, rather than relying on precise sentiment classifications for each individual article, our approach focuses on capturing broader daily sentiment fluctuations, making it less susceptible to discrepancies between lexicons.

B. Controlling for Past Returns

To further test the robustness of our findings, we extend the analysis from Section IV.A, which examined how a country's local news predicts its market returns, by incorporating the previous day's return as an additional predictor. Table 5 presents the updated results alongside those from

Table 1, allowing for direct comparison.

The findings reveal that in China, Germany, the United Kingdom, Canada, Australia, and Switzerland, adding lagged returns reduces predictive performance. In contrast, the remaining countries experience modest but consistent improvements in both statistical and economic significance when past returns are included. This pattern suggests that our primary results are not merely capturing short-term return dynamics. Instead, sentiment measures provide unique predictive value that is not fully explained by past market movements.

C. Translation Bias

Measuring news sentiment often faces challenges when applied to multilingual contexts, as lexicons and sentiment measures may not transfer seamlessly across languages. One common issue is translation bias, where differences in sentiment measurement arise depending on whether texts are analyzed in their native language or translated into another (typically English) before analysis.

To address translation bias, we compare Hedonometer-based sentiment scores computed in the native language (Chinese, German, French, Korean, and Spanish) with those computed on the English-translated text.[11] The Hedonometer is a sentiment analysis tool that assigns words a score along a single "happiness" dimension (Dodds, Harris, Kloumann, Bliss, and Danforth, 2011). We examine to what extent the out-of-sample market timing performance differs when we use sentiment scores (Hedonometer) based on the news in its native language versus translated to English.

In Table 6, we present out-of-sample market timing Sharpe ratios for news sentiment strategies using translated versus native news. Overall, we find that the market timing performance improves

[11] GDELT does not provide SWN-based sentiment scores in the native language of the news, but only to the English-translated version of the news.

in three out of the five languages when the texts are first translated into English before applying the Hedonometer. For example, the Sharpe ratio when using native Spanish news to predict the Spanish equity market equals 0.92 and increases to 0.95 when using Spanish news translated to English. Our findings suggest that translation bias is unlikely to drive our main results: native-language and English-translated sentiment deliver market-timing performance of comparable magnitude, with no systematic degradation from translation. These results align with prior research that explores the impact of translation on sentiment analysis performance.

VI. Conclusions

In this paper, we develop a novel, high-dimensional approach to measuring news-based sentiment, exploiting both thematic and cross-country perspectives in the Global Database of Events, Language, and Tone (GDELT). By systematically capturing how one country's media reports on another across 260 distinct topics, our framework provides granular sentiment linkages that extend well beyond traditional one-dimensional or purely local sentiment measures. We demonstrate that these linkages are economically meaningful for timing equity market returns across 14 major economies, with local sentiment especially influential in North America and Europe, and global (incoming and outgoing) sentiment linkages proving critical in markets such as Japan and China.

We further show that a large share of cross-country predictability can already be recovered from a single information hub: U.S.-originating news. U.S.-sourced sentiment strategies outperform buy-and-hold across all markets and often improve upon local-news strategies, highlighting the global reach of U.S. media in transmitting price-relevant information. At the same time, for most countries, performance increases further when we expand from U.S.-sourced signals to the full global sentiment network, indicating incremental information in broader cross-border news flows.

Our findings are robust to using alternative lexicons, controlling for lagged returns, and accounting for potential translation biases. Taken together, these conclusions highlight the global interplay of economic and geopolitical narratives in shaping market outcomes, emphasizing that additional insights arise when one examines both local sentiment and cross-border news flows. While future work might extend this analysis to firm-level or sector-specific linkages, or to more advanced natural language processing techniques, our findings suggest that high-dimensional sentiment measures offer a rich and underutilized source for understanding global asset pricing.

References

Adämmer, Philipp, and Rainer A. Schüssler, 2020, Forecasting the equity premium: mind the news!, *Review of Finance* 24, 1313–1355.

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani, 2010, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*.

Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor Sentiment in the Stock Market, *Journal of Economic Perspectives* 21, 129–152.

Bekaert, Geert, Robert J. Hodrick, and Xiaoyan Zhang, 2009, International stock return comovements, *Journal of Finance* 64, 2591–2626.

Bhargava, Rajeev, Xiaoxia Lou, Gideon Ozik, Ronnie Sadka, and Travis Whitmore, 2023, Quantifying Narratives and Their Impact on Financial Markets, *Journal of Portfolio Management* 49.

Breiman, Leo, 1996, Bagging predictors, *Machine learning* 24, 123–140.

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

Bühlmann, Peter, and Bin Yu, 2002, Analyzing bagging, *Annals of Statistics* 30, 927–961.

Bybee, Leland, Bryan Kelly, Asaf Manela, and Dacheng Xiu, 2024, Business News and Business Cycles, *Journal of Finance* 79, 3105–3147.

Bybee, Leland, Bryan Kelly, and Yinan Su, 2023, Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text, *Review of Financial Studies* 36, 4759–4787.

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

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108, 905–939.

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

Carrieri, Francesca, Ines Chaieb, and Vihang Errunza, 2013, Do implicit barriers matter for globalization?, *Review of Financial Studies* 26, 1694–1739.

Cenesizoglu, Tolga, and Allan Timmermann, 2012, Do return prediction models add economic value?, *Journal of Banking & Finance* 36, 2974–2987.

Choi, Darwin, Wenxi Jiang, and Chao Zhang, 2025, Alpha Go Everywhere: Machine Learning and International Stock Returns, *Review of Asset Pricing Studies* 15, 288–331.

Cookson, J. Anthony, Runjing Lu, William Mullins, and Marina Niessner, 2024, The social signal, *Journal of Financial Economics* 158, 103870.

Cruz, Fermín L., José A. Troyano, Beatriz Pontes, and F. Javier Ortega, 2014, Building layered, multilingual sentiment lexicons at synset and lemma levels, *Expert Systems with Applications* 41, 5984–5994.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The Sum of All FEARS Investor Sentiment and Asset Prices, *The Review of Financial Studies* 28, 1–32.

De Jong, Frank, and Frans A. De Roon, 2005, Time-varying market integration and expected returns in emerging markets, *Journal of Financial Economics* 78, 583–613.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise Trader Risk in Financial Markets, *Journal of Political Economy* 98, 703–738.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth, 2011, Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter, *PLoS One* 6.

Dougal, Casey, Joseph Engelberg, Diego García, and Christopher A. Parsons, 2012, Journalists and the Stock Market, *Review of Financial Studies* 25, 639–679.

Fraiberger, Samuel P., Do Lee, Damien Puy, and Romain Ranciere, 2021, Media sentiment and international asset prices, *Journal of International Economics* 133, 103526.

Fruchterman, Thomas M. J., and Edward M. Reingold, 1991, Graph drawing by force-directed placement, *Software: Practice and Experience* 21, 1129–1164.

García, Diego, 2013, Sentiment during Recessions, *Journal of Finance* 68, 1267–1300.

Grinsztajn, Léo, Edouard Oyallon, and Gaël Varoquaux, 2022, Why do tree-based models still outperform deep learning on typical tabular data?, in *Advances in Neural Information Processing Systems* vol. 35 pp. 507–520.

Hau, Harald, 2011, Global versus Local Asset Pricing: A New Test of Market Integration, *Review of Financial Studies* 24, 3891–3940.

Hellum, Oliver, Lasse Heje Pedersen, and Anders Rønn-Nielsen, 2025, How Global is Predictability? The Power of Financial Transfer Learning, *Working paper, SSRN* 4620157.

Heston, Steven L., and Nitish Ranjan Sinha, 2017, News vs. Sentiment: Predicting Stock Returns from News Stories, *Financial Analysts Journal* 73, 67–83.

Hoerl, Arthur E, and Robert W Kennard, 1970, Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics* 12, 55–67.

