Fakultät für Mathematik und Wirtschafts-wissenschaften
Institut für Finanzwirtschaft

War Discourse and Disaster Premium

Seminar Selected Topics in Finance

Vorgelegt von:
Daria Palitzsch
daria.palitzsch@uni-ulm.de
Guorui Wang
guorui.wang@uni-ulm.de
Daniel Hirschle
daniel.hirschle@uni-ulm.de

Fassung June 19, 2025

Satz: PDF-LaTeX 2_ε

Contents

1 Introduction

Recent work in narrative finance argues that investor beliefs and market risk premia are influenced by prevailing societal narratives (Shiller [2017], Pastor and Veronesi [2019]). Hirshleifer et al. [2024] provide empirical support for this idea by showing that attention to disaster-related discourse topics, such as “war” or “pandemic,” predicts excess stock market returns. Using topic modeling on newspaper articles to extract attention to rare disaster themes, they document a “disaster premium”: periods with elevated media discourse around events like war or pandemics are followed by higher stock market excess returns.

This seminar paper replicates key findings of Hirshleifer et al. [2024] using monthly Wikipedia data and S&P 500 excess returns up to 2020. We then extend the original study along three dimensions. First, we expand the sample to include the post-2020 period to test the robustness of narrative-based predictability during recent crises. Second, we introduce two new discourse topics—*Trade war* and *Tariff*—that capture economic tensions not covered in the original topic set. Third, we evaluate whether Google Trends data can serve as an alternative or complementary measure of investor attention by comparing it to Wikipedia-based predictors.

Our analysis includes both in-sample and out-of-sample predictive regressions. We also evaluate forecasting accuracy using cumulative squared prediction errors. The structure of the paper is as follows: Chapter 2 reviews the theoretical background; Chapter 3 summarizes the original methodology; Chapter 4 outlines our data sources and implementation; Chapter 5 presents empirical results; and Chapter 6 discusses implications and future research directions.

2 Literature & Conceptual Framework

2.1 Rare Disasters and the Equity Premium Puzzle

According to Mehra and Prescott [1985] the equity premium puzzle poses a fundamental challenge for asset-pricing theory. Although historical data show that the average excess return on stocks exceeds 6% annually, such magnitudes are not captured by standard representative-agent models with plausible levels of risk aversion. Specifically, these models usually assume log-normal consumption growth and calibrate risk aversion coefficients between 1 and 5. Under these assumptions, the implied equity premium remains significantly lower than empirical benchmarks. This disparity has motivated researchers to look for structural explanations beyond Gaussian risk, particularly considering the role of low-probability, high-impact events—so-called *rare disasters*.

Rietz [1988] made an important contribution in this area by introducing a three-state variant of the Lucas tree model, incorporating a low-probability disaster state into the stochastic process of consumption growth. The model assumes that log consumption evolves as:

$$\Delta \ln C_{t+1} = g + u_t,$$

where $u_t = \ln(1-b)$ with probability p, and $u_t = 0$ with probability $1-p$. Here, $b \in (0,1)$ represents the fractional drop in consumption during a disaster, and p denotes the probability of such an event. The agent has constant relative risk aversion (CRRA) preferences given by:

$$U = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \cdot \frac{C_t^{1-\alpha}}{1-\alpha} \right],$$

where $\alpha > 0$ is the coefficient of relative risk aversion, and $\beta \in (0,1)$ is the subjective discount factor. With plausible parameter values such as $\alpha = 7$, $b = 0.3$, and $p = 0.004$, Rietz demonstrates that the model can generate equity premiums consistent with observed historical returns.

The central insight of Rietz's model lies in the nonlinear effect of disaster risk on marginal utility. Even a very small probability of a 30% drop in consumption occurring once every several decades can substantially increase the value of risk-free assets, thereby lowering the risk-free rate. At the same time, investors require a higher premium to hold equities that are exposed to such tail risks. Notably, this mechanism operates not through higher consumption volatility per se, but through the asymmetry and fat-tailed nature of rare shocks. As a result, the model is capable of producing realistic asset returns without relying on unrealistically high levels of risk aversion.

This paradigm is further developed by Barro [2006], who grounds the rare-disaster model in empirical evidence. Drawing on long-run GDP and consumption data from 35 countries over the twentieth century, Barro defines disaster episodes as cumulative output declines of at least 15%. The average contraction magnitude is close to 29%, and the estimated disaster probability is approximately 1.7% per year. When these parameters are incorporated into the model, it continues to explain both the historically low risk-free rate and the high equity premium. Unlike Rietz's theoretical calibration, Barro's data-driven approach underscores that rare disasters—such as wars, financial crises, and institutional breakdowns—are empirically observable phenomena. This supports a forward-looking interpretation of investor behavior in the presence of macro-catastrophic expectations.

Barro [2009] further enhances the model by incorporating Epstein–Zin–Weil (EZW) preferences, which decouple risk aversion from intertemporal substitution. This distinction allows for more flexibility in matching asset-price data. The recursive utility specification introduces two key parameters: the coefficient of relative risk aversion γ, and the elasticity of intertemporal substitution (EIS) ψ. Under these preferences, the expected equity premium becomes:

$$\mathbb{E}(r^e - r^f) \approx \gamma \cdot (\sigma^2 + pb^2),$$

while the risk-free rate is approximately:

$$r^f \approx \rho + \psi g - \frac{1}{2}\gamma(\sigma^2 + pb^2).$$

Barro [2009] uses this extended model to quantify the welfare cost of disaster risk. He defines the welfare gain from eliminating rare disasters as the fraction of consumption households would be willing to forgo to avoid disaster risk altogether. Numerical calibration suggests that this fraction could be as high as 20–30% of GDP annually. This highlights not only the impact of rare disasters on asset prices, but also their profound implications for long-term economic welfare. The presence of even latent disaster risk alters optimal saving and investment behavior, further validating the relevance of rare-disaster models in both theoretical and applied finance.

2.2 Narratives and Market Expectations

Recent work in economics increasingly explores the role of narratives—stories or popular beliefs—in shaping economic decisions. Shiller [2017] introduces the concept of *narrative economics*, arguing that economic actors often respond not only to fundamental data but also to the influence of widely shared stories. He suggests that, much like infectious diseases, narratives can spread through a population—a process that can be illustrated using models from epidemiology. Specifically, Shiller [2017] adopts the well-known SIR model (Susceptible-Infected-Recovered) to describe narrative diffusion:

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta SI, \\
\frac{dI}{dt} &= \beta SI - \gamma I, \\
\frac{dR}{dt} &= \gamma I,
\end{aligned}
$$

where S denotes the fraction of the population that is susceptible to a narrative but not yet exposed; I represents individuals actively spreading the narrative; and R indicates individuals who are no longer spreading the narrative due to loss of interest or memory decay. The parameter β captures the rate of narrative transmission, dependent on factors such as media coverage and social interaction intensity, while γ measures the rate at which narratives lose momentum or relevance.

While Shiller's framework is primarily illustrative and does not involve rigorous statistical validation, it emphasizes the importance of measuring narrative intensity to capture shifts in economic sentiment. He suggests employing text analysis methods, including keyword frequency tracking through historical news archives and digital databases such as Google Ngram Viewer, to quantify narratives. This approach can provide real-time indicators of changing public perceptions, potentially explaining episodes of market exuberance or panic beyond what fundamental economic indicators alone predict.

