Interpretable Systematic Risk around the Clock*

Songrun He

First draft July 2025

Abstract

In this paper, I present the first comprehensive, around-the-clock analysis of systematic jump risk by combining high-frequency market data with contemporaneous news narratives identified as the underlying causes of market jumps. These narratives are retrieved and classified using a state-of-the-art open-source reasoning LLM. Decomposing market risk into interpretable jump categories reveals significant heterogeneity in risk premia, with macroeconomic news commanding the largest and most persistent premium. Leveraging this insight, I construct an annually rebalanced real-time strategy that hedges the most priced jump risk, achieving an out-of-sample Sharpe ratio of 0.84 and delivering significant alphas relative to standard factor models. The results highlight the value of around-the-clock analysis and LLM-based narrative understanding for identifying and managing priced risks in real time.

JEL Classification: C58, G11, G12, G14

Keywords: Systematic jump risk, High-frequency data, Large language models, Fama-MacBeth regression

*Songrun He is at Washington University in St. Louis (h.songrun@wustl.edu). I am indebted to my advisors, Asaf Manela (co-chair) and Guofu Zhou (co-chair), for their invaluable guidance, help, and support. I am also very grateful to Kai Li and Linying Lv for their very helpful comments.

1 Introduction

Understanding the sources and pricing of aggregate market systematic risk is one central question in financial economics. There has been significant progress in using textual analysis to better understand systematic risk and ex-ante compensation (Manela and Moreira, 2017; Bybee et al., 2023, 2024). One major feature and component of systematic risk is that it features large jumps, which typically result from major realizations of news events, and jumps are a continuous-time construct and therefore can only be identified using high-frequency data.

Aleti and Bollerslev (2025) provide the first high-frequency analysis of news events driving the high-frequency systematic jumps during the intraday trading period from 9:30 a.m. to 4:00 p.m. However, focusing only on the intraday period might suffer from an omitted variable bias, where a significant component of systematic risk materializes during the overnight period (Glasserman et al., 2025). What remains missing in the literature is a comprehensive, around-the-clock analysis that captures all systematic jump events and their associated news drivers. Such a holistic approach is essential for forming an unbiased and complete understanding of systematic jump risk and its ex-ante pricing.

I combine three recent advances to provide the first comprehensive analysis of all systematic jump events linked to contemporaneous real-time high-frequency news text from Dow Jones Newswire in the U.S. equity market. First, I exploit around-the-clock high-frequency data on both the cash equity market and the *S&P 500 E-mini* futures, achieving nearly 24-hour coverage of the U.S. market over 1997–2020.

Second, I adapt the continuous-time Fama–MacBeth regression of Aït-Sahalia, Jacod, and Xiu (2025) to decompose systematic risk into continuous and *topic-specific* jump components, constructing "pure-play" hedging portfolios whose betas isolate each component using a large panel of high-frequency S&P 1500 stock returns.

Third, I harness state-of-the-art open-source reasoning LLM *DeepSeek-R1* to retrieve contemporaneous high-frequency news narratives triggering each jump and assign each jump to one of the five mutually exclusive economic topics identified by the LLM: macroeconomic news, corporate bellwethers, international spillovers, policy announcements, and geopolitical events.

The new and more advanced analytical tools, along with the comprehensive analytical framework, yield several new findings.

Firstly, there is significant heterogeneity in risk premia in jumps belonging to different economic categories. Unlike Aleti and Bollerslev (2025), who uncovers the monetary policy as the most important component for risk premia, the comprehensive analysis, including overnight news, reveals the significant role played by macroeconomic news and large macro data surprises. The macro jump risk hedging portfolio commands an annual premium of 3.5% and a Sharpe ratio of 0.78, which surpasses the market's Sharpe ratio of 0.53 in the same period. Other types of jump risks, including the monetary policy jump risk, earn smaller or statistically insignificant premia once macro jumps are controlled for.

Secondly, because of the enhanced language understanding and reasoning ability of the LLM, classifying jumps into distinct economic categories adds significant value to investors. A real-time strategy that, each December, selects the pure-play jump-topic hedging portfolio with the most significant risk premia and holds it for the next year attains an out-of-sample Sharpe ratio of 0.84 with highly significant alphas against Fama and French (2018) six-factor models. Placebo strategies with randomly assigned topics never match this performance, highlighting the incremental value created by LLM-based narrative understanding.

These findings contribute to the asset pricing literature in three distinct ways. First, I offer the first around-the-clock empirical decomposition of priced systematic jump risk using high-frequency market data, complementing and extending prior studies that focus solely on intraday movement and may overlook overnight dynamics. Second, I introduce a novel integration of large language models into asset pricing, demonstrating that LLMs can provide economically meaningful causal narrative retrieval and classification of news-based systematic jump risk in a manner that enhances both interpretability and out-of-sample investment performance. Third, I develop a fully automated framework for contemporaneous identification of the economic narrative driving each market jump using open-source LLMs, enabling transparent and replicable mapping from raw high-frequency news to interpretable sources of systematic risk.

My work also relates to two broad strands of literature. First, a rich body of research in high-frequency financial econometrics investigates the differential risk premia associated with various beta estimates (Bollerslev et al., 2016; Aït-Sahalia et al., 2025; Bollerslev et al., 2025). High-frequency asset returns enable precise identification of betas, as the probability of idiosyncratic jumps vanishes with finer sampling intervals (Li et al., 2017). This literature documents significant risk compensation for jump betas. Building on these findings, my paper demonstrates that linking high-frequency return jumps to news text—combined with classifying these jumps into distinct categories using advanced LLMs—enables the identification of heterogeneous risk premia. This categorization enhances the construction of real-time investment strategies that outperform portfolios based on the original systematic factor.

Second, the growing adoption of LLMs in asset pricing has given rise to a rapidly expanding literature focused on their use for interpretable financial insights. One line of research leverages textual embeddings derived from LLMs (e.g., Jha et al. (2025), Chen et al. (2022), Sarkar (2024), Lv (2024), He et al. (2025)). Another line uses prompt-based methods to directly instruct LLMs to perform specific tasks (e.g., Lopez-Lira and Tang (2023), Bybee (2023), Chen et al. (2025), Beckmann et al. (2024)). Building on this emerging literature, my paper highlights a new application of LLMs for narrative retrieval in the context of high-frequency return jumps, showing that this approach not only reveals heterogeneity in risk pricing but also yields substantial improvements in portfolio performance.

In this paper, I combine around-the-clock analysis with state-of-the-art LLMs to offer a comprehensive perspective on systematic risk. The ability to fully attribute market jumps to specific types of economic news in real time opens up new possibilities for constructing interpretable and adaptive investment strategies.

The rest of the paper is organized as follows. Section 2 introduces the methodology. Section 3 shows the data used in my study. Section 4 presents the main empirical results of my analysis.

Section 5 concludes.

2 Methodology

In this section, I first describe the organizing framework for analyzing jump risk premia associated with different systematic news topics. Next, I present the methodology for empirical estimation of the model. Finally, I provide details on the use of the DeepSeek-R1 model for jump narrative retrieval and topic assignment.

2.1 Organizing Framework

To study systematic risk and link it precisely to news events, I consider a continuous-time setting with a large panel of assets driven by a systematic risk factor, dF_t, following the setup in Aït-Sahalia et al. (2025).

Firstly, the dF_t can be decomposed into the continuous part and the jump part in a continuous-time scenario:

$$dF_t = \lambda_t^C dt + \sum_{k=1}^{K} \lambda_t^{J,k} dt + dF_t^C + \sum_{k=1}^{K} dF_t^{J,k}, \tag{1}$$

where λ_t^C and λ_t^J are the risk premia associated with continuous movement (dF_t^C) and discontinuous movement (dF_t^J) in the factor. The superscript k indexes different categories of jump risks triggered by different types of news.