Karolyi, G. Andrew, and René M. Stulz, 2003, Chapter 16 - Are financial assets priced locally or globally?, in *Financial Markets and Asset Pricing* (Elsevier, Amsterdam).

Kelly, Bryan, Semyon Malamud, and Kangying Zhou, 2024, The Virtue of Complexity in Return Prediction, *Journal of Finance* 79, 459–503.

Lee, Hojoon, Xiaoxia Lou, Gideon Ozik, and Ronnie Sadka, 2025, Narrative Momentum, *Working paper, SSRN* 4912496.

Leitch, Gordon, and J. Ernest Tanner, 1991, Economic Forecast Evaluation: Profits Versus the Conventional Error Measures, *American Economic Review* 81, 580–590.

Lewis, Karen K., 2011, Global Asset Pricing, *Annual Review of Financial Economics* 3, 435–466.

Lo, Andrew W., 2002, The Statistics of Sharpe Ratios, *Financial Analysts Journal* 58, 36–52.

Loughran, Tim, and Bill McDonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

Mai, Dat, and Kuntara Pukthuanthong, 2021, Economic narratives and market outcomes: a semi-supervised topic modeling approach, *Working paper, SSRN* 3990324.

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

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Nagel, Stefan, 2021, *Machine Learning in Asset Pricing*. (Princeton University Press New Jersey).

Newey, Whitney K., and Kenneth D. West, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

Pukthuanthong, Kuntara, and Richard Roll, 2009, Global market integration: An alternative measure and its application, *Journal of Financial Economics* 94, 214–232.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2013, International Stock Return Predictability: What Is the Role of the United States?, *Journal of Finance* 68, 1633–1662.

Rapach, David E., and Guofu Zhou, 2013, Chapter 6 - Forecasting Stock Returns, in Graham Elliott, and Allan Timmermann, eds.: *Handbook of Economic Forecasting* (Elsevier, Amsterdam).

Revelas, Christos, Otilia Boldea, and Bas J. M. Werker, 2024, When does Subagging Work?, *Working paper, arXiv* 2404.01832.

Richmond, Robert J., 2019, Trade Network Centrality and Currency Risk Premia, *Journal of Finance* 74, 1315–1361.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson, 2022, Measuring news sentiment, *Journal of Econometrics* 228, 221–243.

Shen, Zhouyu, and Dacheng Xiu, 2025, Can Machines Learn Weak Signals?, *Working paper, NBER* 33421.

Tetlock, Paul C., 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance* 62, 1139–1168.

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

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann, 2005, Recognizing contextual polarity in phrase-level sentiment analysis, in *Proceedings of human language technology conference and conference on empirical methods in natural language processing* pp. 347–354.

World Bank, 2023, World Development Indicators: GDP (current US$), Accessed: 2024-08-06.

Young, Lori, and Stuart Soroka, 2012, Affective News: The Automated Coding of Sentiment in Political Texts, *Political Communication* 29, 205–231.

Zhou, Guofu, 2018, Measuring Investor Sentiment, *Annual Review of Financial Economics* 10, 239–259.

Zou, Hui, and Trevor Hastie, 2005, Regularization and Variable Selection Via the Elastic Net, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67, 301–320.

Figure 2: English Versus Translingual Articles Over Time

The figure shows the proportion of articles in GDELT for each month from February 2015 through December 2023 categorized by English-based articles versus translingual articles.

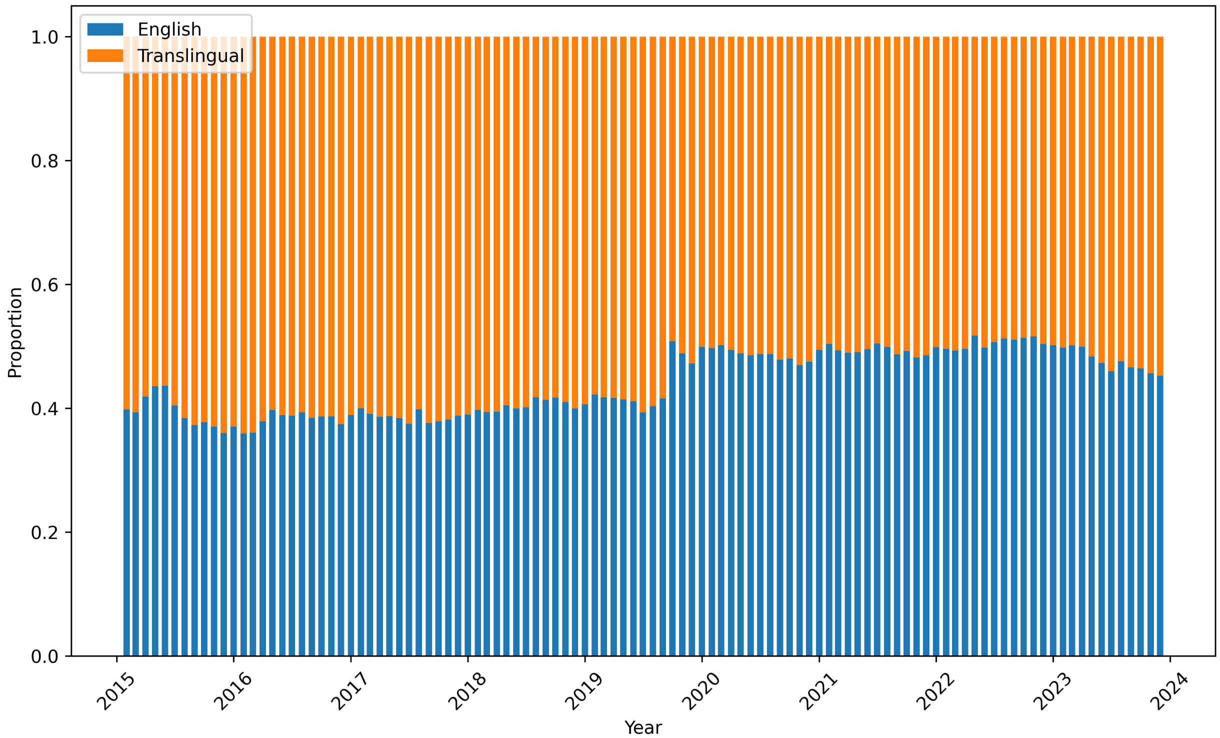
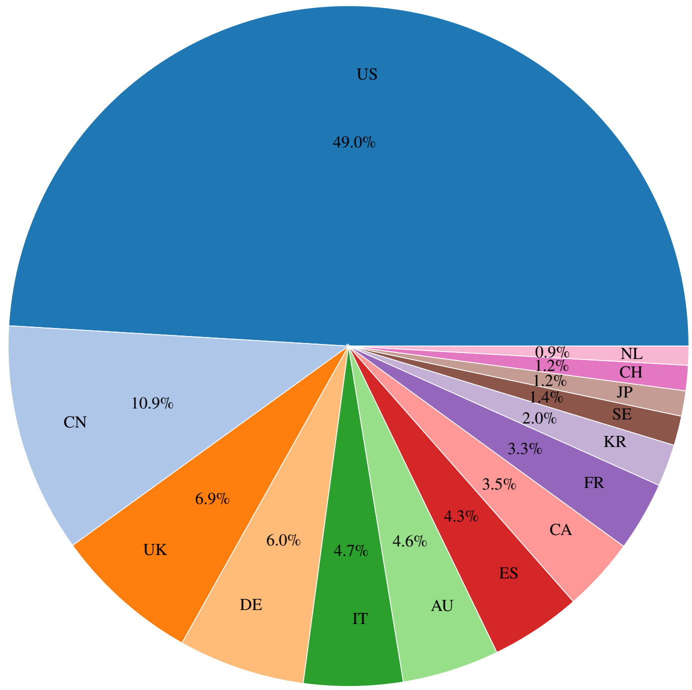


Figure 3: Country and Theme Distribution of Articles

The upper (lower) figure shows the country (theme) distribution in our sample. The sample period runs from February 2015 through December 2023.