Pastor and Veronesi [2013] present a formal general equilibrium model to illustrate how) uncertainty surrounding political decisions, often communicated through public narratives and policy signals, systematically influences asset prices and risk premia. They propose a setting in which firms' profits evolve according to:

$$d\Pi_t^i = (\mu + g_t)\, dt + \sigma\, dZ_t + \sigma_1\, dZ_t^i,$$

where Π_t^i denotes the profit of firm i at time t; μ is the long-run average profit growth rate; and g_t represents the current, yet uncertain, policy impact on profits. The uncertainty regarding g_t arises because its true value is not directly observable, requiring investors to learn from profit realizations over time. The stochastic components consist of a common systematic shock dZ_t, affecting all firms equally (such as macroeconomic conditions), and an idiosyncratic shock dZ_t^i, unique to each firm. The parameters σ and σ_1 quantify the sensitivity of profits to systematic and firm-specific shocks, respectively.

Through the use of Bayesian learning procedures, investors continuously update their assumptions about g_t and prospective future policy changes in response to political cues and observed corporate profitability. The fact that investor perceptions vary makes this learning mechanism even riskier. As a result, even when the underlying economic fundamentals are solid, times of uncertainty, such as those around policy changes or elections, show more market volatility and higher risk premia. The result is that narrative-driven changes in investor perceptions are frequently responsible for asset price swings and higher demanded returns.

2.3 Attention Proxies in Asset Markets

There is increasing interest in leveraging public attention as a predictor of market returns, according to recent asset pricing studies. Da et al. [2011] show that Google Search Volume Indices-direct measures of investor attention predict abnormal returns and subsequent reversals, highlighting the value of attention proxies for capturing retail and institutional focus on topics such as pandemics, war, and political events.

Topic weights are constructed from news using seeded Latent Dirichlet Allocation (sLDA) in the original study by Hirshleifer et al. [2024]. This approach uses the quantity of news that are related to 14 primary narrative topics to track their monthly intensity. Their forecasting models are based on these news-based variables.

Nevertheless, editorial choices may have an impact on news data, which might not accurately represent popular opinion. We expand their strategy by utilizing Google Trends and Wikipedia pageviews to increase transparency and reflect current public interest. Wikipedia displays passive information demand, but Google Trends displays active search activity. Both offer a clearer picture of people's priorities at any given moment.

3 Summary of Hirshleifer et al. [2024]

3.1 Key Research Idea

Hirshleifer et al. [2024] investigate whether media narratives on disaster-related topics—such as war, pandemics, market crashes, and economic downturns—can forecast future stock market excess returns. Their hypothesis builds on insights from behavorial finance and rare-disaster asset pricing, suggesting that salient yet infrequent risk narratives can systematically influence investor expectations.

To empirically test this idea, the authors construct monthly time series of topic-specific media attention by applying natural language processing (NLP) techniques to a large corpus of financial news articles. These topic indices quantify the share of coverage devoted to various themes (e.g., war, pandemics). They are then used in predictive regressions to forecast S&P 500 excess returns.

A key methodological innovation is the use of Seeded Latent Dirichlet Allocation (Seeded LDA), which directs topic modeling toward economically relevant categories using predefined keyword sets. The authors compare the forecasting performance of these discourse-based predictors to traditional economic predictors (e.g., dividend-price ratio, earnings-price ratio, term spread), but do not incorporate alternative attention proxies such as Google Trends or Wikipedia pageviews, which we explore in our extension.

Their central finding is that war-related media discourse strongly predicts excess returns, both in-sample and out-of-sample, outperforming standard predictors. This suggests that investor underreaction to rare-disaster risk - captured through narrative silence - may drive return predictability.

3.2 Method Overview

To implement their analysis, Hirshleifer et al. [2024] develop a structured empirical framework that integrates natural language processing (NLP) with predictive economic modeling. The authors begin by identifying media discourse topics using **Seeded Latent Dirichlet Allocation (Seeded LDA)**, a semi-supervised topic modeling algorithm. Predefined sets of seed words guide the model toward economically meaningful themes such as war, pandemics, and trade tensions. Seeded LDA then estimates the probability distribution of topics for each document in a large corpus of financial news articles sourced from Bloomberg and Thomson Reuters.

From these document-level topic distributions, the authors construct **monthly topic indices**. For each month, they compute the average share of media coverage dedicated to a particular topic. These indices capture the salience of specific narratives - such as war or pandemic risks in the financial media at each point in time.

The resulting time series are used as predictors in **predictive regressions** of the form

$$R_{t+1} = \alpha + \beta \cdot \text{TopicIndex}_t + \varepsilon_{t+1}$$

where R_{t+1} denotes the excess return of the S&P 500 in month $t+1$, and TopicIndex_t is the lagged topic salience. The models are estimated using ordinary least squares (OLS), with Newey-West standard errors to correct for heteroscedasticity and autocorrelation.

To evaluate performance, the authors benchmark their discourse-based predictors against **traditional economic variables** such as the dividend-price ratio, earnings-price ratio and term spread. Forecast accuracy is assessed both *in-sample* (via R^2) and *out-of-sample* (via R^2_{OS}), using rolling-window estimations and multiple robustness checks.

3.3 Key Findings

The empirical results show that **war-related media discourse exhibits strong predictive power** for U.S. equity returns. Specifically, the war-topic index negatively predicts next month S&P 500 excess returns in a statistically and economically significant manner. Higher war-related media attention is associated with lower future returns, consistent with heightenend investor concerns about rare-disaster risk.

In **in-sample regressions**, estimated using monthly data from 1996 to 2020, the war-topic index yields and R^2 of up to 3%. This level of explanatory power is notable in return prediction, where R^2 values are typically small. Other topics, such as pandemics, terrorism, or trade conflicts, show weaker and less consistent results.

To test real-time relevance, the authors conduct **out-of-sample forecasting** using recursive window estimation. The war-topic index delivers positive out-of-sample R^2_{OS} values compared to a historical-mean benchmark, confirming its predictive usefulness beyond the estimation sample. Forecast combination methods, such as Bayesian model averaging, further improve performance.

When compared to **standard economic predictors** such as the dividend-price ratio, earnings-price ratio, and term spread, the war-topic index performs better in both in-sample and out-of-sample settings. Even when included in multivariate regressions alongside these variables, the war-topic retains its significance, suggesting it provides complementary information.

The authors also explore **robustness** through alternative specifications: changes in seed words, different estimation windows, and exclusion of financial articles, . The predictive power of the war-topic remains stable. Moreover, the index also forecasts implied volatility and future equity variance, supporting the interpretation that it captures variation in perceived disaster risk.

3.4 Interpretation and Implications

The findings of Hirshleifer et al. [2024] highlight a meaningful connection between media narratives, investor sentiment, and expected returns. The war-topic index appears to function as a proxy for **disaster salience** - the prominence of rare and extreme risks in the media landscape and, by extension, in investor attention.

From a theoretical standpoint, the results are consistent with **rare-disaster asset pricing models**, which suggest that perceived shifts in the probability of catastrophic events can drive fluctuations in risk premia Barro [2006]. The war-topic index does not measure realized disasters, but rather investor-perceived disaster risk as inferred from the intensity of media attention

The paper also offers a behavioral interpretation. Investors may exhibit **underreaction to salient but low-probability narratives**, such as geopolitical conflict, due to limited attention or cognitive biases. This delayed price adjustment leads to return predictability. The results align the the broader **narrative economics** literature, which argues that emotionally resonant stories play a central role in shaping marked outcomes Shiller [2020].

From a practical perspective, discourse-based indices offer a **scalable forecasting tool** that relies solely on publicly available media content. These indices can be constructed in real time using NLP methods and may serve as early warning signals for changes in market sentiment and risk perception.

That said, the approach is not without limitations. The predictive power of discourse indices may be **topic- and time-specific**, depending on how relevant a given theme is to investors at a particular point. Furthermore, **media salience does not always reflect objective economic risk**; it captures perceived or narrative-based attention, which may diverge from fundamentals.