With the setup of factors, I can model the individual asset's excess returns as:

$$dR_t = \left(\beta_t^C \lambda_t^C + \sum_{k=1}^{K} \beta_t^{J,k} \lambda_t^{J,k} \right) dt + \beta_t^C dF_t^C + \sum_{k=1}^{K} \beta_t^{J,k} dF_t^{J,k} + dR_t^I, \tag{2}$$

where $\beta_t^C \in \mathbb{R}^{N \times 1}$ and $\beta_t^J \in \mathbb{R}^{N \times K}$ are the betas of the individual asset with the continuous and jump movements of the factor, and dR_t^I represents the idiosyncratic return movement of the asset unspanned by the factor.

Since dF_t^J and dF_t^C are non-tradeable factors, I use continuous Fama-MacBeth regression developed by Aït-Sahalia et al. (2025) to build hedging portfolios for these distinct sources of risks.

Specifically, once β_t^C and β_t^J are known, let $\beta_t = [\mathbf{1}, \beta_t^C, \beta_t^J] \in \mathbb{R}^{N\times(K+2)}$, we can construct the hedging portfolios as:

$$(\beta_t'\beta_t)^{-1}\beta_t' dR_t \equiv W_t' dR_t, \tag{3}$$

where $W_t \in \mathbb{R}^{N\times(K+2)}$ is the portfolio weight matrix of the $K+2$ factors. The second to last $K+1$ factors satisfy the unique 'pure-play' property following the argument of Fama and French (2020) and Chib et al. (2023):

$$\begin{aligned} w_j'\mathbf{1} &= 0, \quad \forall j = 2, \cdots K+2, \\ w_j'\beta_t^j &= 1, \quad \forall j, \\ w_j'\beta_t^k &= 0, \quad \forall j \neq k, \end{aligned} \tag{4}$$

where w_j is the j-th column of W_t matrix and β_t^j is the j-th column of β_t.[1]

Equation 4 states that each hedging portfolio: (1) has portfolio weights summing to 0; (2) has unit β exposure to its own sources of risk; (3) has zero β exposure to other sources of risk. Therefore, the Fama-MacBeth regression provides a way to isolate pure-play risk-exposure and jointly control all the other types of risks.

Moreover, if news text can provide context for the reasons of systematic jumps, the Fama-MacBeth framework allows for interpretable attribution of systematic risks grouped into different categories.

2.2 Empirical Estimation

The previous section offers a rationale for studying Fama-MacBeth hedging portfolios at the population level. In this part, I discuss in detail the empirical estimation of the continuous-time Fama-MacBeth models.

[1] I provide a proof to these pure-play properties of the Fama-MacBeth factors in Appendix A.

In the first-pass time-series regression, I estimate the factor loadings. There is an important aspect of continuous-time models compared to the low-frequency counterpart. That is to distinguish the jump movement against the continuous movements in factor, as the exposure and risk compensation can be different for the two components.

For this task, I follow the convention in high-frequency econometrics (Aït-Sahalia and Jacod, 2014). If the movement in factor returns is larger than the following threshold, I classify the return as a jump:

$$\widehat{F}_{t,i}^{J} = F_{t,i} \times 1_{\{|F_{t,i}| \geq u_n \sqrt{\tau_i TV_t} \Delta_n^{\varpi}\}}, \tag{5}$$

where u_n is a scaling constant, τ_i is the time-of-the-day volatility adjustment factor, TV_t stands for truncated variance for trading day t, Δ_n is the sampling interval length, and ϖ is the exponent parameter. The TV_t is estimated by considering factor returns, truncating large movements:

$$TV_t = \sum_{i=1}^{n} |F_{t,i}|^2 1_{\{|F_{t,i}| \leq u_n \sqrt{\tau_i BV_t} \Delta_n^{\varpi}\}}, \tag{6}$$

where $BV_t = \frac{\pi}{2} \frac{n}{n-1} \sum_{i=2}^{n} |F_{t,i-1}||F_{t,i}|$. Following Aleti and Bollerslev (2025), I use a truncation threshold $u_n = 3$ and an exponent $\varpi = 0.49$ in the empirical identification of jump movements.

After identifying the jump movement, I can link it to contemporaneous news text and identify economically meaningful groups of jumps triggered by different types of news events. I leave the discussion of using textual analysis to identify meaningful groups of jumps to the next subsection. For now, take these topic groups as given, and I can write the topic-specific jumps as:

$$\widehat{F}_{t,i}^{J,k} = \widehat{F}_{t,i}^{J} \times 1_{\{\text{Jump News}_{t,i} \in \text{Topic}_k\}}. \tag{7}$$

Because jumps are rare, I exploit every jump observed up to and including t to estimate $\beta_t^{J,k}$, following the procedure proposed in Li et al. (2017). Let

$$\mathcal{J}_t(k) = \left\{ (\tau, i) : \tau \leq t, \widehat{F}_{\tau,i}^{J,k} \neq 0 \right\},$$

and stack the corresponding factor and return vectors:

$$\mathbf{F}_t^{J,k} = (\widehat{F}_{\tau,i}^{J,k})_{(\tau,i)\in\mathcal{J}_t(k)}, \quad \Delta^n \mathbf{R}_{m,t}^J = (\Delta^n R_{m,\tau,i})_{(\tau,i)\in\mathcal{J}_t(k)}.$$

The real-time jump beta for asset m and topic k at time t is then:

$$\widehat{\beta}_{m,t}^{J,k} = \left(\mathbf{F}_t^{J,k\top}\mathbf{F}_t^{J,k}\right)^{-1}\mathbf{F}_t^{J,k\top}\Delta^n\mathbf{R}_{m,t}^J. \tag{8}$$

Effectively, the estimator isolates the observations where the systematic factor jumps and uses these time periods to uncover the β for the assets.

These jump movements in factors enable precise identification of βs because, at such times, nearly all large asset-level price changes are driven by the corresponding large move in the factor. As the sampling interval shrinks, the probability that an idiosyncratic jump coincides with a systematic one converges to zero. As a result, even if the observations used are small, the estimates can be very tight with low standard errors. I then concatenate everything together and let $\widehat{\beta}_t^J \in \mathbb{R}^{N_t\times K}$ denote the matrix of jump betas for the total number of N_t assets at time t across K different jump categories.

On the other hand, I estimate the continuous β_t^C of assets using a local window, which is similar to the low-frequency counterpart as in Lewellen and Nagel (2006). Different from the jump betas, continuous betas can be estimated using a rolling window because there are hundreds of observations each month that provide sufficient statistical power. A shorter window captures evolving betas without oversmoothing. The use of high-frequency data also facilitates precise estimation with low standard errors—a theoretical advantage noted in Merton (1980) and empirically validated by Aït-Sahalia et al. (2020).

Define the set:

$$\mathcal{C}_t = \{(\tau,i) : t-l < \tau \leq t, \mid F_{\tau,i} \mid < u_n\sqrt{\tau_i TV_\tau}\Delta_n^{\varpi}\},$$

where l is the parameter controlling the rolling window length. Stack all continuous movements in factors and asset returns:

$$\mathbf{F}_t^C = (F_{\tau,i})_{(\tau,i)\in\mathcal{C}_t}, \quad \Delta^n\mathbf{R}_{m,t}^C = \left(\Delta^n R_{m,\tau,i}^C\right)_{(\tau,i)\in\mathcal{C}_t}.$$

The continuous beta can be estimated as:

$$\widehat{\beta}_{m,t}^C = \left(\mathbf{F}_t^{C\top}\mathbf{F}_t^C\right)^{-1}\mathbf{F}_t^{C\top}\Delta^n\mathbf{R}_{m,t}^C. \tag{9}$$

After this step, let $\widehat{\beta}_t^C$ be the vector stacking all continuous beta estimates.

In the second-pass cross-sectional regression, I use real-time β estimates as the portfolio weights to form the Fama-MacBeth hedging portfolios as in Equation (3). Let $\widehat{\beta}_t = [\mathbf{1}, \widehat{\beta}_t^C, \widehat{\beta}_t^J]$ be stacked β matrix of all assets available at time t. The hedging portfolio can be formed as:

$$(\widehat{\beta}_t'\widehat{\beta}_t)^{-1}\widehat{\beta}_t'\Delta_i^n R_t \in \mathbb{R}^{(K+1)\times 1}. \tag{10}$$

These $K+1$ hedging portfolios, according to Fama and French (2020), have unique pure-play properties that isolate portfolios hedging different systematic risks.