Panel A: Country Distribution of Articles



Panel B: Theme Distribution of Articles

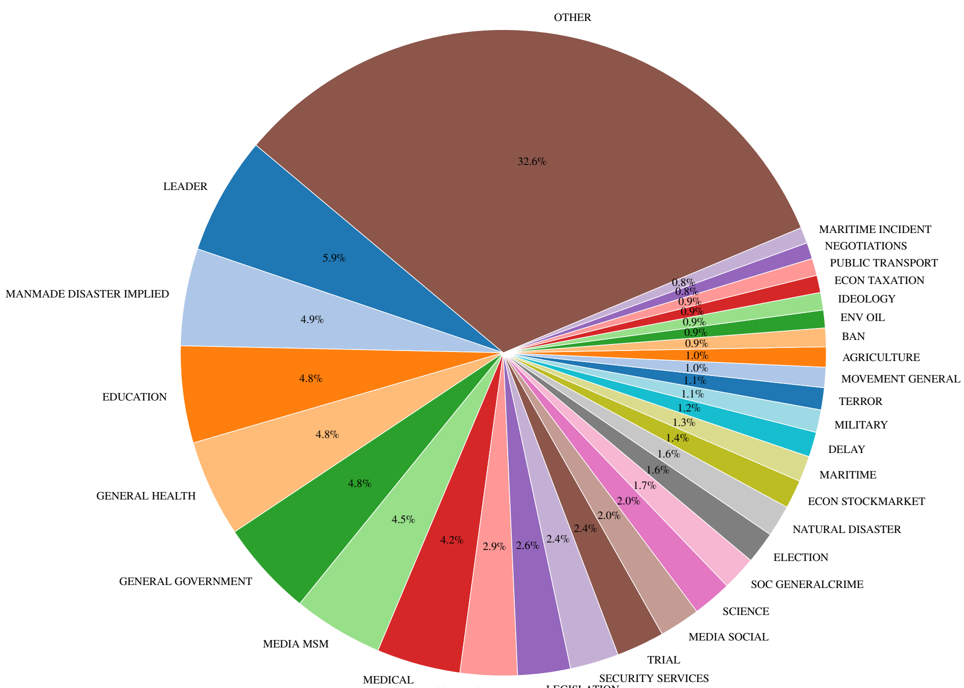


Figure 4: Time-Varying Coverage of Themes

The figure shows the top 20 themes per month in our sample, where the U.S. is the target country. The sample period runs from February 2015 through December 2023.

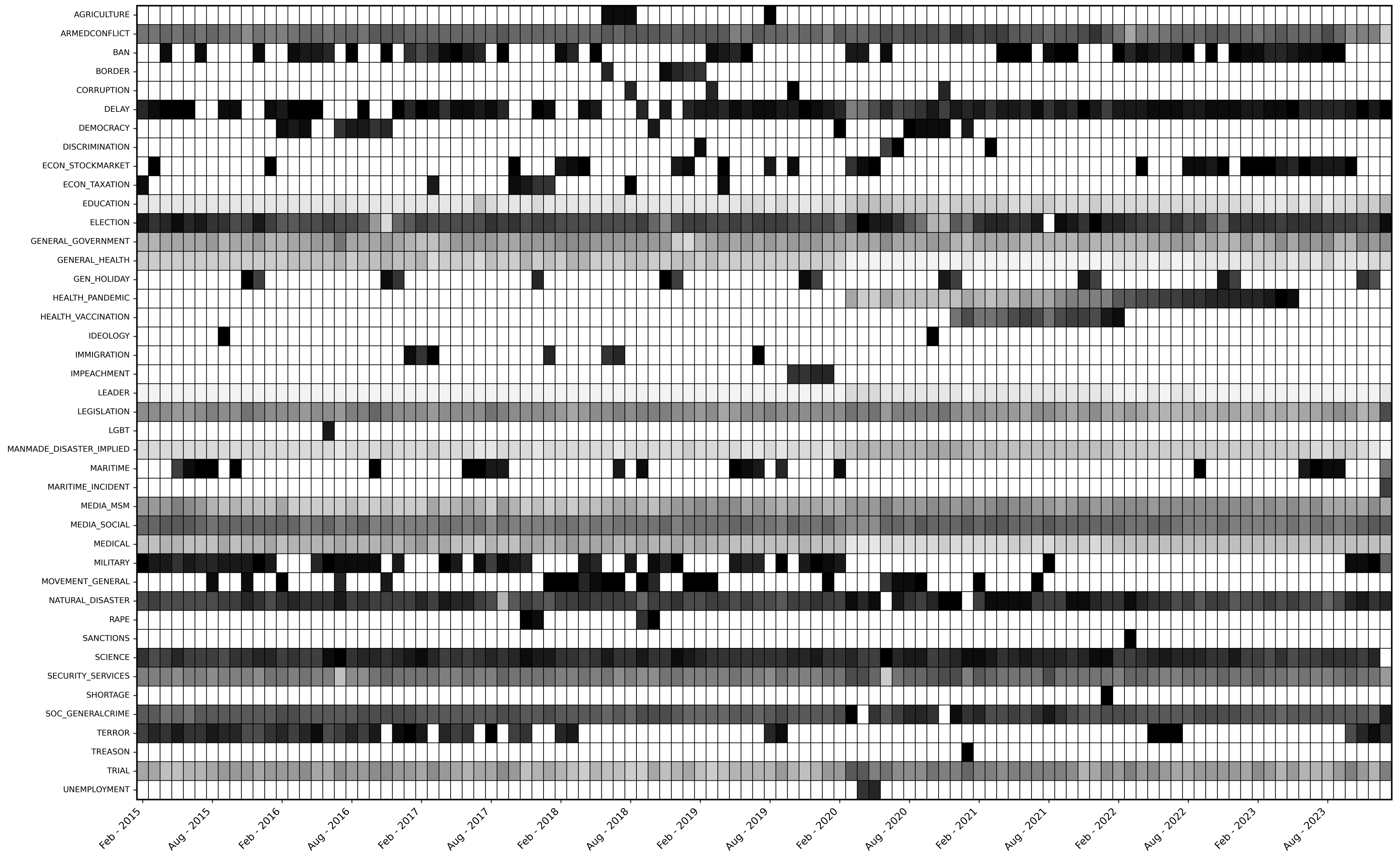


Figure 5: Cross-Country News Linkages

The figure shows the number of articles that each source country publishes about each target country as a fraction of the total number of news articles published by the source country. The sample period runs from February 2015 through December 2023.