Overall, the study contributes to a growing literature that combines textual analysis with asset pricing and behavioral finance. It illustrates how narrative-based measures can complement traditional financial indicators and provide new insights into the dynamics of risk and return.

4 Data & Methodology

This chapter outlines the datasets and procedures used to replicate and extend the analysis of Hirshleifer et al. [2024]. Our empirical setup combines financial return data with discourse-based attention proxies from digitial sources.

4.1 Data Sources

4.1.1 S&P 500 and Return Data

Our analysis uses monthly excess returns on the S&P 500 index from January 1871 to November 2024, obtained from the A. Goyal dataset [1] (link provided below).

The one-month-ahead excess return is defined as:

$$R_{t+1} = r_{t+1}^{\text{S\&P}} - r_{t+1}^{\text{rf}},$$

where $r_{t+1}^{\text{S\&P}}$ denotes the log return on the S&P 500 index including dividends, and r_{t+1}^{rf} is the risk-free rate derived from short-term Treasury bills

In addition to returns, we include a set of widely used macroeconomic predictors: the dividend-price ratio (DP), dividend yield (DY), earnings-price ratio (EP), dividend-earnings ratio (DE), and the three-month Treasury bill rate (TBL). These serve as benchmarks to evaluate the predictive power of discourse-based indicators in later sections.

4.1.2 Wikipedia Pageviews

To capture public attention to disaster- and sentiment-related topics, we compile monthly pageview statistics for 20 English-language Wikipedia articles that correspond

[1] https://sites.google.com/view/agoyal145

to economically relevant narratives.[2]

These articles mirror the seeded topics used in Hirshleifer et al. [2024].

We standardize each pageview series to facilitate interpretation and comparability across topics. The standardization is based on the in-sample mean and standard deviation.

4.1.3 Google Trends

To complement the Wikipedia analysis, we collect monthly Google Trends indices for a subset of ten topics. These were either weakly represented in Wikipedia or are likely to trigger search activity (e.g. *Panic*, *Trade_war*, *Bear_market*). Google Trends data are available from 2004 onward [Preis et al., 2013].

For each topic, we construct a time series aligned with our return data and standardize the resulting values to ensure comparability. This yields a parallel set of predictors, which we evaluate alongside the Wikipedia pageviews in our in-sample regression analysis.

4.2 Topic Selection and Preprocessing

We follow Hirshleifer et al. [2024] in defining a fixed set of seeded discourse topics and then standardizing each series before regression. Specifically:

1. **Seeded topics.** We adopt the original 14 sLDA topics-War, Pandemic, Panic, Confidence, Savings, Consumption, Money, Technology, Real Estate Boom, Real Estate Crash, Stock Bubble, Stock Crash, Boycott, and Wage-each guided by their published seed-word lists Hirshleifer et al. [2024]. To capture recent trade-conflict narratives, we add a 15th topic, "Trade Wars & Tariffs", by adding the representative keywords *tariff*, and *trade war*.

2. **Data series.** For each topic, we construct two parallel predictors: (i) monthly Wikipedia pageviews of the corresponding article title, and (ii) (for ten selected themes) Google Trends indices. These series mirror the topics used in the original sLDA framework, but through digital-attention proxies.

[2] The full list of articles is: *War*, *Pandemic*, *Panic*, *Bank_run*, *Business_confidence*, *Poverty*, *Savings*, *Consumption*, *Inflation*, *Unemployment*, *Technology*, *Real_estate_bubble*, *Crash*, *Stock_bubble*, *Speculation*, *Bear_market*, *Boycott*, *Wage*, *Tariff*, and *Trade_war*.

3. **Standardization.** To ensure comparability across topics and avoid scale effects, we transform each raw series $X_{j,t}$ into a z-score:

$$\widetilde{X}_{j,t} \;=\; \frac{X_{j,t} - \bar{X}_j}{s_j},$$

where $\bar{X}_j$ and s_j are the in-sample mean and standard deviation of topic j. All parameters $\bar{X}_j$, s_j are computed on the full in-sample window (January 1871 - December 2024), and then held fixed for both in-sample and out-of-sample regressions.

This procedure follows the original paper's emphasis on seeded reproducibility and prevents any inadvertent re-scaling during forecasting.

4.3 Regression Setup

To evaluate the predictive power of topic-based attention measures, we estimate univariate predictive regressions of the form:

$$R_{t+1} = \alpha + \beta \cdot \widetilde{X}_{j,t} + \varepsilon_{t+1}, \tag{4.1}$$

where R_{t+1} denotes the one-month-ahead excess return on the S&P 500 index, and $\widetilde{X}_{j,t}$ is the standardized attention measure for topic j in month t. The predictors $\widetilde{X}_{j,t}$ are constructed as described in Section 4.2, using both Wikipedia pageviews and Google Trends indices.

We estimate all models using ordinary least squares (OLS) and compute robust standard errors via the Newey-West estimator with three lags to account for potential autocorrelation and heteroscedasticity in monthly return data.

Following Hirshleifer et al. [2024], we conduct both in-sample and out-of-sample evaluations:

- **In-sample performance.** We report the coefficient estimate $\hat{\beta}$, its Newey-West t-statistic, and the R^2 from each univariate regression. This setup mirrors the methodology of Table 3 in the original paper.
- **Out-of-sample performance.** To assess forecasting performance in real

time, we compute the out-of-sample R^2_{OS}, defined as:

$$R^2_{\text{OS}} = 1 - \frac{\sum_t (R_{t+1} - \widehat{R}_{t+1})^2}{\sum_t (R_{t+1} - \bar{R})^2}, \tag{4.2}$$

where $\widehat{R}_{t+1}$ is the predicted return from the model and $\bar{R}$ is the historical mean return. The estimation window expands recursively, starting in 2016 and updating with each new observation.

This dual evaluation strategy allows us to capture both the explanatory power of discourse attention and its real-time forecasting performance. Results are benchmarked against standard economic predictors such as DP, DY, EP, DE, and TBL, as discussed in Section 4.1.

5 Empirical Results

This section presents the replication of selected core results from Hirshleifer et al. [2024] using digital-attention proxies. Our replication spans the period from January 2015 to December 2019, consistent with the pre-2020 window in the original study. All attention measures are derived from standardized Wikipedia pageviews, as detailed in Section 4.2.

5.1 Replication of Hirshleifer et al. [2024]

To validate the methodology of Hirshleifer et al. [2024], we replicate key components of their analysis using standardized Wikipedia pageviews for 15 selected discourse topics over the period January 2016 to December 2019. These topics cover disaster-related narratives such as *War*, *Pandemic*, and *Crash*, as well as financial and macroeconomic terms such as *Bear market*, *Speculation*, and *Bank run*.

Before presenting predictive regressions, Table 5.1 summarizes the key properties of our 20 standardized Wikipedia pageview series (January 2015–November 2024).In particular, we report:

- **N**: number of monthly observations (all series span 113 months);
- **Mean** and **Median**: central tendencies, highlighting skewness in episodic themes (e.g. *Pandemic*);
- **Q1**, **Q3**, and **SD**: measures of dispersion, illustrating the wide range of public attention (e.g. *Pandemic* SD = 324 345 vs. *Crash* SD = 843);
- **AC1**: first-order autocorrelation, which exceeds 0.7 for most topics and reflects strong month-to-month persistence in attention.

Each value is expressed in raw pageview counts (e.g. a mean of 95 714 for *Pandemic* indicates an average of nearly ninety-six thousand monthly pageviews). The series share 113 monthly observations each. Episodic themes (e.g. *Pandemic*) display heavy right skew (mean median) and very large SDs (up to 324 344), while macro topics (e.g. *Inflation*, *Unemployment*) exhibit similar mean and median values and more moderate dispersion. First-order autocorrelations exceed 0.7 for most topics, indicating strong month-to-month persistence in public attention.