Another key advantage of the high-frequency analytical framework lies in the inference on the risk premia. Unlike the low-frequency models, where Shanken adjustment (Shanken, 1992) is typically required to account for estimation errors in βs, as the sampling interval shrinks, according to the double asymptotic theory developed by Aït-Sahalia et al. (2025), we can treat β as if they are observed without errors.

After obtaining these hedging portfolios, I can form estimates and conduct inference for the unconditional risk premia of different interpretable risk factors. Specifically, use $\lambda^C \equiv \mathbb{E}[\lambda_t^C]$ and $\lambda^J \equiv \mathbb{E}[\lambda_t^J]$ to denote the unconditional continuous risk premia and jump risk premia, respectively. The estimates for these unconditional risk premia can be obtained by averaging the hedging portfolios' returns:

$$
\begin{aligned}
\widehat{\lambda}^{C} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[(\widehat{\beta}_t'\widehat{\beta}_t)^{-1}\widehat{\beta}_t'\Delta_i^n R_t\right]_2, \\
\widehat{\lambda}^{J,k} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[(\widehat{\beta}_t'\widehat{\beta}_t)^{-1}\widehat{\beta}_t'\Delta_i^n R_t\right]_{k+2},
\end{aligned}
\tag{11}
$$

where the subscripts index for the second entry and $(k+2)$-th entry, respectively, in factor return vector. The standard error and confidence interval can be constructed with the volatility of the hedging portfolio, and the final t-stat is closely related to the Sharpe ratio of the topic-specific hedging portfolio.

2.3 Narrative Retrieval and Topic Classification

The previous section discusses the empirical estimation of the Fama-MacBeth regression model assuming a given categorization of jumps. Obtaining an economically meaningful division of the jump categories can be critical for generating heterogeneity in risk exposure and risk prices.

To achieve this goal, I link market jumps to high-frequency newswire data and apply state-of-the-art large language models to analyze the concurrent news in the 15-minute interval at the jump time.

Even though the high-frequency data is helpful for reducing the volume of concurrent news, there can still be hundreds of news stories released in the time window of the market jump. To sift through the large amount of text, the language model needs to possess strong reasoning skills to identify news stories that are both systematic and aligned with the movement in the market.

Moreover, proprietary LLMs present a significant hurdle for replication studies, and feeding newswire text to these models through the API may violate the copyright agreement of the data vendor.

With these two considerations, I self-host and deploy the state-of-the-art open-source reasoning language model, the DeepSeek-R1-0528, to analyze the concurrent news during the market jump. The DeepSeek-R1 is a 685 billion parameter mixture of experts model with the strongest reasoning capabilities among open-source language models.[2]

I then feed the model with concurrent news events, time, market response direction, and magnitude, and ask the model to identify the likely cause of the jump from the news stories using the following prompt.

Prompt 1 (Narrative Retrieval): From {event start time} to {event end time} ET, the US market {increases/decreases} by {event ret}%. Listed below are the news headlines in this period from the Dow Jones Newswire. Can you find what is likely causing the jump? Output your answer in JSON in the format of {"News_id": list[int], "Explanation": str}. If there is no plausible news accounting for the jump, output "News_id" as an empty list.

{News IDs followed by news headlines.}

The retrieved news narratives and explanations offer interpretable signals for what is moving the market. With this narrative retrieval step, I obtain the list of news relevant to explaining the contemporaneous market jump and the reasoning logic behind such attribution through the LLM.

After the narrative retrieval step, I apply the language model to obtain the topic categories for each jump event. The topic's overall categories should satisfy three goals: (1) nearly all jumps can be classified as one of the categories; (2) the categories should be broad so that there are enough jumps within the category for identifying jump βs; (3) the categories should be mutually exclusive.

With the three goals, I design the following prompt to obtain overall topic categories for all the jump events.

[2]According to the LLM arena https://lmarena.ai/leaderboard, the DeepSeek-R1-0528 model ranks 6th overall and ranks 1st among open-weight models. I discuss in more detail on hosting the model in the Online Appendix B.

Prompt 2 (Overall Jump Topic Categories): Please read the provided explanations and narratives for why the U.S. market jumps. Help me classify them into comprehensive and mutually exclusive topics. Ensure nearly all jumps can be classified into one of the topics. Output a JSON file in the format of {"Topic_Name": str, "Topic Definition": str, "Text_ID": list[int]}.

{Narrative IDs followed by explanation text generated in **Prompt 1**.}

I feed all explanations with a non-empty news ID list generated from Prompt 1 to this topic categorization step through the same DeepSeek R1 model. After merging related topics, the procedure yields five distinct broad topic categories for all the jump events in the market. To allow for non-classified jumps, I include a 'None of the Above' category. I provide details on these categories and their definitions in the Table 1.

After obtaining the overall jump categories, I use the same DeepSeek R1 model to zoom in on each individual jump and classify them into one of the categories. This more focused classification step enables the model to produce more accurate and consistent classification results.

Prompt 3 (Jump Classification): Please read the provided explanation and narrative, as well as the relevant news for a U.S. equity market jump event, and classify the jump into one of the following six topics:

{Topic IDs, Topic Names, Topic Definitions listed in Table 1, identified with **Prompt 2**}

Output your response in JSON in the format of {"Topic_Category": int, "Explanation": str}. Here is the explanation followed by the relevant news:

{Explanation: Explanation text generated using **Prompt 1**. Relevant News: News identified as relevant using **Prompt 1**.}

After this step, I assign each market jump to one of the economically distinct categories defined in Table 1. This categorization enables me to apply the decomposition framework in Equation (2) to break down asset exposures across different sources of risk. As a result, I can construct hedging portfolios that yield interpretable risk premia, fully decomposing the overall market risk.

3 Data

In this section, I present the dataset used in my study. The first part introduces the high-frequency panel data on S&P 1500 constituents returns. The second part presents the construction of the around-the-clock market factor. The third part shows the high-frequency newswire data from the Dow Jones.

3.1 High Frequency Return Panel Data

Successfully recovering risk premia requires access to a long time span of data, as well as a broad cross-section of liquid assets to construct hedging portfolios that accurately mimic jump risks. To this end, I compile a large panel of high-frequency return data for S&P 1500 constituent companies using TAQ milisecond data from WRDS, covering an extensive sample period from September 1997 to May 2020—nearly 23 years.[3]

Crucially, the S&P 1500 membership for each firm is determined using the Compustat *idxcst_his* table, which records historical index constituents.[4] This ensures that, at each point in time, only firms that were actually included in the S&P 1500 are used for constructing returns, reflecting the real-time available investment universe. After merging the index membership with CRSP and applying standard exchange and share code filters, my final sample contains 3,488 unique companies, providing both a long time-series and a rich cross-section for robust asset pricing analysis.

To clean the high-frequency data and mitigate the effects of market microstructure noise, I follow the procedure outlined in Da and Xiu (2021). Appendix C provides further details on the

[3] I start the sample from September 1997 because, as will be introduced later, the S&P 500 E-mini futures product was introduced to the market since then.

[4] The `idxcst_his` table was removed from WRDS in July 2020, which limits my sample period to dates prior to its removal.

data preprocessing, including the aggregation of millisecond-level trades to 15-minute intervals for the main empirical analysis.

Finally, because the TAQ data only cover intraday observations during regular trading hours (9:30 a.m. to 4:00 p.m. ET), I supplement these data by linking TAQ with CRSP to construct a comprehensive panel of returns that incorporates stock splits, dividend payments, and overnight price movements. Specifically, I use the daily open and close prices from CRSP—rather than TAQ—to ensure accurate measurement of daily returns. The overnight return is then computed as the ratio of the CRSP close-to-close gross return to the open-to-close gross return. This approach ensures that all adjustments for splits and dividends are fully captured in the overnight return component. Further details on this procedure and the linking process are provided in Appendix C.