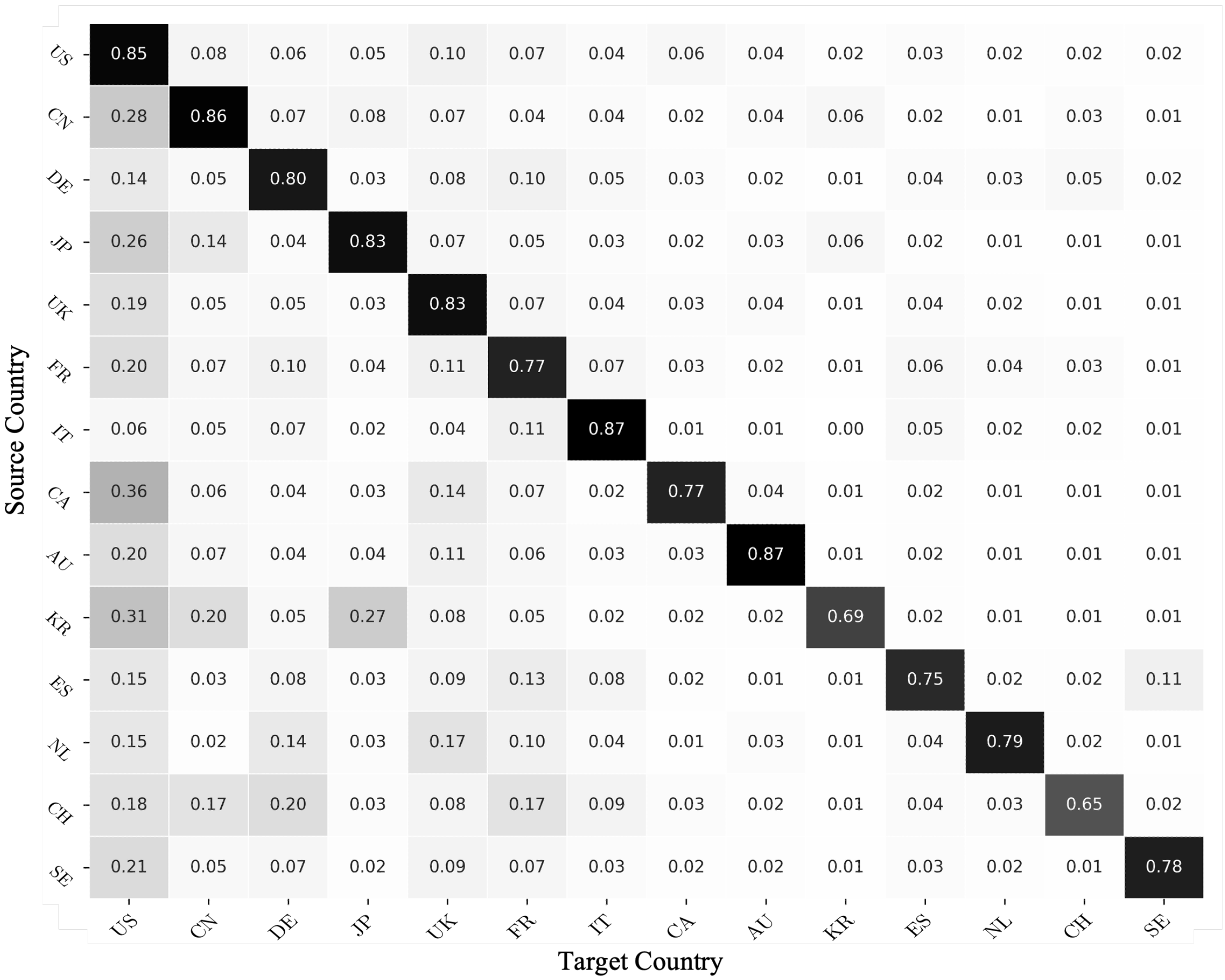
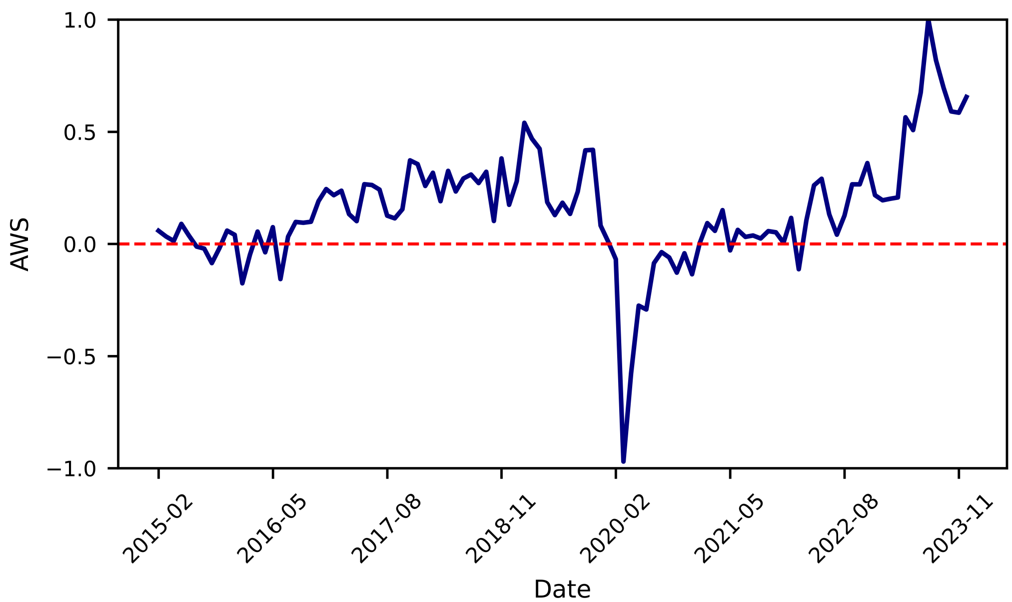


Figure 6: Net Attention-Weighted Sentiment Over Time

The figure shows the net attention-weighted sentiment (AWS) over time for the "Stock Market" theme (panel A) and the "Health Pandemic" theme (panel B). Daily net AWS values are averaged within each calendar month to produce the monthly series shown. Each AWS series is divided by its maximum absolute value, standardizing it between −1 and 1 to facilitate visual comparison across themes. The red dashed line marks zero. The sample period runs from February 2015 through December 2023.

Panel A: Stock Market



Panel B: Health Pandemic

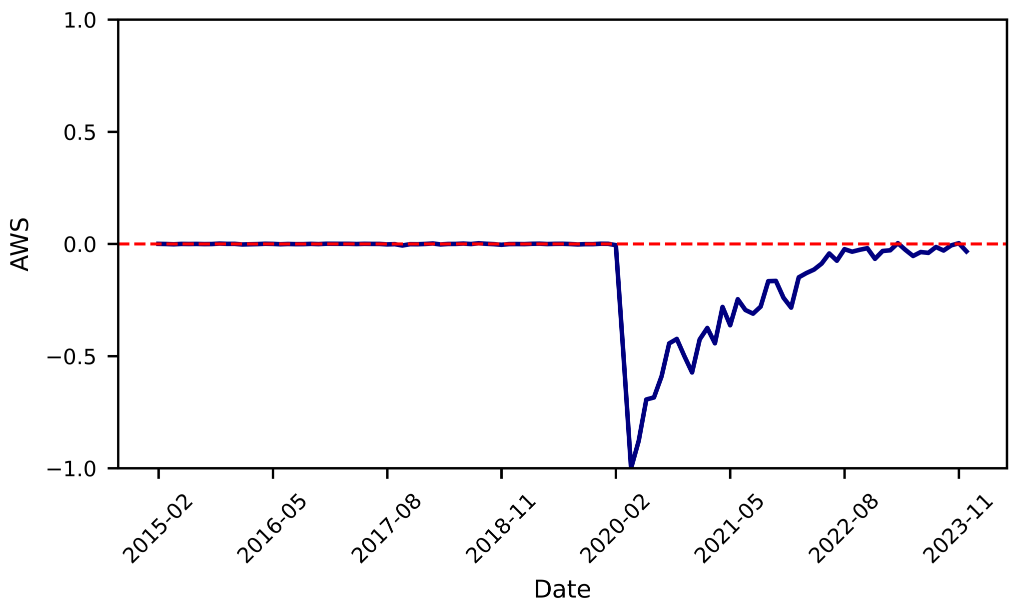


Figure 7: Time-Varying Variable Importance

The figure shows the top five most important variables, ranked by the local model for the U.S. in each quarter, that appear more than once, revealing key patterns in their contribution to forecast accuracy. The sample period runs from February 2015 through December 2023.