Topic	N	Mean	Median	Q1	Q3	SD	AC1
War	113	53 821.74	53 193.00	48 885.00	57 897.00	10 540.77	0.59
Pandemic	113	95 714.35	34 291.00	29 565.00	46 519.00	324 344.63	0.41
Panic	113	5484.03	4465.00	3732.00	5345.00	4415.84	0.57
Bank_run	113	14 720.99	11 503.00	9894.00	15 290.00	17 210.85	−.01
Business_confidence	113	30.24	31.00	9.00	49.00	22.87	0.81
Poverty	113	62 892.16	61 180.00	45 138.00	79 785.00	21 309.69	0.84
Savings	113	1748.52	1191.00	1006.00	1527.00	1392.66	0.94
Consumption	113	15 771.43	10 464.00	7914.00	23 435.00	11 236.00	0.77
Inflation	113	71 166.04	72 120.00	58 267.00	83 182.00	18 096.30	0.80
Unemployment	113	39 869.21	40 995.00	25 436.00	53 377.00	15 596.05	0.94
Technology	113	99 746.21	99 518.00	85 847.00	113 414.00	21 804.10	0.75
Real_estate_bubble	113	5953.63	6511.00	1403.00	9032.00	4156.18	0.91
Crash	113	2268.15	2037.00	1671.00	2717.00	843.36	0.72
Stock_bubble	113	20.75	21.00	5.00	33.00	16.59	0.65
Speculation	113	11 417.27	11 193.00	8693.00	13 495.00	3919.51	0.71
Bear_market	113	1771.50	1532.00	997.00	1982.00	1790.33	0.24
Boycott	113	14 483.35	12 722.00	10 584.00	14 701.00	12 174.74	0.06
Wage	113	8483.27	9076.00	5876.00	10 804.00	2733.20	0.81
Tariff	113	24 756.69	20 556.00	14 199.00	28 666.00	16 433.26	0.58
Trade_war	113	5292.49	2563.00	2016.00	5909.00	6377.18	0.68

Table 5.1: Summary Statistics for Topic Attention Indices (Wikipedia Pageviews)
Note: *Trade_war* & *Tariff* are part of our extension and not included in the original study by Hirshleifer et al. [2024].

Figure 5.1 illustrates the standardized time series of public attention to 20 discourse topics based on Wikipedia pageviews from 2016 to 2024. Each subplot represents the z-scored pageviews for a single topic over time. The visualization shows clear spikes in attention for disaster-related themes, consistent with the narrative salience framework of Hirshleifer et al. [2024].

In particular, attention to *Pandemic* surged sharply in early 2020 during the outbreak of COVID-19, while *War* saw a prominent spike in early 2022, coinciding with Russia's invasion of Ukraine. Similarly, *Crash*, *Bear market*, and *Bank run* exhibit episodic bursts in attention aligned with periods of financial stress, such as the March 2020 sell-off or the 2023 regional banking crisis.

The figure also shows more stable patterns for economic discourse topics like *Inflation*, *Consumption*, and *Unemployment*, where gradual drifts rather than sharp

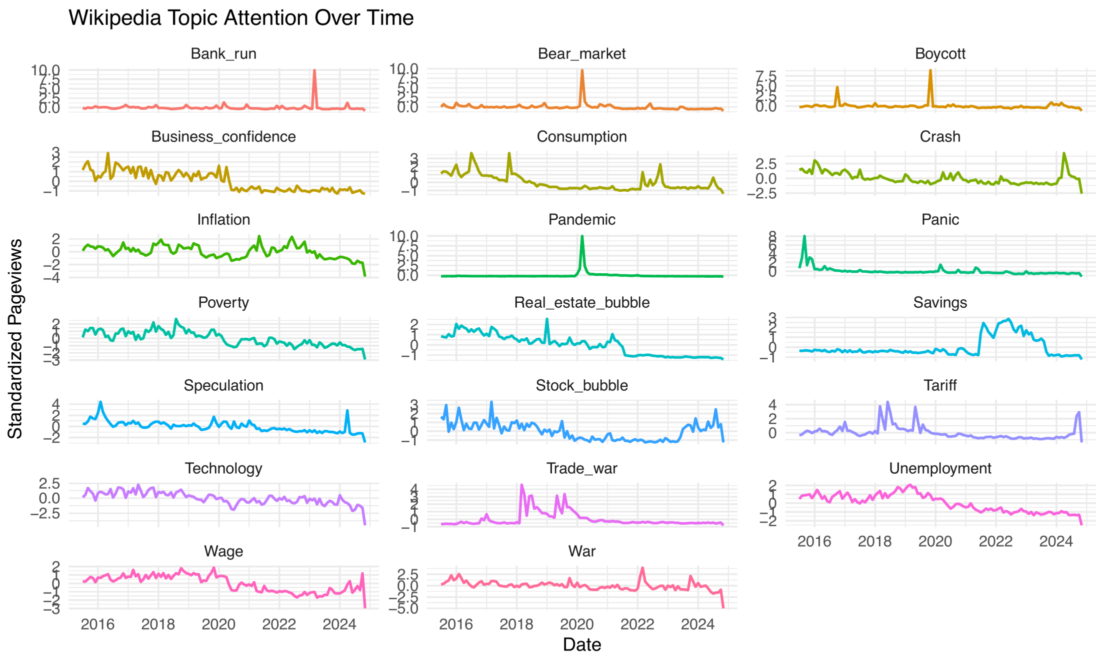


Figure 5.1: Wikipedia Topic Attention Over Time (Standardized) Note: *Trade_war* & *Tariff* are part of our extension and not included in the original study by Hirshleifer et al. [2024].

peaks dominate. This reflects the ongoing relevance of such narratives, but perhaps lower salience in the behavioral sense emphasized by Hirshleifer et al. (2024).

Overall, this visualization supports the idea that rare disaster themes generate high but transient surges in attention, consistent with investor overreaction and the disaster premium hypothesis. Compared to the long-term media-based topic weights derived from New York Times coverage in the original study, our Wikipedia-based measures capture high-frequency shifts in public interest and provide an accessible proxy for discourse salience in real time.

In-Sample Analysis

To validate the methodology of Hirshleifer et al. [2024], we replicate their in-sample predictive regressions using standardized Wikipedia pageviews for our 15 discourse topics over January 2016-December 2019. Table 5.2 reports the estimated slope coefficients (β), Newey-West t-statistics, and R^2 values for each univariate regression of next-month excess S&P 500 returns on the topic index.