3.2 High Frequency Market Returns and S&P 500 E-mini Futures

To obtain the systematic factor, I consider the market return of the U.S. equities. It is also critical to have full high-frequency observations of the systematic factor. This allows for full decomposition of the systematic risks without leaving any jump risks outside the picture.

To this purpose, I construct both the high-frequency market returns for intraday variations, and I use the S&P 500 E-mini futures returns to obtain the overnight variations of the index because this product is traded around the clock.

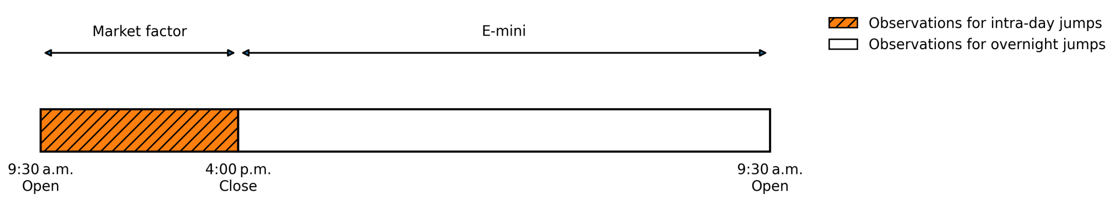

Figure 1 Intraday and Overnight Jump Identification Timeline
This figure presents the timeline and instruments we used to identify the market jumps for the intraday period and overnight periods

Firstly, for the intraday market return factor, I follow the construction procedure of Mkt-Rf factor from Fama and French (1993), which is based on value-weighted returns of common stocks listed on the NYSE, NASDAQ, and AMEX.

For the overnight market return factor, I leverage the S&P 500 E-mini futures data from the CME DataMine database, which provides the tick-level high-frequency data of the product with a long span of history.[5] The E-mini futures contract is the most liquid equity index futures product globally, trading nearly 24 hours a day and facilitating continuous price discovery outside of regular U.S. equity market hours. Its deep liquidity and global participation make it a primary venue for incorporating and reflecting new information—especially systematic news events—during periods when the underlying cash equity market is closed. By utilizing E-mini futures, I capture the overnight market response to news releases, geopolitical developments, and macroeconomic events, ensuring a comprehensive measure of market-wide return dynamics around the clock. Similar to the intraday market factor, I sample the S&P 500 E-mini futures at a 15-minute frequency.

Figure 1 provides an illustration of the timeline I used to identify intraday and overnight jumps. Specifically, I divide each trading day into two parts: the intraday observations from 9:30 a.m. to 4:00 p.m. ET and the overnight period that spans from 4:00 p.m. to the next day's opening at 9:30 a.m. I then calculate the truncated estimator for realized volatility defined in Equation (6) using separately the intraday and overnight observations. I then use the two RV estimators as the truncation threshold for identifying intraday and overnight jumps as defined in Equation (5).

I don't merge the intraday and overnight observations because the two have different diurnal patterns. Also, the separate estimation enables direct comparison with high-frequency finance literature, which focuses on the intraday component of the return to identify jumps (Aït-Sahalia and Jacod, 2014).

I then use a threshold of 0.5% to filter out large jump observations for empirical analysis. In the robustness check section, I provide evidence on alternative thresholds for large jumps.

[5] I discuss in detail the procedure for data cleaning and constructing a continuously rolling-over return series for the E-mini futures in Appendix C.

3.3 Dow Jones Newswire

I use the Dow Jones Newswire to retrieve contemporaneous news released around the time of systematic market jumps. The Dow Jones Newswire is a real-time, timely, and comprehensive news service widely used by institutional investors via platforms such as Bloomberg.

This dataset offers two key advantages. First, it provides precise time stamps indicating exactly when a news item reaches the market, enabling accurate alignment with high-frequency return data. Second, it offers broad and reliable coverage of market-relevant news through reputable media outlets such as The Wall Street Journal, Barron's, MarketWatch, among others.

These features make the Dow Jones Newswire particularly well-suited for identifying market-moving narratives in conjunction with high-frequency financial data.

In contrast to Aleti and Bollerslev (2025), who apply extensive filtering using the anchor phrase methodology,[6] I retain all news items released during the identified market jump intervals. Their approach filters the newswire to isolate systematic content, which requires substantial pre-processing. In my approach, I take advantage of recent advances in LLMs and directly prompt the model to process the raw news text and identify truly market-relevant items without pre-filtering.

To ensure that all relevant news fits within the LLM's context window, I restrict attention to jumps occurring in 15-minute intervals and exclude those during futures maintenance windows or after-hours periods. Over 95% of the identified jumps occur within 15-minute windows,[7] so this restriction maintains comprehensive coverage of systematic jump events.

4 Empirical Results

In this section, I present the main empirical results. Firstly, I provide a summary of evidence on what news triggers a market jump. Next, I present risk-premia estimates of ex-ante compensation

[6] See Hoberg and Manela (2025) for a detailed overview of the anchor phrase method.

[7] Some jumps fall in intervals with length larger than 15 minutes because these are times for brief trading halts, futures daily clearing sessions, or weekend closures with no high-frequency price data. These account for a small fraction of total identified jumps and typically reflect mechanical rather than news-driven price adjustments.

for bearing different risks. Then I show investors can improve their utility using LLM to understand the systematic risk compensation in real time and build a portfolio to outperform the market. Lastly, I dive into the mechanism and forces driving the difference in risk compensation.

4.1 What Triggers Market Jump?

Figure 2 presents a visualization of all market jumps occurring in my sample. I divide them into two categories: (1) jumps occurring in the intraday period; (2) jumps occurring in the overnight period.

One prominent feature stands out from the figure: the vast majority of market jumps occur overnight. I find that the intraday jumps account for less than 8.7% of the total observations. The evidence highlights the importance of taking a holistic, 'around-the-clock' view to include overnight observations to better understand systematic jump risk affecting the equity market. Otherwise, the risk premia estimates might suffer from omitted variable bias.

Linking the news with the jumps, I run each jump and corresponding news through **Prompt 1** to **Prompt 3**. With the overall categories listed in Table 1, I map each jump into one of the categories. As a first step sanity check, I examine whether the topic classification from the LLM is consistent with the theme of each topic.

Figure 3 plots the word clouds of the news items identified as relevant by the LLM for triggering the jump. I find consistent patterns within each topic category and distinct word distributions across categories. The top words occurring within each cluster match the theme of the topic, suggesting the LLM does a good job in allocating each jump-triggering news to relevant categories.

Next, Table 2 provides an overview of different types of news events' contributions to driving the stock market jump. Firstly, I find that more than 88% of jump events can be explicitly linked to news stories. The number is consistent with findings from Baker et al. (2021). The evidence suggests that at the aggregate market level, the systematic jumps are mainly triggered by public information rather than private trading with hidden information.

Among the jumps that can be linked to news, I find that the unclassified category only accounts for less than 3% of the total observations. This means the overall categories identified by LLM using **Prompt 2** are comprehensive to cover most of the jump observations.

For the five groups of classified jumps, the 'international market spillovers' topic accounts for most of the observations, representing 33% of total jumps. This is followed by the 'U.S. macro data surprises' category, which takes up 1/5 of the total jump observations. Another prominent category is the spillover from bellwether or systematically important firms' earnings, accounting for 16% of total jumps. The last two categories, policy and geopolitical tensions, each contribute around 7 to 8% of total observations. However, the policy-triggered jumps are more volatile than other categories. In the R^2 space, it ranks third in variation accounted, only below international spillover and macro data surprises.

The prominent role played by international market spillover and macro data release again suggests the importance of overnight observations for understanding U.S. market jump risks. This is because international news usually happens when the U.S. market is closed, and the macro data releases usually happen before the U.S. market opens.