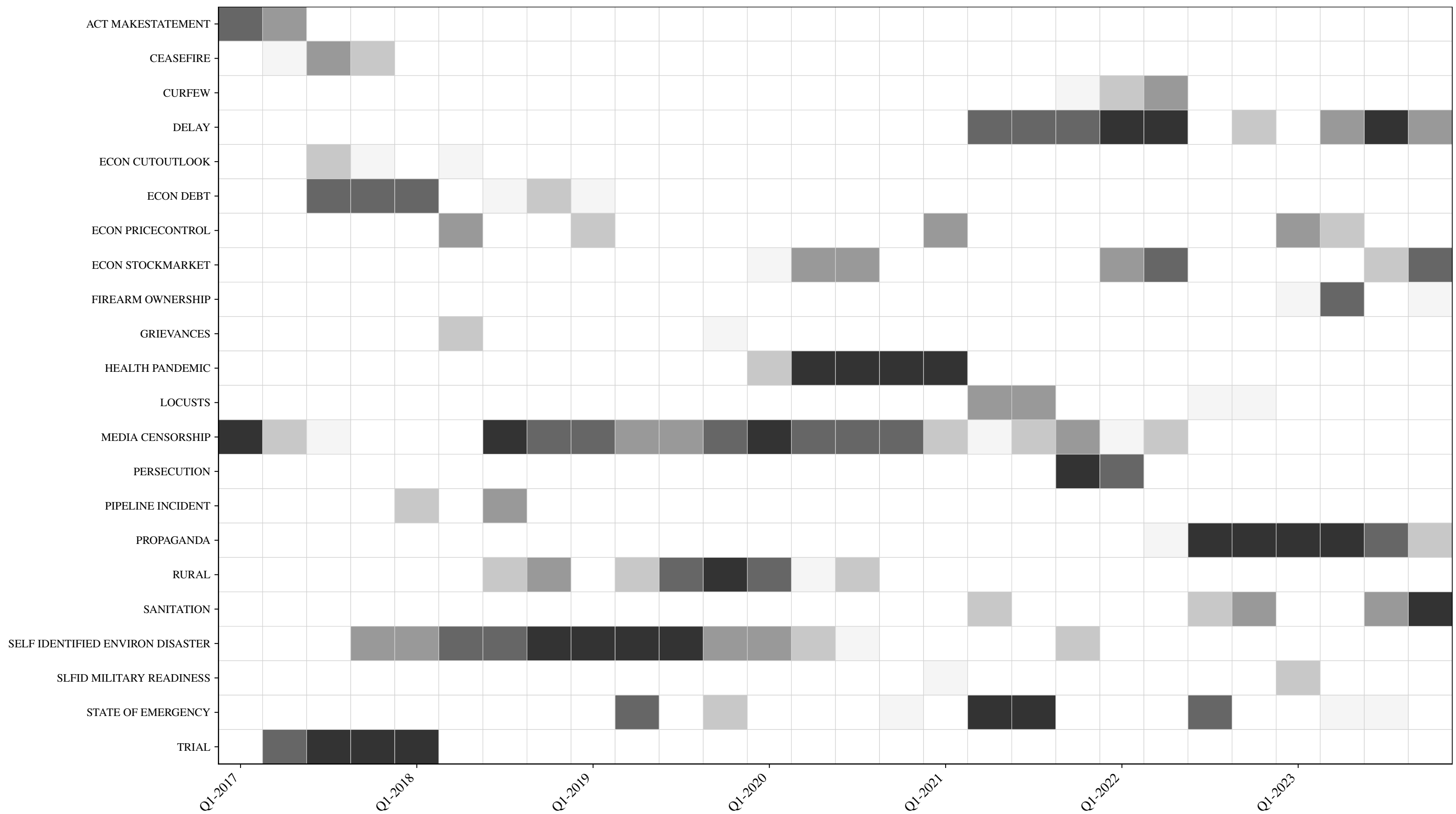


Figure 8: Global Versus Local: Spanning Regressions

The figure presents the annualized spanning alpha for each market. We regress the global sentiment strategy return on the local strategy return and report the intercept (annualized in % on the left y-axis) and the corresponding t-statistic (right y-axis). Reported in parentheses are Newey and West (1987) HAC t-statistics with 7 lags. The sample period starts in February 2015. The forecasting period runs from February 2017 till December 2023.

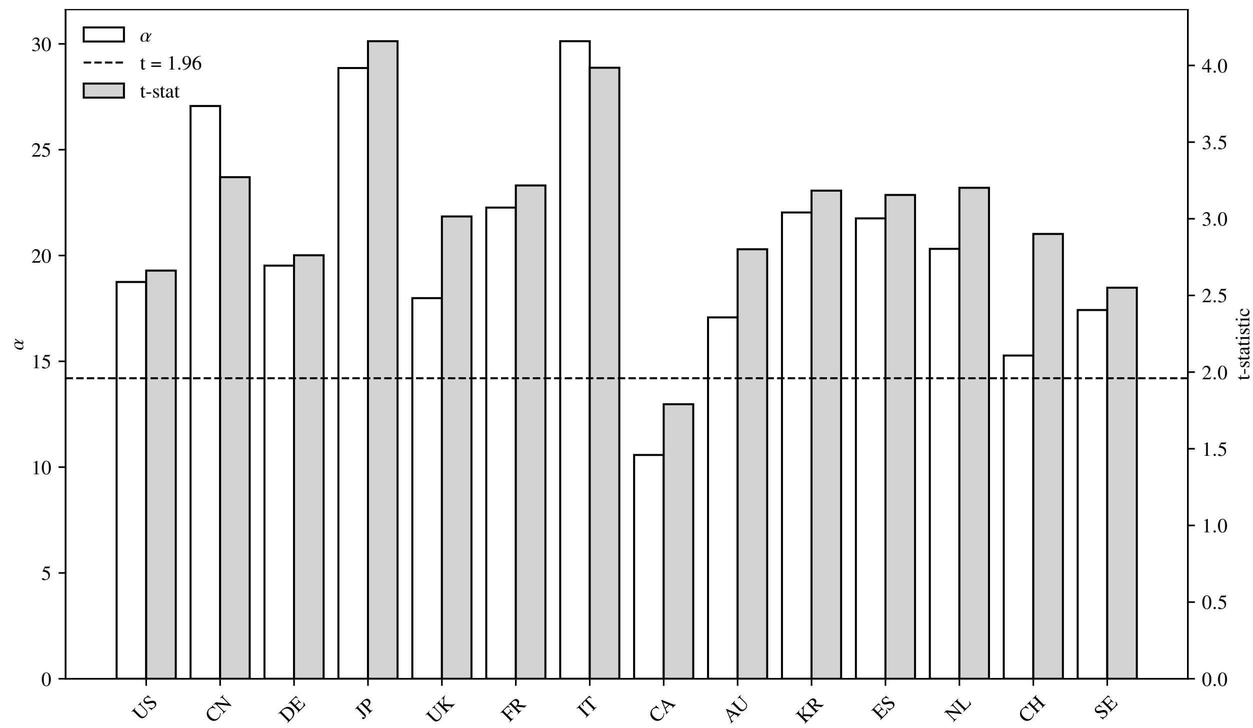


Figure 9: Theme Aggregation and Regularization

The figure shows the annualized Sharpe ratios of news sentiment strategies that: use local aggregated news (Local Agg), use global aggregated news (Global Agg), local or theme-specific news using Elastic Nets (Local and Global ELN), and our baseline strategies using Random Forests. The out-of-sample forecasting period runs from February 2017 to December 2023.