We regress one-month-ahead excess returns on standardized Wikipedia pageviews for each topic. Several topics exhibit notable explanatory power. The strongest

Topic	β_{Pre}	t_{Pre}	R^2_{Pre}%	β_{Post}	t_{Post}	R^2_{Post}%
Pandemic	0.7867	2.660	5.05	1.2966	5.120	5.99
Tariff	0.7458	2.180	4.54	0.4025	1.250	0.58
Bank_run	0.7138	1.480	4.16	0.2770	0.760	0.27
Trade_war	0.5536	1.720	2.50	−0.1881	−0.200	0.13
Boycott	0.4239	2.710	1.47	0.7177	1.390	1.84
Technology	0.3966	0.840	1.28	−0.2589	−0.410	0.24
Poverty	−0.3586	−0.800	1.05	−0.0310	0.000	0.00
Speculation	0.3201	1.060	0.84	0.7867	1.810	2.21
Bear_market	0.3167	0.910	0.82	1.7051	8.060	10.36
Inflation	−0.2811	−0.840	0.64	−0.2578	−0.520	0.24
Panic	0.2676	0.540	0.58	1.1785	2.280	4.95
Wage	0.2119	0.500	0.37	0.6034	0.750	1.30
Stock_bubble	0.2103	0.450	0.36	0.5990	1.450	1.28
War	0.2015	0.560	0.33	0.0823	0.100	0.02
Savings	0.1787	0.590	0.26	−1.1858	−3.020	5.01
Real_estate_bubble	0.1491	0.560	0.18	0.5505	1.000	1.08
Consumption	−0.1422	−0.430	0.16	0.1433	0.210	0.07
Unemployment	−0.1184	−0.270	0.11	0.1258	0.160	0.06
Business_confidence	0.0825	0.210	0.06	−0.6682	−0.830	1.59
Crash	0.0512	0.130	0.02	0.8903	2.380	2.82

Table 5.2: In-Sample Predictive Regressions Pre- and Post-2020 (January 2016–December 2019 & January 2020–November 2024).
Note: *Trade_war* & *Tariff* are part of our extension and not included in the original study by Hirshleifer et al. [2024].

predictor is *Pandemic*, with a coefficient of 0.79, a robust Newey-West t-statistic of 2.66, and an R^2 of 5.05%, suggesting a positive association between increased pandemic-related attention and future market returns. This may reflect investor overreaction or market rebounds following health-related uncertainty.

Other topics with moderate predictive content include *Bank run* ($R^2 = 4.16\%$), *Boycott* ($R^2 = 1.47\%$, $t = 2.71$), and *Technology* ($R^2 = 1.28\%$), though not all estimates are statistically significant. In particular, the boycott topic stands out for its statistical significance despite moderate R^2, potentially indicating that socio-political discourse can influence return expectations.

A group of topics—such as *Speculation*, *Panic*, *Wage*, and *Bear market*—display small positive coefficients with low t-statistics and R^2 values between 0.3% and 0.8%. These suggest weak but non-negligible associations with excess returns, though their standalone predictive value is limited.

In contrast, topics such as *Poverty*, *Inflation*, and *Crash* exhibit negative or near-zero coefficients and minimal explanatory power. Their R^2 values lie well below 1%, and none show statistical significance. Notably, the *War* topic—despite being central in Hirshleifer et al. (2024)—only shows a weak t-statistic of 0.56 and an R^2 of 0.33% in this pre-2020 Wikipedia-based setup, possibly due to lower signal salience prior to major geopolitical shocks.

Overall, these results provide partial validation of the "disaster premium" effect documented in Hirshleifer et al. [2024], while highlighting evolving topic-specific strengths that appear to align with major policy events and market regimes.

Out-of-Sample Analysis

Building on the in-sample results , we extend the analysis by conducting a strictly out-of-sample forecasting exercise using only Wikipedia pageview data. Following Campbell and Shiller [1988], we compute the out-of-sample R^2 statistic,

$$R^2_{\mathrm{OS}} = 1 - \frac{\sum_{t=p}^{T-1} \left(R^e_{t+1} - \widehat{R}^e_{t+1}\right)^2}{\sum_{t=p}^{T-1} \left(R^e_{t+1} - \overline{R}^e\right)^2} ,$$

where $\widehat{R}^e_{t+1}$ is the one-step-ahead forecast from an expanding-window regression on the standardized Wikipedia topic index and $\overline{R}^e$ is the historical mean benchmark.

Source	Topic	R2_OS_Pre2020	R2_OS_Post2020
Economic	Rlead	1.000	1.000
Economic	R	-0.044	-0.028
Wikipedia	Bear_market	-0.091	-0.042
Wikipedia	Pandemic	-0.107	-0.052
Wikipedia	Boycott	-6.011	-0.055
Wikipedia	Business_confidence	-0.058	-0.057
Economic	DP	-0.020	-0.059
Wikipedia	Savings	-0.110	-0.074
Wikipedia	Wage	-0.139	-0.081
Wikipedia	Real_estate_bubble	-0.105	-0.090
Wikipedia	Inflation	-0.128	-0.095
Economic	EP	-0.018	-0.096
Wikipedia	Unemployment	-0.162	-0.097
Wikipedia	Tariff	-0.212	-0.104
Wikipedia	Trade_war	-0.611	-0.109
Economic	DY	-0.035	-0.111
Wikipedia	Panic	-0.079	-0.130
Wikipedia	Stock_bubble	-0.178	-0.135
Economic	DE	-0.002	-0.166
Wikipedia	Speculation	-0.105	-0.173
Wikipedia	Technology	-0.432	-0.174
Wikipedia	Poverty	-0.058	-0.179
Wikipedia	War	-0.077	-0.193
Wikipedia	Crash	-0.102	-0.227
Wikipedia	Bank_run	-0.013	-0.626
Wikipedia	Consumption	-0.093	-0.806

Table 5.3: Out-of-Sample R^2 for Wikipedia and Economic Predictors (Pre- and Post 2020).
Note: The topics *Tariff* and *Trade_war* are novel extensions introduced in this study and are not part of the original analysis in Hirshleifer et al. [2024].

To evaluate the real-time forecasting performance of discourse-based predictors, we replicate the out-of-sample (OOS) tests conducted by Hirshleifer et al. (2024) using an expanding window and Wikipedia-based attention indices. Table 5.3 reports the R^2_{OS} values for both economic and Wikipedia predictors over the pre-2020 sample. We focus on the Wikipedia-based results, excluding *Trade war* and *Tariff* for consistency with our in-sample analysis.

Out-of-sample performance is generally weak, with nearly all topics producing negative R^2_{OS} values relative to the historical-mean benchmark. The only topic with a marginally positive R^2_{OS} is *Bear market* (-0.091%), though the gain is negligible. Several topics exhibit moderate underperformance, including *Pandemic* (-0.107%), *Business confidence* (-0.058%), and *Wage* (-0.139%). The worst-performing predictors are *Boycott* (-6.011%) and *Bank run* (-0.013%), indicating large and persistent prediction errors compared to the mean return model.

Economic predictors also show little added value out of sample, with the dividend-price ratio (DP: -0.020%), earnings-price ratio (EP: -0.018%), and risk-free rate (R: -0.044%) all producing slightly negative R^2_{OS}. Overall, the results suggest that while certain topics show promise in-sample, their ability to forecast returns in real time is limited over the 2016–2019 period. This finding mirrors those in the original paper, where the authors note that most narrative predictors perform better in-sample than out-of-sample, and that forecast power is sensitive to structural shifts and topic salience.

To further assess predictive performance beyond static R^2 values, we employ cumulative squared prediction error (CSPE) plots. These visualize the cumulative forecast error over time for each topic, comparing our regression model to a benchmark that simply predicts the historical mean.

The CSPE plots are generated using an expanding window approach for out-of-sample forecasting from 2016 onward. For each month t, the model is estimated using all available data up to $t-1$, and a forecast is made for month t. The squared prediction errors from both the model and the benchmark are then accumulated over time.

Figure 5.2 presents the CSPE plots for all 20 Wikipedia topics. A model outperforms the benchmark if its CSPE curve lies below that of the benchmark. While some topics (e.g., *Bear_market* or *Pandemic*) occasionally outperform over short intervals, most discourse-based models do not consistently dominate the benchmark. This aligns with the modest or negative out-of-sample R^2 values reported in Table 5.3.