In the next section, I apply the Fama-MacBeth regression approach to quantify the importance of the different types of risk for ex-ante risk premia.

4.2 What Risks are Priced?

To quantify risk prices, I first estimate the real-time jump and continuous betas of the S&P 1500 panel of stocks using Equations (8) and (9). Continuous betas are updated monthly using a 1-month rolling estimation window, and jump betas are updated annually using an expanding estimation window.

Figure 4 plots the time series of percentile estimates for both continuous and topic-specific jump betas. The median continuous beta hovers around 1, exhibiting notable time-series variation. In contrast, the topic jump betas display more muted variation over time due to their lower update frequency and the use of expanding window estimation.

Notably, the jump betas for macroeconomic, corporate, and international topics rise significantly following the 2008 financial crisis. This increase likely reflects the surge in systematic jump events during the crisis and heightened comovement between individual stocks and the corresponding jump risk factors. Toward the end of the sample period, the geopolitics jump beta also rises, consistent with elevated geopolitical tensions during the Trump administration.

Using the real-time beta estimates, I then form the jump hedging portfolios using Equation (10) and calculate risk-premia estimates using Equation (11). Table 3 presents the estimates, standard errors, and Sharpe ratios of topic hedging portfolios. Firstly, the continuous risk premia is large in magnitude, accounting for more than 47% of the total market risk premia. However, because of its high volatility, the hedging portfolio has low Sharpe ratios.

Among the topic-specific jump risks, the macroeconomic category commands the highest premium, with an annualized return of 3.54% and a t-statistic of 2.78. Importantly, the macro jump-hedging portfolio achieves a Sharpe ratio of 0.78, outperforming the contemporaneous market's Sharpe ratio of 0.53, reflecting its high return and relatively low volatility after controlling for other systematic jump risk exposures and continuous risk exposures.

Other topic-specific jump risks with positive risk premia include the corporate bellwether and international market spillover categories, with annualized returns of 2.32% and 0.87%, respectively. While the international topic accounts for the largest share of contemporaneous jump events, it offers limited explanatory power for ex-ante risk compensation. The corporate topic delivers a better risk-return tradeoff than the international category, but the estimated premium is statistically insignificant, suggesting that investors may find it difficult to identify and act on this risk factor in real time.

Figure 5 displays the cumulative returns of the three jump risk-hedging portfolios. Both the corporate and international portfolios exhibit higher volatility than the macro portfolio. Given their lower premia and greater volatility, the macroeconomic topic emerges as the only one that delivers a statistically and economically significant source of priced jump risk.

An important question is whether priced jump risk arises from overnight or intraday returns. Table 4 provides supporting evidence. Panel A examines portfolios that hold market exposure exclusively during either the overnight or intraday window. The results show that nearly all U.S. equity risk premia accrue overnight, underscoring the importance of overnight risk. Panel B applies the real-time Fama–MacBeth regression framework to separately estimate risk premia for continuous returns, overnight jumps, and intraday jumps. Again, the evidence points to overnight jump risk as the primary source of compensation.

To further validate this result, I consider the GANs-based SDF constructed in Aleti and Bollerslev (2025), following the methodology of Chen et al. (2024). The Fama–MacBeth regression confirms that overnight jump risk earns the most significant premium. These findings suggest that focusing solely on intraday returns may lead to biased estimates of priced risk.

4.3 Real-time Jump Risk Management

Building on the risk premia estimates in Section 4.2, a natural question is whether investors can identify the heterogeneity in risk compensation across different types of systematic risk in real-time and construct hedging portfolios that outperform the market.

In this section, I evaluate the performance of a real-time optimal jump-topic hedging strategy. At the end of each year, I use Equation (11) to estimate the cross-sectional prices of jump risk, substituting in the most up-to-date jump betas computed using all available data. I then select the jump topic with the highest Sharpe ratio, i.e., the most significantly priced jump risk, as the basis for the hedging portfolio in the subsequent year.

Figure 6 illustrates the performance of this real-time maximum Sharpe ratio jump-topic hedging portfolio. The macroeconomic (macro) topic consistently emerges as the dominant source of priced jump risk from the early years of the sample and continues to exhibit strong performance toward the end. The resulting portfolio delivers an out-of-sample Sharpe ratio of 0.84, substantially exceeding the market Sharpe ratio of 0.53.

To assess whether these jump-topic hedging portfolios capture new dimensions of risk beyond traditional asset pricing factors, I regress their returns on standard risk factors commonly used in the literature. Table 5 reports the results for both the macro hedging portfolio and the real-time selected topic portfolio. In all specifications, the estimated alphas remain highly significant, indicating that these portfolios are not simply repackaging exposures to known risk factors but rather capture distinct sources of priced jump risk.

An important follow-up question concerns the economic value added by using large language models (LLMs) to classify jumps into distinct categories based on contemporaneous news. To address this, I conduct a placebo analysis in which I randomly assign each jump to one of the six categories, drawn from a uniform distribution. Using the same methodology, I construct real-time hedging portfolios that invest in the jump category with the highest Sharpe ratio within the estimation sample.

Figure 6 includes 20 such placebo strategies, shown as faint lines corresponding to random seeds from 1 to 20. None of the placebo portfolios achieves a Sharpe ratio as high as the LLM-based strategy. The average Sharpe ratio across placebo portfolios is 0.24, which is significantly lower than that of the market. The highest Sharpe ratio among the placebo strategies is 0.66, over 20% lower than the LLM-based optimal portfolio.

Taken together, these results suggest that LLMs add substantial economic value by identifying jump events linked to similar underlying economic risks. Because these risks exhibit stable pricing over time, investors can exploit real-time information to construct hedging portfolios that deliver superior out-of-sample performance relative to the market.

4.4 What Drives Macro Jump Risk?

The preceding analyses identify macroeconomic jump risk as the most significantly priced source of systematic jump risk, both statistically and economically. A natural follow-up question is: what specific macroeconomic indicators are responsible for these market-moving jumps? To shed light on the underlying drivers, I examine the detailed composition of macro-related news that triggers U.S. market jumps.

Table 6 presents a breakdown of macroeconomic jump events into high-level categories and sub-categories based on LLM analysis of contemporaneous news articles.

Three primary macro categories dominate: Labor Market, Inflation, and Growth and Real Activity. Among these, the labor market stands out as the most frequent driver, accounting for 50 of the 153 macro jumps. Within this group, the Non-farm Payroll (NFP) report, commonly referred to as the Employment Situation release, alone explains 34 jumps. This reflects the report's central role in shaping investor expectations about the real economy and the future stock market.

In total, 153 market jump events within the macroeconomic category are successfully attributed to specific indicators using LLM-guided classification. This evidence reinforces the notion that systematic jump risk priced in the market stems from a well-defined set of high-frequency economic data releases. These releases tend to occur in the pre-market period, further validating the earlier finding that overnight jump risk is the primary driver of the equity risk premia in the U.S.

5 Conclusion

In this paper, I present the first comprehensive, around-the-clock analysis of systematic jump risk in the U.S. equity market by integrating high-frequency financial data with contemporaneous news narratives retrieved and interpreted using a state-of-the-art LLM. By decomposing systematic risk into interpretable, topic-specific jump components, I provide new insights into the sources, pricing, and management of jump risk.

The empirical evidence yields several key findings. First, this classification reveals significant heterogeneity in risk premia: macroeconomic jump risk commands a sizeable and statistically significant premium, outperforming the market in terms of Sharpe ratio, while other types of risk carry limited compensation.

Second, I demonstrate that this interpretable risk decomposition has substantial economic value for investors. A real-time trading strategy that dynamically allocates to the most significantly priced jump-topic portfolio each year achieves an out-of-sample Sharpe ratio of 0.84, exceeding that of the aggregate market. Importantly, placebo analyses confirm that this performance cannot be replicated by randomly assigned jump categories, underscoring the value added by LLM-based narrative understanding.