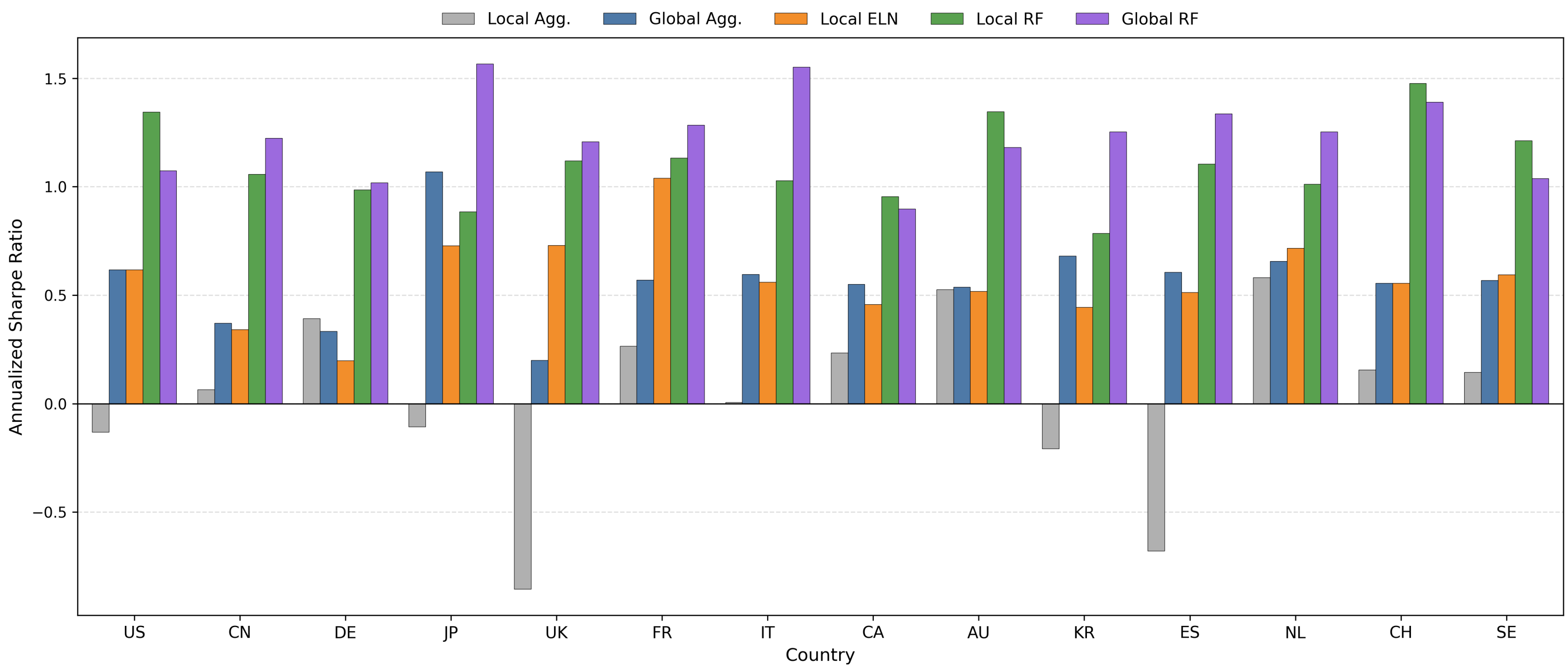


Table 1: Local News Sentiment and Equity Market Returns

This table presents performance statistics of out-of-sample local news sentiment strategies for equity markets. We report the Sharpe ratio (gross and net), the Sharpe ratio of the buy-and-hold strategy, CAPM alpha (gross and net), market beta, information ratio, and maximum drawdown of the strategy and benchmark. Reported in parentheses are t-statistics. Sharpe-ratio t-statistics follow Lo (2002) and alpha and beta t-statistics use Newey and West (1987) HAC standard errors with 7 lags. The sample period begins in February 2015, and the forecasting period spans from February 2017 to December 2023.

	SR_{gross}	SR_{net}	$SR_{B\&H}$	α	α_{net}	β	IR	DD_{max}	$DD_{max,B\&H}$
US	1.34	1.11	0.62	20.41	15.98	0.41	0.67	-23.68	-34.15
	(3.59)	(2.96)	(1.65)	(3.11)	(2.43)	(5.08)			
CN	1.06	0.82	0.01	23.32	18.17	0.12	0.79	-35.05	-50.33
	(2.82)	(2.20)	(0.02)	(2.55)	(1.98)	(2.02)			
DE	0.99	0.73	0.34	18.43	13.54	0.02	0.46	-37.35	-38.53
	(2.63)	(1.94)	(0.92)	(2.56)	(1.88)	(0.32)			
JP	0.88	0.61	0.63	14.83	9.78	0.14	0.20	-33.96	-30.80
	(2.36)	(1.64)	(1.67)	(1.98)	(1.31)	(2.58)			
UK	1.12	0.81	0.29	17.99	13.10	-0.03	0.57	-18.97	-35.18
	(2.99)	(2.17)	(0.78)	(2.33)	(1.69)	(-0.27)			
FR	1.13	0.85	0.57	23.11	17.88	-0.20	0.36	-27.77	-38.67
	(3.03)	(2.27)	(1.52)	(3.20)	(2.47)	(-1.97)			
IT	1.03	0.80	0.59	20.46	15.76	0.03	0.31	-24.12	-41.06
	(2.74)	(2.12)	(1.59)	(2.15)	(1.66)	(0.22)			
CA	0.95	0.70	0.46	14.57	10.38	0.14	0.38	-18.58	-35.88
	(2.55)	(1.87)	(1.22)	(2.00)	(1.43)	(1.00)			
AU	1.35	1.05	0.48	20.94	16.13	0.13	0.65	-23.85	-36.61
	(3.59)	(2.81)	(1.29)	(2.96)	(2.28)	(1.25)			
KR	0.78	0.53	0.34	14.50	9.70	0.01	0.32	-20.92	-41.05
	(2.10)	(1.40)	(0.90)	(2.31)	(1.54)	(0.10)			
ES	1.11	0.83	0.31	20.93	15.82	-0.06	0.54	-23.93	-39.62
	(2.95)	(2.22)	(0.84)	(2.56)	(1.94)	(-0.57)			
NL	1.01	0.74	0.63	16.48	11.81	0.06	0.28	-19.45	-35.77
	(2.70)	(1.97)	(1.67)	(2.57)	(1.84)	(0.60)			
CH	1.48	1.13	0.56	21.08	16.05	0.01	0.65	-18.62	-27.11
	(3.94)	(3.00)	(1.48)	(3.36)	(2.55)	(0.13)			
SE	1.21	0.97	0.57	20.30	15.91	0.17	0.50	-17.94	-31.33
	(3.23)	(2.59)	(1.52)	(3.13)	(2.45)	(2.65)			

Table 2: Global News Sentiment and Equity Market Returns

This table presents performance statistics of out-of-sample global news sentiment strategies for equity markets. We report the Sharpe ratio (gross and net), the Sharpe ratio of the buy-and-hold strategy, CAPM alpha (gross and net), market beta, information ratio, and maximum drawdown of the strategy and benchmark. Reported in parentheses are t-statistics. Sharpe-ratio t-statistics follow Lo (2002) and alpha and beta t-statistics use Newey and West (1987) HAC standard errors with 7 lags. The sample period begins in February 2015, and the forecasting period spans from February 2017 to December 2023.