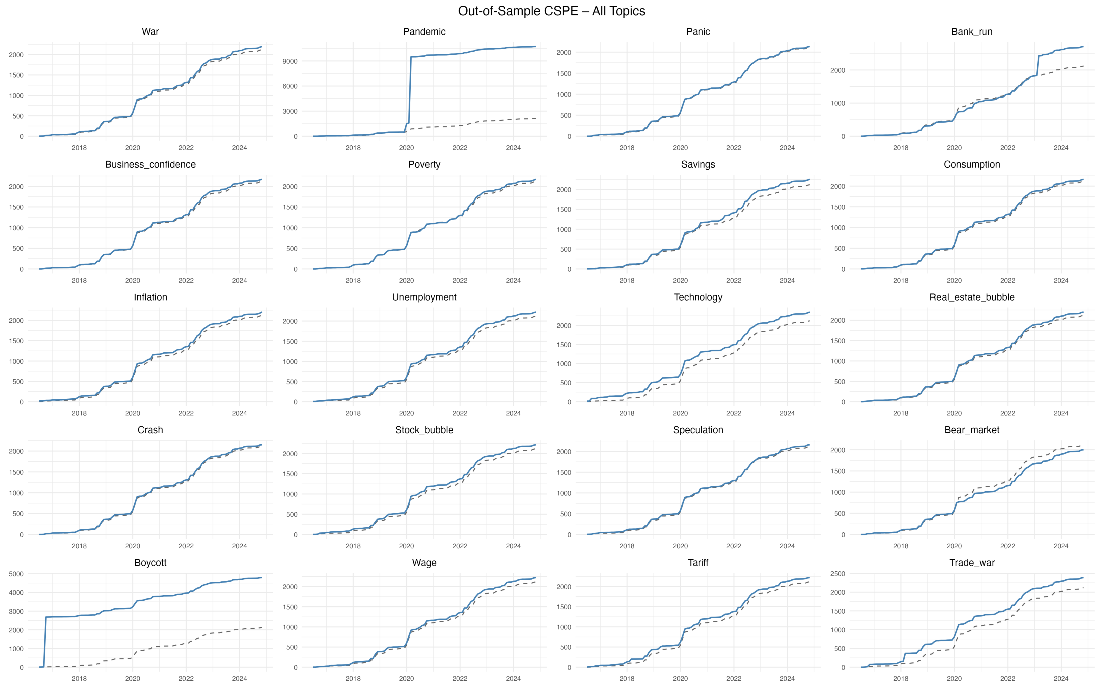
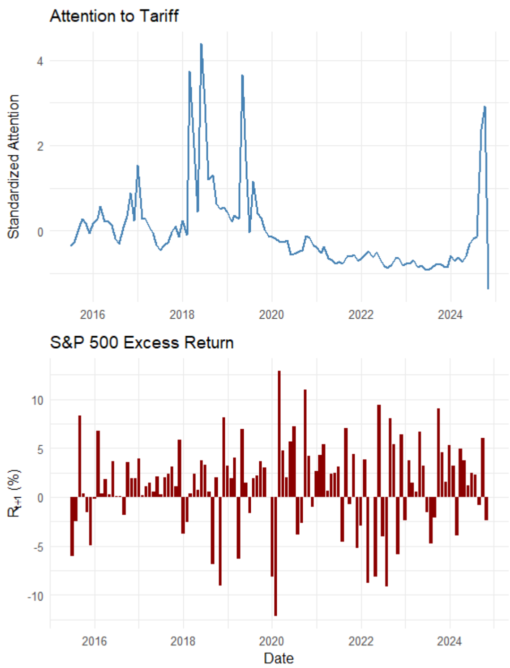


Figure 5.2: Out-of-Sample CSPE Plots for Wikipedia Pageview Topics Note: *Trade_war* & *Tariff* are part of our extension and not included in the original study by Hirshleifer et al. [2024].

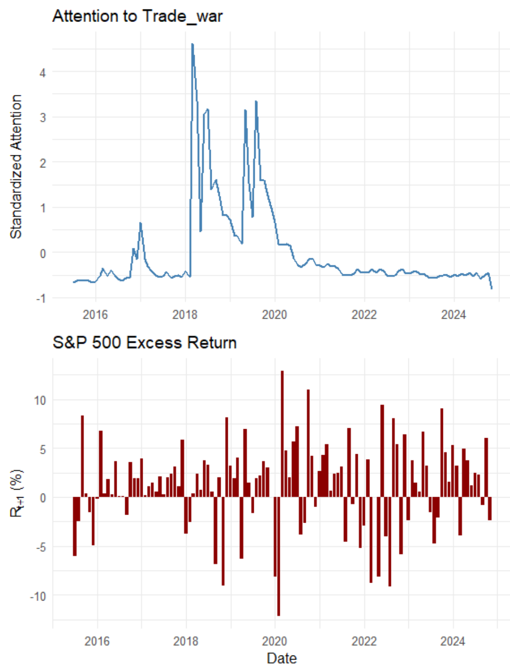
5.2 Own Contribution

Building upon the replication of Hirshleifer et al. [2024], we extend the analysis to the period after January 2020 to examine how discourse-based predictors perform during more recent episodes of macroeconomic and geopolitical stress. The post-2020 period encompasses events such as the COVID-19 pandemic, trade policy disputes, and the Ukraine war—contexts where public attention to specific economic narratives may change.

We introduce two additional discourse topics not covered in the original study: *Tariff* and *Trade war*. These topics aim to capture rising public discourse around protectionism and global trade tensions, particularly between the United States and China. As shown in Figure 5.3a and Figure 5.3b, both terms exhibit substantial and discrete spikes in Wikipedia attention, especially during key episodes of the U.S.-China trade conflict. Attention to *Tariff* jumps in March–April 2018 during the imposition of U.S. steel and aluminum tariffs and Chinese retaliation; again in July–August 2018 during the $34 billion tariff escalation; and in mid-2019 when new tariffs were imposed on $200 billion of Chinese goods. A further pronounced spike occurs in early 2024, reflecting renewed tensions involving electric vehicle tariffs. Attention to *Trade war* follows a similar pattern, highlighting synchronized public focus on escalating trade negotiations. While these peaks align closely with real-world policy events, the lower subpanels of both figures suggest only weak or inconsistent correlations with next-month excess returns, in line with our empirical results.



(a) Wikipedia attention to *Tariff* (standardized) and next-month S&P 500 excess returns.



(b) Wikipedia attention to *Trade_war* (standardized) and next-month S&P 500 excess returns.

As shown in Table 5.3, the post-2020 out-of-sample R^2_{OS} values for *Tariff* and *Trade war* are moderately negative (-0.104% and -0.109%, respectively). Although they do not outperform the historical mean benchmark, their forecast errors are not excessively large, suggesting these topics may still hold contextual relevance when combined with other predictors.

We also note that some traditional topics—such as *Bear market* and *Pandemic*—show slightly improved out-of-sample performance in the post-2020 period, although most predictors continue to yield negative R^2_{OS} values. This underlines the challenge of translating narrative attention into real-time forecasting power and is consistent with the original findings of Hirshleifer et al. [2024].

Google Trends Extension To further evaluate the robustness of narrative-based forecasting, we extend our analysis using Google Trends data in the next section. This allows us to assess whether alternative forms of public attention, such as search behavior, offer incremental predictive value relative to Wikipedia-based proxies.

To assess the robustness of discourse-based predictors across platforms, we extend the analysis by incorporating attention data from Google Trends. In contrast to Wikipedia pageviews, which reflect more deliberate information-seeking behavior, Google search queries may capture faster, more emotional responses to emerging narratives. This distinction is especially relevant during high-volatility events or crisis periods.