These results contribute to the literature on systematic risk, high-frequency econometrics, and the growing field of LLM applications in finance. Methodologically, I extend the continuous-time Fama–MacBeth regression framework to allow for interpretable decomposition of risk based on contemporaneous news content. Empirically, I provide a transparent and replicable approach for mapping high-frequency news to market movements using open-source LLMs. Practically, the findings offer new tools for constructing real-time, interpretable, and economically significant hedging strategies.

Future work could apply similar techniques to study cross-sectional asset pricing implications, explore firm-level jump exposures, or extend the analysis to global markets and asset classes. As LLMs continue to evolve in their reasoning and interpretability capabilities, their integration with high-frequency financial data promises to unlock even deeper understanding of how markets process information and price risk.

References

Aït-Sahalia, Yacine, and Jean Jacod, 2014, High-frequency financial econometrics, in *High-Frequency Financial Econometrics* (Princeton University Press).

Aït-Sahalia, Yacine, Ilze Kalnina, and Dacheng Xiu, 2020, High-frequency factor models and regressions, *Journal of Econometrics* 216, 86–105.

Aït-Sahalia, Yacine, Jean Jacod, and Dacheng Xiu, 2025, Continuous-time Fama-MacBeth regressions, *The Review of Financial Studies* Forthcoming.

Aleti, Saketh, and Tim Bollerslev, 2025, News and asset pricing: A high-frequency anatomy of the SDF, *The Review of Financial Studies* 38, 712–759.

Baker, Scott, Nicholas Bloom, Steven J Davis, and Marco C Sammon, 2021, What triggers stock market jumps?, *NBER Working Paper No. 28687* .

Beckmann, Lars, Heiner Beckmeyer, Ilias Filippou, Stefan Menze, and Guofu Zhou, 2024, Unusual financial communication: Chatgpt, earnings calls, and financial markets, *Olin Business School Center for Finance & Accounting Research Paper* .

Bollerslev, Tim, Sophia Zhengzi Li, and Viktor Todorov, 2016, Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns, *Journal of Financial Economics* 120, 464–490.

Bollerslev, Tim, Andrew J Patton, and Rogier Quaedvlieg, 2025, Granular betas and risk premium functions, *Journal of Econometrics* 106034.

Bybee, J Leland, 2023, The ghost in the machine: Generating beliefs with large language models, *arXiv preprint arXiv:2305.02823* 349–389.

Bybee, Leland, Bryan Kelly, Asaf Manela, and Dacheng Xiu, 2024, Business news and business cycles, *The Journal of Finance* 79, 3105–3147.

Bybee, Leland, Bryan Kelly, and Yinan Su, 2023, Narrative asset pricing: Interpretable systematic risk factors from news text, *The Review of Financial Studies* 36, 4759–4787.

Chen, Jian, Guohao Tang, Guofu Zhou, and Wu Zhu, 2025, Chatgpt and deepseek: Can they predict the stock market and macroeconomy?, *arXiv preprint arXiv:2502.10008* .

Chen, Luyang, Markus Pelger, and Jason Zhu, 2024, Deep learning in asset pricing, *Management Science* 70, 714–750.

Chen, Yifei, Bryan T Kelly, and Dacheng Xiu, 2022, Expected returns and large language models, *Available at SSRN 4416687* .

Chib, Siddhartha, Yi Chun Lin, Kuntara Pukthuanthong, and Xiaming Zeng, 2023, Slope factors outperform: evidence from a large comparative study, *Available at SSRN 3966807* .

Da, Rui, and Dacheng Xiu, 2021, When moving-average models meet high-frequency data: uniform inference on volatility, *Econometrica* 89, 2787–2825.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

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

Fama, Eugene F, and Kenneth R French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Fama, Eugene F, and Kenneth R French, 2020, Comparing cross-section and time-series factor models, *The Review of Financial Studies* 33, 1891–1926.

Glasserman, Paul, Kriste Krstovski, Paul Laliberte, and Harry Mamaysky, 2025, Does overnight news explain overnight returns?, *arXiv preprint arXiv:2507.04481* .

Guo, Daya, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al., 2025, DeepSeek-R1: Incentivizing reasoning capability in llms via reinforcement learning, *arXiv preprint arXiv:2501.12948* .

He, Songrun, Linying Lv, Asaf Manela, and Jimmy Wu, 2025, Chronologically consistent large language models, *arXiv preprint arXiv:2502.21206* .

Hoberg, Gerard, and Asaf Manela, 2025, The natural language of finance, *Available at SSRN 5119322* .

Holden, Craig W, and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *The Journal of Finance* 69, 1747–1785.

Jha, Manish, Hongyi Liu, and Asaf Manela, 2025, Does finance benefit society? a language embedding approach, *The Review of Financial Studies* hhaf012.

Lewellen, Jonathan, and Stefan Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289–314.

Li, Jia, Viktor Todorov, and George Tauchen, 2017, Jump regressions, *Econometrica* 85, 173–195.

Lopez-Lira, Alejandro, and Yuehua Tang, 2023, Can chatgpt forecast stock price movements? return predictability and large language models, *arXiv preprint arXiv:2304.07619* .

Lv, Linying, 2024, The value of information from sell-side analysts, *arXiv preprint arXiv:2411.13813* .

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

Merton, Robert C, 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Sarkar, Suproteem K, 2024, Economic representations.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *The Review of Financial Studies* 5, 1–33.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al., 2022, Chain-of-thought prompting elicits reasoning in large language models, *Advances in neural information processing systems* 35, 24824–24837.

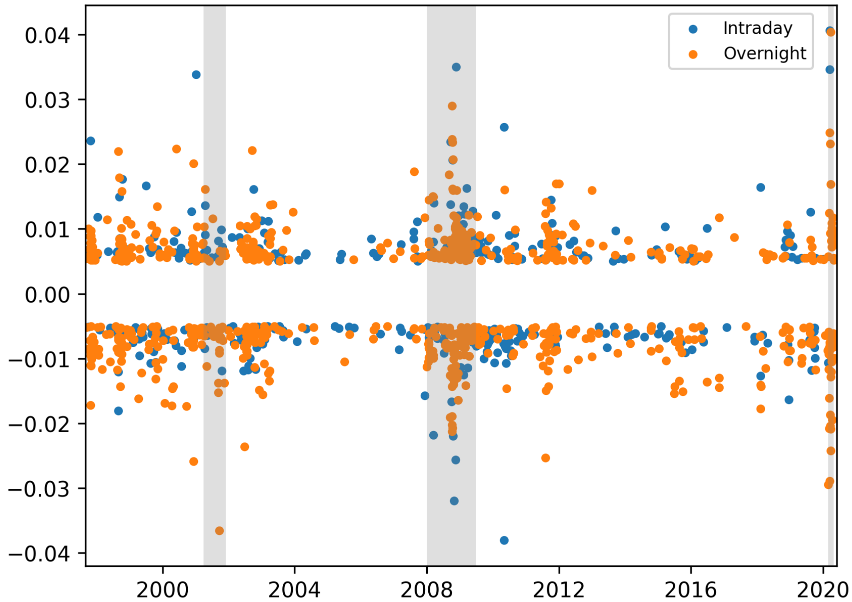

Figure 2 Intraday and Overnight Jump

This figure displays the time series of intraday and overnight jump returns identified using Equation (5). Blue dots indicate intraday jumps, while orange dots represent overnight jumps. Shaded gray areas denote NBER dated recessions. The sample period covers September 1997 to May 2020.