	SR_{gross}	SR_{net}	$SR_{B\&H}$	α	α_{net}	β	IR	DD_{max}	$DD_{max,B\&H}$
US	1.07	0.80	0.62	18.01	12.96	0.19	0.36	-25.52	-34.15
	(2.87)	(2.15)	(1.65)	(2.43)	(1.75)	(2.38)			
CN	1.22	0.99	0.01	26.98	21.79	0.03	0.87	-27.83	-50.33
	(3.26)	(2.64)	(0.02)	(3.29)	(2.66)	(0.49)			
DE	1.02	0.76	0.34	18.26	13.33	0.15	0.52	-24.85	-38.53
	(2.72)	(2.02)	(0.92)	(2.30)	(1.68)	(1.73)			
JP	1.57	1.31	0.63	27.30	22.57	0.14	0.71	-19.07	-30.80
	(4.17)	(3.49)	(1.67)	(3.46)	(2.87)	(2.02)			
UK	1.21	0.87	0.29	18.53	13.19	0.16	0.70	-16.25	-35.18
	(3.22)	(2.33)	(0.78)	(3.08)	(2.19)	(2.19)			
FR	1.28	1.01	0.57	21.85	16.77	0.18	0.56	-16.34	-38.67
	(3.43)	(2.69)	(1.52)	(2.74)	(2.10)	(1.71)			
IT	1.55	1.31	0.59	31.33	26.43	0.00	0.67	-18.55	-41.06
	(4.14)	(3.49)	(1.59)	(3.95)	(3.33)	(0.00)			
CA	0.90	0.62	0.46	11.96	7.40	0.36	0.39	-14.72	-35.88
	(2.40)	(1.65)	(1.22)	(2.17)	(1.34)	(3.62)			
AU	1.18	0.89	0.48	20.21	15.39	-0.12	0.47	-13.97	-36.61
	(3.15)	(2.36)	(1.29)	(2.82)	(2.15)	(-1.35)			
KR	1.25	0.97	0.34	21.98	16.66	0.20	0.72	-31.19	-41.05
	(3.34)	(2.58)	(0.90)	(3.12)	(2.36)	(2.73)			
ES	1.34	1.04	0.31	25.37	19.85	-0.09	0.69	-26.05	-39.62
	(3.56)	(2.77)	(0.84)	(3.19)	(2.50)	(-0.87)			
NL	1.25	0.95	0.63	21.10	16.02	0.01	0.44	-24.15	-35.77
	(3.34)	(2.54)	(1.67)	(2.55)	(1.94)	(0.11)			
CH	1.39	1.06	0.56	18.56	13.77	0.18	0.65	-17.00	-27.11
	(3.71)	(2.82)	(1.48)	(2.78)	(2.07)	(1.44)			
SE	1.04	0.79	0.57	18.45	13.95	0.05	0.34	-19.34	-31.33
	(2.77)	(2.11)	(1.52)	(2.72)	(2.06)	(0.53)			

Table 3: U.S.-Sourced News Sentiment and Equity Market Returns

This table presents performance statistics of out-of-sample United States sourced news sentiment strategies for equity markets. We report the Sharpe ratio (gross and net), the Sharpe ratio of the buy-and-hold strategy, CAPM alpha (gross and net), market beta, information ratio, and maximum drawdown of the strategy and benchmark. Reported in parentheses are t-statistics. Sharpe-ratio t-statistics follow Lo (2002) and alpha and beta t-statistics use Newey and West (1987) HAC standard errors with 7 lags. The sample period begins in February 2015, and the forecasting period spans from February 2017 to December 2023.

	SR_{gross}	SR_{net}	$SR_{B\&H}$	α	α_{net}	β	IR	DD_{max}	$DD_{max,B\&H}$
US	1.34	1.11	0.62	20.41	15.98	0.41	0.67	-23.68	-34.15
	(3.59)	(2.96)	(1.65)	(3.11)	(2.43)	(5.08)			
CN	1.31	1.08	0.01	28.90	23.75	-0.00	0.92	-25.74	-50.33
	(3.50)	(2.87)	(0.02)	(3.26)	(2.68)	(-0.00)			
DE	0.75	0.48	0.34	14.17	9.13	-0.02	0.28	-32.53	-38.53
	(1.99)	(1.28)	(0.92)	(1.93)	(1.24)	(-0.17)			
JP	1.11	0.85	0.63	19.02	14.21	0.13	0.36	-17.76	-30.80
	(2.96)	(2.27)	(1.67)	(2.56)	(1.92)	(1.48)			
UK	1.16	0.86	0.29	17.81	13.06	0.14	0.66	-19.29	-35.18
	(3.09)	(2.29)	(0.78)	(2.56)	(1.88)	(1.24)			
FR	0.83	0.54	0.57	13.70	8.30	0.17	0.20	-24.64	-38.67
	(2.23)	(1.45)	(1.52)	(1.60)	(0.97)	(1.28)			
IT	0.98	0.73	0.59	17.67	12.54	0.18	0.30	-26.62	-41.06
	(2.62)	(1.95)	(1.59)	(2.27)	(1.61)	(2.53)			
CA	1.16	0.84	0.46	19.50	14.29	-0.09	0.47	-14.10	-35.88
	(3.09)	(2.24)	(1.22)	(2.49)	(1.82)	(-0.56)			
AU	1.28	0.98	0.48	18.35	13.46	0.32	0.69	-23.64	-36.61
	(3.42)	(2.62)	(1.29)	(2.89)	(2.12)	(2.85)			
KR	0.93	0.69	0.34	14.90	10.55	0.36	0.52	-28.90	-41.05
	(2.47)	(1.85)	(0.90)	(2.37)	(1.68)	(7.02)			
ES	1.38	1.11	0.31	24.62	19.54	0.18	0.83	-20.48	-39.62
	(3.69)	(2.96)	(0.84)	(3.34)	(2.65)	(2.16)			
NL	1.67	1.42	0.63	24.80	20.56	0.33	0.90	-19.28	-35.77
	(4.46)	(3.79)	(1.67)	(3.44)	(2.86)	(3.55)			
CH	0.96	0.60	0.56	13.51	8.28	0.04	0.29	-13.83	-27.11
	(2.56)	(1.59)	(1.48)	(2.63)	(1.62)	(0.34)			
SE	1.37	1.13	0.57	22.43	17.98	0.24	0.65	-30.22	-31.33
	(3.66)	(3.01)	(1.52)	(2.93)	(2.35)	(2.75)			

Table 4: Alternative Sentiment Measures

This table reports the annualized Sharpe ratio of an out-of-sample local news sentiment strategy applied to the United States using a range of sentiment measures. Reported in parentheses are Sharpe-ratio t-statistics that follow Lo (2002). The out-of-sample period runs from February 2017 till December 2023.

	Sharpe Ratio
SentiWordNet3	1.34 (3.59)
Loughran McDonald	1.15 (3.07)
Subjectivity Lexicon	1.31 (3.50)
ML-SENTICON	1.34 (3.58)
Lexicoder	1.33 (3.56)

Table 5: Local News and Accounting for Lagged Returns

This table presents performance statistics of out-of-sample local news sentiment strategies for equity markets, accounting for lagged returns. We account for lagged returns by incorporating a one-day lagged return in the set of predictors. We report the Sharpe ratio (gross and net), the Sharpe ratio of the buy-and-hold strategy, CAPM alpha (gross and net), market beta, information ratio, and maximum drawdown of the strategy and benchmark. Reported in parentheses are t-statistics. Sharpe-ratio t-statistics follow Lo (2002) and alpha and beta t-statistics use Newey and West (1987) HAC standard errors with 7 lags. The sample period begins in February 2015, and the forecasting period spans from February 2017 to December 2023.