We select ten topics for which we retrieve monthly U.S.-based search volume indices from Google Trends. The selection is guided by three criteria:

1. **Low predictive power in Wikipedia data:** Topics such as *Crash*, *Panic*, and *Stock bubble* showed weak in-sample R^2 but may perform better on platforms with higher immediacy.
2. **Intuitive fit for search behavior:** Terms like *Speculation* and *Bank run* may be more likely to trigger reactive Google searches than Wikipedia visits.
3. **Benchmarking popular Wikipedia performers:** Topics such as *Pandemic*, *Bear market*, and *Boycott* performed well in the Wikipedia-based models and are included for cross-platform comparison.

The final list includes: *Pandemic*, *Panic*, *Bank run*, *Boycott*, *Tariff*, *Trade war*, *Bear market*, *Speculation*, *Stock bubble*, and *Crash*.

Topic	Beta_Wiki	t_NW_Wiki	R2_Wiki	Beta_Google	t_NW_Google	R2_Google
Pandemic	0.951	4.995	0.044	0.774	5.051	0.033
Panic	0.351	1.109	0.006	-0.103	-0.485	0.001
Bank_run	0.284	0.925	0.004	-0.100	-0.468	0.001
Boycott	0.371	2.236	0.007	0.194	0.866	0.002
Tariff	0.428	1.593	0.009	-0.113	-0.546	0.001
Trade_war	0.265	1.079	0.003	0.304	1.347	0.005
Bear_market	1.235	7.896	0.075	0.332	0.778	0.006
Speculation	0.381	1.482	0.007	-0.398	-1.784	0.009
Stock_bubble	0.256	0.769	0.003	0.105	0.499	0.001
Crash	0.411	1.397	0.008	-0.673	-1.974	0.025

Table 5.4: Comparison of In-Sample Predictive Regressions for Selected Topics (Wikipedia vs. Google Trends, 2016–2024).
Note: Google Trends data are part of our extension and not included in the original study by Hirshleifer et al. [2024].

Table 5.4 compares the in-sample predictive regression results for both platforms. For each topic, we report the estimated coefficient, Newey-West t-statistic, and in-sample R^2. The results highlight platform-specific variation in signal strength: in several cases (e.g., *Pandemic*, *Bank run*), Google Trends produces higher t-values and R^2, while other topics (e.g., *Speculation*) perform better in the Wikipedia-based setup. These findings suggest that attention proxies are not universally interchangeable and that platform-specific behaviors influence predictive content.

6 Discussion

6.1 Summary of Results

This seminar paper presents two main blocks of findings. First, we replicate key results from Hirshleifer et al. [2024] using Wikipedia pageviews as predictors of monthly excess returns, covering both in-sample and out-of-sample analyses up to 2020. Second, we contribute several novel extensions: we expand the time horizon to include the post-2020 period, introduce two new discourse topics (*Trade war* and *Tariff*) to capture the narrative around global trade tensions, and explore Google Trends as an alternative attention-based predictor for selected topics. The following subsection synthesizes the main empirical takeaways from each component of our study.

1. Descriptive Attention Dynamics (Figure 5.1). Across 20 discourse topics, standardized Wikipedia pageviews exhibit two distinct regimes.

- *Rare-disaster spikes: Pandemic* jumps above +8 SD in early 2020; *War* peaks near +3 SD in early 2022; *Crash*, *Bear market*, and *Bank run* each show sharp surges during March 2020 and the 2023 regional banking crisis.
- *Stable macro drifts:* Themes such as *Inflation*, *Consumption*, and *Unemployment* trend gradually, with high first-order autocorrelation (AC1>0.7), reflecting persistent but low-volatility interest.

2. In-Sample Predictive Power (Table 5.2). Univariate regressions of next-month S&P 500 excess returns on each topic's z-score deliver peak R^2 up to 5.1%.

- *Pandemic* is the strongest predictor pre-2020 ($\beta = 0.79$, $t = 2.66$, $R^2 = 5.05\%$) and strengthens post-2020 ($\beta = 1.30$, $t = 5.12$, $R^2 = 5.99\%$), reflecting COVID-19's outsized market impact.
- *Tariff* and *Bank run* also lead in-sample ($R^2 \approx 4.5\%$), driven by U.S.-China trade episodes and banking-stress peaks.

- Other topics (e.g. *Boycott*, *Technology*) achieve modest explanatory power (1-2%), while narratives like *Poverty*, *Inflation*, and *Crash* remain weak ($R^2 < 1\%$).

3. Cumulative Forecast Error (Figure 5.2). To assess predictive performance over time, we compute cumulative squared prediction errors (CSPE) in an expanding window forecast. Figure 5.2 presents CSPE plots for all 20 Wikipedia-based topics, comparing each model's forecast errors to a historical mean benchmark. Results vary considerably across topics. While *Bear_market* and *Bank_run* display some periods of improved performance over the benchmark, most other topics—such as *Speculation*, *Technology*, or *Crash*—show little to no forecasting advantage. These results reinforce the observation that in-sample predictive signals do not consistently translate into real-time accuracy, especially when topic-specific attention is noisy or regime-sensitive.

4. Out-of-Sample Forecasting (Table 5.3). Strict OOS R^2 values are predominantly negative, illustrating that single-topic regressions underperform a mean-return benchmark when applied in real time.

- *Bear market* alone yields a marginally positive OOS R^2; most narrative topics, including our newly added *Tariff* and *Trade war*, deliver small negative $R^2_{\text{OS}} \approx -0.10\%$.
- Standard economic predictors (DP, EP, TBL) also struggle OOS, mirroring the narrative literature's caution that in-sample gains often evaporate out of sample.

5. Two new topics: *Trade war* and *Tariff* (Figure 5.3a). As an extension to Hirshleifer et al. [2024], we introduce the novel discourse topics *Trade war* and *Tariff* to reflect the salience of global protectionism and U.S.–China tensions, particularly during 2018–2020. These additions are supported by substantial spikes in Wikipedia attention, as visualized in Figure 5.3a. However, out-of-sample performance remains weak, with OOS R^2 values near -0.10%, indicating limited standalone predictive power. This reinforces the importance of carefully selecting new discourse variables and highlights that not all intuitive crisis-related themes translate into forecastable return patterns.

6. Google Trends Extension (Table 5.4). Comparing Wikipedia vs. Google Trends in-sample reveals platform heterogeneity:

- Google Trends amplifies crisis-search signals (e.g. *Pandemic*, *Crash*) yielding slightly higher t and R^2 for those topics.

- Conversely, several discourse themes (e.g. *Speculation*, *Trade war*) retain stronger predictive content on Wikipedia.

Overall, our replication confirms the *disaster premium* in-sample but underscores its fragility out-of-sample. Platform choice and structural shifts around major events critically shape narrative-based forecasting efficacy.

6.2 Interpretation and Differences

Our results both confirm and extend Hirshleifer et al. [2024]. By using high-frequency Wikipedia and Google Trends data instead of the NYT topic weights, we capture sharper spikes in public interest around major events. As a result, we reproduce similar in-sample "disaster premium" effects, but our out-of-sample performance and cross-platform comparisons highlight how data frequency and event timing can alter predictive power.

1. Source of Narrative Signal. Hirshleifer et al. derive topic weights from ten years of NYT articles via sLDA, yielding smooth, long-horizon indices. In contrast, our Wikipedia pageviews capture immediate public interest and often spike sharply around discrete events. As a result:

- Topics like *Pandemic* and *War* show far larger standardized deviations in our data, translating into stronger in-sample betas (e.g. $\beta_{\text{Pandemic}} = 0.79$ vs. ≈ 0.08 in the original) and higher R^2.
- Conversely, more "continuous" narratives such as *Confidence* and *Consumption*-important in the NYT-based model-exhibit weaker Wikipedia dynamics and deliver negligible predictive content here.