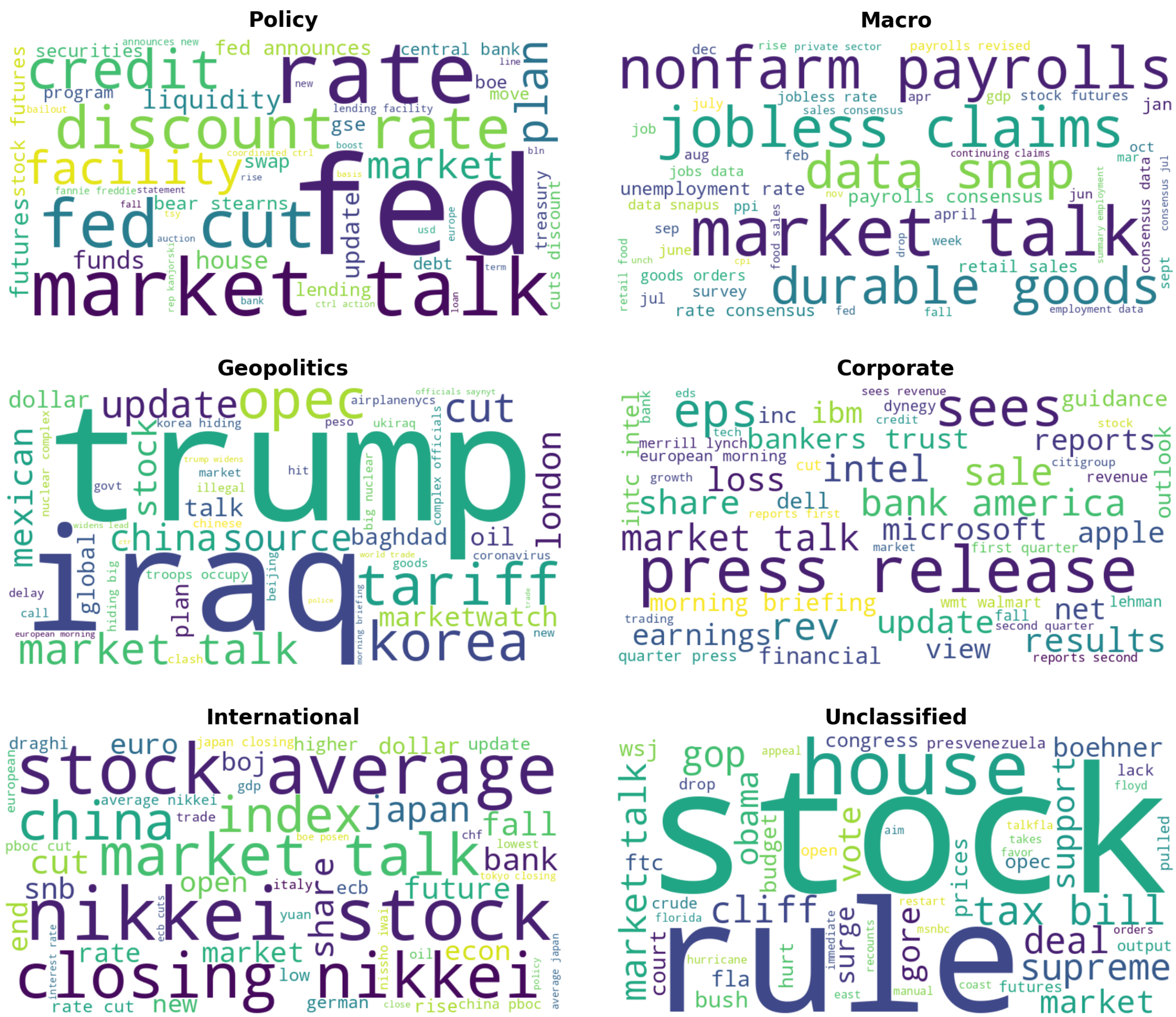

Figure 3 Word Clouds for Different Systematic Risk Topics

This figure displays word clouds generated from news headlines associated with each category of market jumps listed in Table 1. The word size reflects the frequency of each term within the headlines attributed to a given topic. The sample spans from September 1997 to May 2020.

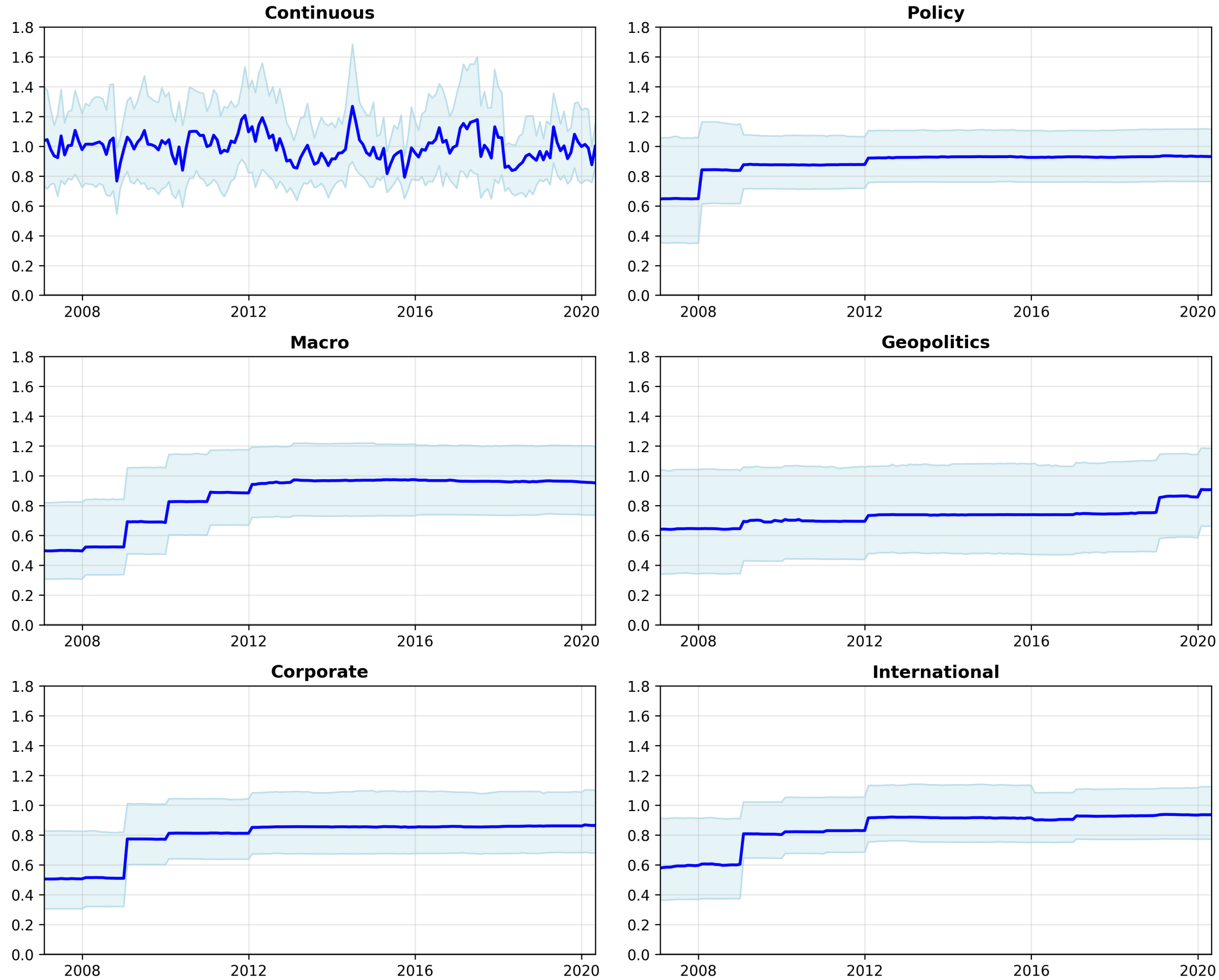

Figure 4 Continuous and Jump Betas over Time

This figure displays the time series of percentile estimates for continuous betas and topic jump betas. The first panel presents the continuous beta estimates, while the following five panels show the jump beta estimates for the five interpretable topics listed in Table 1. The sample covers the out-of-sample estimation period from January 2007 to May 2020. The blue line denotes the cross-sectional median, and the light blue shaded area indicates the interquartile range (25th to 75th percentiles). Continuous betas are updated monthly using a rolling window, whereas jump betas are updated annually using an expanding window.