	SR_{gross}	SR_{net}	$SR_{B\&H}$	α	α_{net}	β	IR	DD_{max}	$DD_{max,B\&H}$
US	1.50	1.26	0.62	25.96	21.38	0.19	0.69	-21.73	-34.15
	(4.01)	(3.36)	(1.65)	(3.77)	(3.10)	(8.10)			
CN	0.88	0.66	0.01	19.49	14.55	-0.07	0.60	-43.38	-50.33
	(2.36)	(1.76)	(0.02)	(2.36)	(1.76)	(-2.91)			
DE	0.57	0.31	0.34	10.48	5.58	0.03	0.16	-47.52	-38.53
	(1.52)	(0.82)	(0.92)	(1.48)	(0.79)	(1.42)			
JP	1.19	0.91	0.63	20.78	15.64	0.11	0.42	-21.66	-30.80
	(3.18)	(2.44)	(1.68)	(3.01)	(2.26)	(4.69)			
UK	0.75	0.43	0.29	11.37	6.18	0.14	0.35	-26.36	-35.18
	(2.02)	(1.15)	(0.78)	(1.92)	(1.04)	(6.21)			
FR	1.32	1.07	0.57	25.41	20.73	-0.09	0.51	-22.08	-38.67
	(3.54)	(2.86)	(1.52)	(3.69)	(3.01)	(-3.84)			
IT	1.12	0.89	0.59	22.62	17.96	-0.00	0.37	-25.05	-41.06
	(2.99)	(2.37)	(1.59)	(2.98)	(2.37)	(-0.03)			
CA	0.75	0.48	0.46	10.68	6.36	0.20	0.23	-32.69	-35.88
	(2.00)	(1.29)	(1.22)	(1.78)	(1.06)	(8.82)			
AU	1.09	0.80	0.48	15.98	11.22	0.23	0.49	-27.19	-36.61
	(2.91)	(2.13)	(1.29)	(2.69)	(1.89)	(9.98)			
KR	0.99	0.72	0.34	17.89	12.75	0.09	0.49	-18.65	-41.05
	(2.66)	(1.92)	(0.90)	(2.59)	(1.84)	(3.81)			
ES	1.27	1.02	0.31	22.97	18.30	0.11	0.72	-22.00	-39.62
	(3.40)	(2.72)	(0.84)	(3.32)	(2.65)	(4.79)			
NL	1.09	0.81	0.63	17.54	12.84	0.09	0.34	-25.87	-35.77
	(2.91)	(2.17)	(1.67)	(2.77)	(2.03)	(3.63)			
CH	1.35	1.01	0.56	17.74	12.88	0.20	0.62	-12.42	-27.11
	(3.60)	(2.69)	(1.48)	(3.37)	(2.44)	(8.71)			
SE	1.33	1.07	0.57	22.29	17.66	0.18	0.59	-17.76	-31.33
	(3.55)	(2.87)	(1.52)	(3.32)	(2.63)	(7.88)			

Table 6: Native Versus Translated Local News

This table presents performance statistics of out-of-sample local news sentiment strategies for equity markets using either translated (to English) news or native news. We report the Sharpe ratio, CAPM alpha, market beta, and information ratio of the strategies. The columns with (ENG) indicate that the news was first translated to English and then the sentiment was calculated using the English version of Hedonometer. The columns without the (ENG) tag indicate that the sentiment was calculated directly using the native language version of Hedonometer. Reported in parentheses are t-statistics. Sharpe-ratio t-statistics follow Lo (2002) and alpha and beta t-statistics use Newey and West (1987) HAC standard errors with 7 lags. The out-of-sample period spans from February 2017 to December 2023.

	SR_{gross}^{ENG}	SR_{gross}	α^{ENG}	α	IR^{ENG}	IR
CN	1.09	1.12	24.03	24.73	0.77	0.77
	(2.91)	(2.99)	(2.91)	(3.00)		
DE	1.44	1.22	25.54	22.73	0.87	0.63
	(3.84)	(3.26)	(3.72)	(3.23)		
FR	0.86	0.88	14.55	13.80	0.22	0.25
	(2.29)	(2.36)	(2.11)	(2.05)		
KR	1.44	1.28	25.66	22.98	0.85	0.70
	(3.85)	(3.42)	(3.75)	(3.33)		
ES	0.95	0.92	17.91	17.91	0.44	0.40
	(2.54)	(2.45)	(2.57)	(2.60)		

Appendices

Table A1: Sample of Equity Futures Contracts

Reported are the country name ("Country"), local close time ("Close"), timezone ("TZ"), name of the market index ("Index"), the Bloomberg ticker ("BB"), and the annualized Sharpe ratio ("Sharpe"). All data are obtained from Bloomberg from February 2015 till December 2023.

Country	Close	TZ	Index	BB	Sharpe
United States	17:00	America/Chicago	S&P 500	ES	0.58
China	16:30	Asia/Hong_Kong	HSI	HI	0.31
Germany	22:05	Europe/Berlin	DAX	GX	0.30
Japan	15:16	Asia/Tokyo	Nikkei 225	NK	0.51
United Kingdom	21:00	Europe/London	FTSE 100	Z	0.31
France	22:00	Europe/Paris	CAC 40	CF	0.49
Italy	22:00	Europe/Rome	FTSE MIB	ST	0.43
Canada	16:30	America/Montreal	TSX 60	PT	0.42
Australia	16:30	Australia/Sydney	ASX SPI 200	XP	0.38
South Korea	15:46	Asia/Seoul	KOSPI 200	KM	0.33
Spain	20:00	Europe/Madrid	IBEX 35	IB	0.22
Netherlands	22:00	Europe/Amsterdam	AEX	EO	0.55
Switzerland	22:05	Europe/Zurich	SMI	SM	0.44
Sweden	17:25	Europe/Stockholm	OMXS 30	QC	0.46

Figure A1: Global Importance Over Other Countries

The figure shows the sum of standardized variable importance of a given country (on the x-axis) over other countries in the global model estimation. These sums are used to scale node sizes in Figure 1.

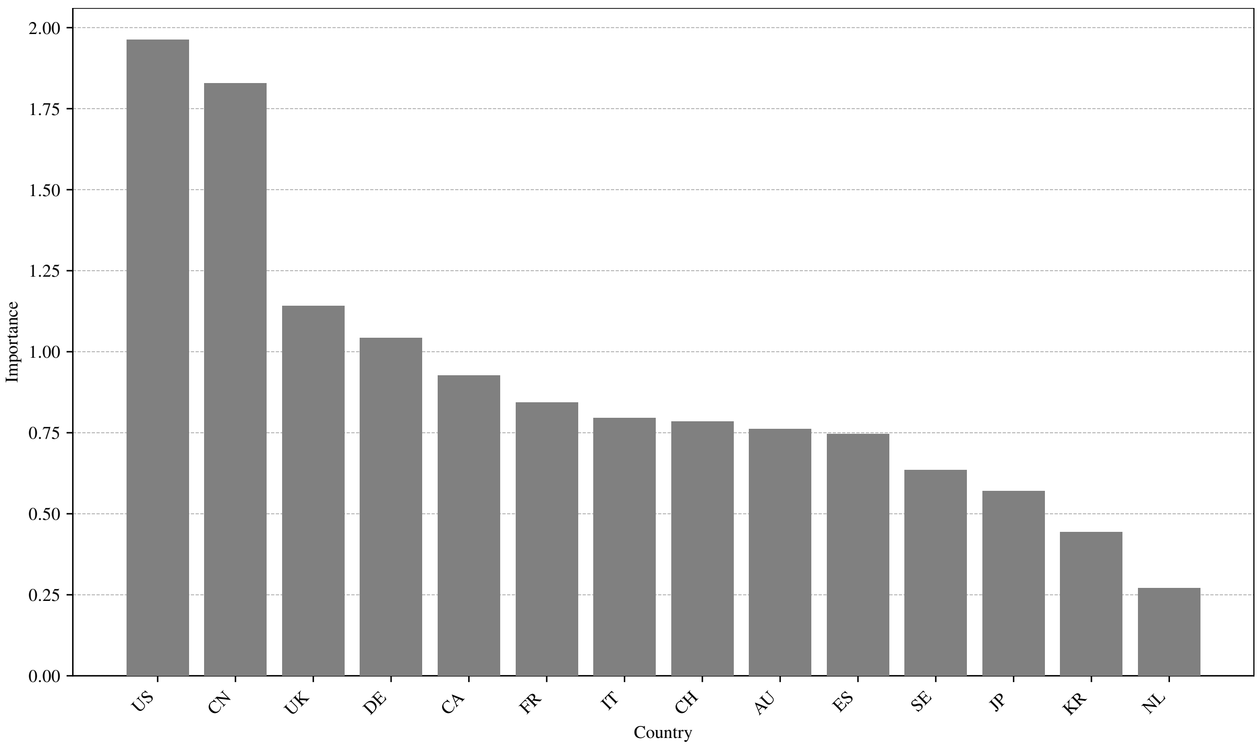


Figure A2: Cumulative Variable Importance of Top 5 Local Themes

The figure plots the cross-country average (over the 14 local Random Forest models) of the share of total gain importance held by each model's five most important themes, evaluated at every retrain. The out-of-sample period runs from February 2017 to December 2023.