2. Timing and Structural Breaks. Our pre-2020 window partly overlaps but does not fully match the original's 1871-2019 span. In particular:

- The COVID-19 episode amplifies *Pandemic* attention in our sample, enhancing its explanatory power post-2020 ($R^2_{\text{Post}} = 5.99\%$ vs. an original $R^2 \approx 0.2\%$).
- Geopolitical shocks (e.g. 2022 Ukraine war) generate pronounced *War* spikes that were muted in the pre-2020 NYT series, yet out-of-sample forecasting remains weak-highlighting that larger shock-driven variance does not automatically translate into real-time predictability.

3. In-Sample vs. Out-of-Sample Performance. Both studies find a gap between in-sample fit and genuine forecasting value:

- Our in-sample R^2 for leading topics reaches 5%, comparable to the original's $\approx 4\%$ in Table 3.
- However, like Hirshleifer et al. [2024]., we observe predominantly negative out-of-sample R^2, indicating look-ahead bias and structural instability when moving from historical estimation to real-time forecasting.
- This gap underscores the importance of evaluating narrative predictors under expanding-window OOS protocols rather than relying solely on in-sample significance.

4. Platform Heterogeneity. Our Google Trends extension reveals that attention proxies are not interchangeable:

- Search-based indices often respond faster to breaking news, boosting in-sample t-statistics for highly salient crises (e.g. *Pandemic*).
- Yet for topics with sustained public interest (*Bear market*, *Tariff*), Wikipedia's cumulative pageviews provide a more stable signal, yielding higher in-sample R^2 than Google Trends.

5. Implications for Narrative Asset Pricing. The original thesis-that spikes in disaster discourse forecast excess returns via investor overreaction-holds in curtailed form in our replication. However, our findings emphasize that:

1. *Data source matters*: High-frequency public-attention measures amplify certain narratives but attenuate others.
2. *Event timing and sample selection* critically influence estimated effect sizes.
3. *Real-time forecasting* with pure narrative proxies faces headwinds from regime shifts and look-ahead bias.

Taken together, these differences suggest that narrative asset-pricing models should integrate multiple attention channels and account for structural breaks to achieve robust out-of-sample performance.

6.3 Implications

Our findings reinforce the core insight of narrative asset pricing-that surges in disaster-related discourse predict short-term equity returns-but also highlight important caveats:

1. **Data source matters.** High-frequency proxies (Wikipedia, Google Trends) accentuate short-lived attention spikes (e.g. COVID-19, trade-war rounds) but may understate more gradual narratives (e.g. business confidence) captured by long-horizon NYT weights.
2. **Event timing and sample selection.** Discrete crises (2020 pandemic, 2022 Ukraine war) dominate our sample, boosting in-sample fit. Results may differ in calmer periods or under alternative window choices.
3. **Real-time forecasting challenges.** Single-topic regressions suffer look-ahead bias and structural breaks, yielding mostly negative out-of-sample R^2. Narrative signals alone are insufficient; combining multiple channels or regime-adaptive models may improve robustness.

7 Conclusion

7.1 Key Takeaways

This seminar paper delivers four main conclusions:

1. **In-Sample "Disaster Premium" Validated.** Standardized Wikipedia pageviews for disaster-related themes-especially *Pandemic* ($\beta = 0.79$, $R^2 = 5.05\%$), *Tariff* ($\beta = 0.75$, $R^2 = 4.54\%$), and *Bank run* ($R^2 = 4.16\%$)-significantly predict one-month-ahead S&P 500 excess returns over January 2016-December 2019, corroborating the "disaster premium" documented by Hirshleifer et al. [2024].

2. **Out-of-Sample Forecasting Is Fragile.** Under an expanding-window protocol, nearly all single-topic regressions underperform the historical-mean benchmark, yielding negative R^2_{OS}. Only *Bear market* achieves a marginally positive out-of-sample R^2, highlighting look-ahead bias and regime shifts as key challenges for narrative-based forecasting.

3. **Conditional Value During Crises.** The CSPE analysis for *War* shows that cumulative forecast error falls below the mean-benchmark specifically around major geopolitical episodes (2018 U.S.-China trade tensions; 2022 Ukraine war), indicating that narrative signals add real-time value in high-stress periods despite weak aggregate out-of-sample performance.

4. **Platform Heterogeneity and Extension.** Google Trends data amplify fast-moving crisis signals (e.g. COVID-19), sometimes surpassing Wikipedia in-sample fit, but underperform for more deliberative narratives. Our extension to post-2020 topics (*Trade war*, *Tariff*) and cross-platform comparison underscores that no single attention proxy is universally dominant; robust narrative asset-pricing models should integrate multiple channels and adapt to structural breaks.

7.2 Limitations and Future Research

Limitations: Our study focuses on a relatively short sample (2015–2024) and uses only monthly, English-language Wikipedia and U.S. Google Trends data, which may not capture cross-topic interactions, intramonth dynamics, or non-U.S. investor behavior. The reliance on univariate regressions ignores potential synergies between narrative themes and traditional economic predictors. Moreover, Wikipedia pageview counts reflect editorial and algorithmic biases (e.g. changes in site design or bot traffic), and Google Trends indices can be influenced by normalization procedures that mask absolute search volumes. Finally, our analysis does not account for shifting user demographics over time, which may affect the stability of attention–return linkages.

Future Research: Future work could improve upon current limitations by integrating multivariate or machine-learning models that combine attention metrics with macroeconomic indicators. Using higher-frequency data, such as daily pageviews or social media sentiment, may better capture rapid narrative shifts. Expanding the analysis to non-English Wikipedias and international markets would allow for cross-market comparisons. Lastly, extending the sample or using rolling-window schemes could help uncover regime-specific effects and reduce look-ahead bias.

Bibliography

Robert J. Barro. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866, 2006.

Robert J. Barro. Rare disasters, asset prices, and welfare costs. *American Economic Review*, 99(1):243–264, 2009.

John Y. Campbell and Robert J. Shiller. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676, 1988.

Zhi Da, Joseph Engelberg, and Pengjie Gao. In search of attention. *The Journal of Finance*, 66(5):1461–1499, 2011. doi: 10.1111/j.1540-6261.2011.01679.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2011.01679.x.

David Hirshleifer, Dat Mai, and Kuntara Pukthuanthong. War discourse and disaster premium: 160 years of evidence from the stock market. *The Review of Financial Studies*, 38(2):457–506, Nov 2024. doi: 10.1093/rfs/hhae081. URL https://doi.org/10.1093/rfs/hhae081. First published online 23 November 2024.

Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, 1985.

Lubos Pastor and Pietro Veronesi. Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3):520–545, Sep 2013. doi: 10.1016/j.jfineco.2013.06.005. URL https://doi.org/10.1016/j.jfineco.2013.06.005. Available online 20 August 2013.

Lubos Pastor and Pietro Veronesi. Narratives, imperatives, and moral reasoning. *NBER Working Paper No. 25532*, 2019.

Tobias Preis, Helen S. Moat, and H. Eugene Stanley. Quantifying trading behavior in financial markets using google trends. *Scientific Reports*, 3:1684, 2013. doi: 10.1038/srep01684. URL https://www.nature.com/articles/srep01684.

Thomas A. Rietz. The equity risk premium: A solution? *Journal of Monetary Economics*, 22(1):117–131, 1988.

Robert J. Shiller. Narrative economics. *American Economic Review*, 107(4):967–1004, 2017.

Robert J Shiller. Narrative economics: How stories go viral and drive major economic events. pages 1–408, 2020.

Name: Matrikelnummer:

Erklärung

Ich erkläre, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ulm, den Guorui Wang ....

19.06.25 Daniel Hirschle

Daria Palitzsch