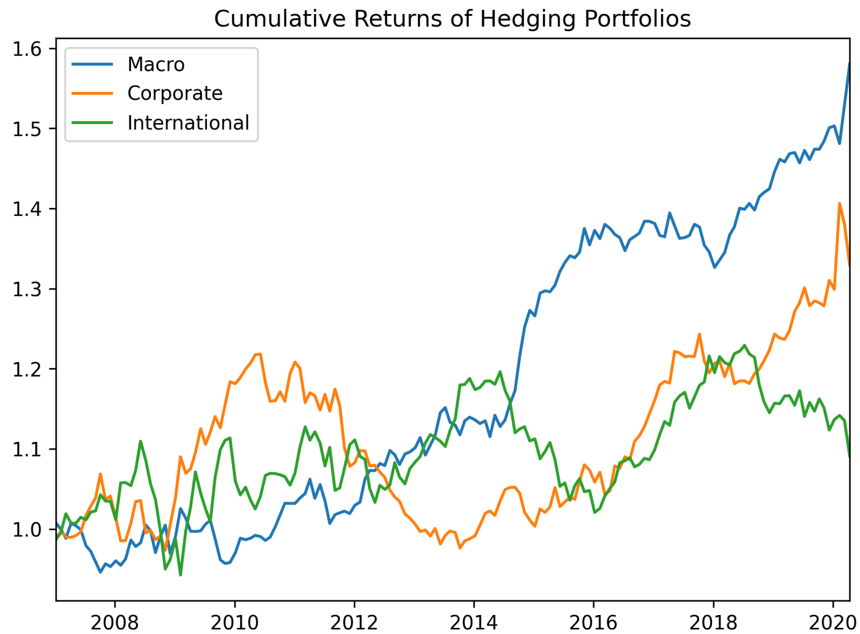

Figure 5 Cumulative Returns of Risk Hedging Portfolios

This figure plots the cumulative returns of portfolios designed to hedge against jump risks associated with three types of systematic events: (1) U.S. macroeconomic data surprises, (2) corporate earnings and forward guidance, and (3) international market spillovers. The sample covers the out-of-sample period from January 2007 to May 2020.

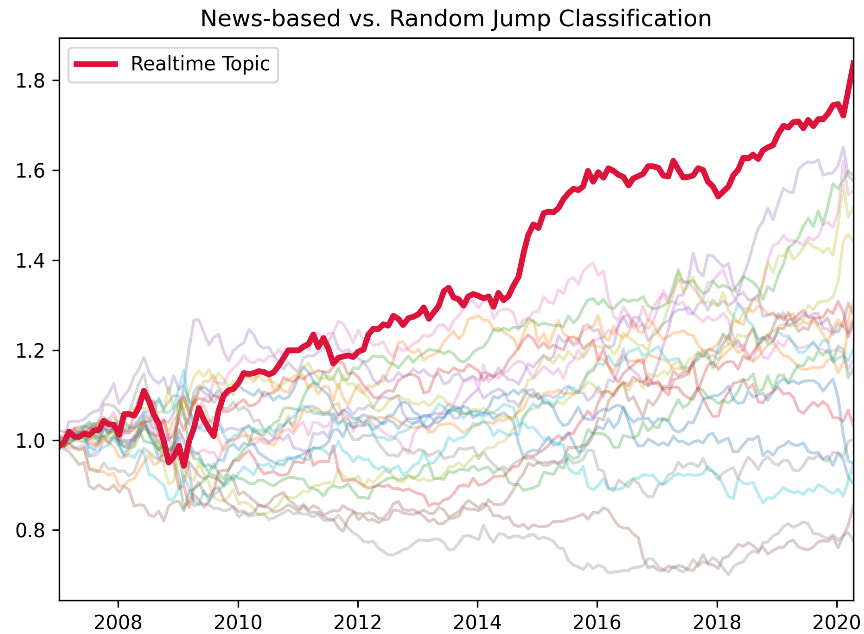

Figure 6 Economic Value of Jump Classification with LLM

This figure plots the cumulative returns of trading strategies that, in real time, invest in the jump risk hedging portfolio with the highest Sharpe ratio. The red line represents the strategy using jump categories classified based on contemporaneous newswire and LLM analysis. The lighter lines depict placebo strategies where jump categories are randomly assigned from a uniform distribution. To ensure comparability, all portfolios are volatility-scaled to match the strategy based on LLM-classified jumps. The sample spans the out-of-sample period from January 2007 to May 2020.

Table 1 Definitions of Topic Categories

This table provides information on the overall jump topic categories identified using the **Prompt 2**. This classification information is then provided in the **Prompt 3** to assign each jump into one of the categories.

Topic ID	Topic Name	Topic Definition
1	*U.S. Policy Actions (Monetary, Fiscal, & Political)*	Federal Reserve rate moves, emergency facilities, policy statements, and major fiscal or political event.
2	*U.S. Macro Data Surprises*	Releases of macroeconomic information such as retail sales, GDP, inflation, payrolls, jobless claims, etc., that diverge sharply from consensus.
3	*Geopolitical & Security Events*	Developments in cross-border negotiations or tensions. Terror attacks, war-risk headlines, or news that eases/tightens military tensions.
4	*Corporate Earnings & Guidance*	Earnings/Warnings from bell-wether firms or industries that drag or lift the whole market.
5	*International Market Spillovers*	Significant moves or outlook changes in major foreign equity markets, commodities, energy prices, or FX rates. Overseas monetary/fiscal policy shifts, trade measures, capital-flow controls, or other cross-border actions that carry global risk implications.
6	*None of the Above*	Material news that do not fit the above definitions.

Table 2 Summary Statistics of the Jumps by Categories

This table reports summary statistics for jumps classified into seven categories. The six main categories—Policy, Macro, Geopolitics, Corporate, International, and Unclassified—correspond to those defined in Table 1. An additional category, 'Unattributable', includes jumps for which no corresponding narrative is identified in the contemporaneous newswire. The final column reports aggregate statistics across all jumps. For each category, the following statistics are reported: the number of jumps (N), the proportion of total jumps (%), the proportion of positive jumps (Prop Pos), the mean return (Mean), mean absolute return (Mean Abs), standard deviation (Std), interquartile range (IQR), skewness (Skew), and the variance explained (R^2), calculated as the sum of squared returns in that category divided by the total sum of squared returns across all jumps. The sample period spans from September 1997 to May 2020.

	Policy	Macro	Geopol.	Corp.	Intl.	Unclass.	Unattrib.	All
N	54	153	63	114	242	17	87	730
%	7.40	20.96	8.63	15.62	33.15	2.33	11.92	100.00
Prop Pos (%)	61.11	49.67	44.44	40.35	38.43	35.29	57.47	45.48
Mean (%)	0.37	-0.04	-0.23	-0.18	-0.20	-0.22	0.13	-0.08
Mean Abs (%)	1.03	0.79	0.93	0.75	0.78	0.90	0.87	0.82
Std (%)	1.19	0.85	1.06	0.79	0.82	1.04	0.96	0.91
IQR (%)	1.52	1.37	1.43	1.34	1.36	1.35	1.45	1.38
Skew	0.27	0.06	-0.38	0.09	0.38	-0.15	-0.10	0.19
R^2	13.46	17.97	11.99	12.06	28.25	3.00	13.28	100.00

Table 3 Risk Premia Estimates

This table reports annualized risk premia estimates, standard errors, and Sharpe ratios for seven Fama-MacBeth hedging portfolios, based on the continuous-time Fama-MacBeth regression described in Equation (11). Risk premia and standard errors are expressed in percentage points. The annualized Sharpe ratio is computed using the corresponding hedging portfolio returns. The final row reports results for the contemporaneous market excess return as a benchmark. The sample covers the out-of-sample period from January 2007 to May 2020.

	Ann RP(%)	Std Err(%)	SR
Continuous	4.06	(5.31)	0.21
Policy	-1.95	(1.65)	-0.33
Macro	3.54	(1.27)	0.78
Geopolitics	-0.67	(1.13)	-0.16
Corporate	2.32	(1.65)	0.39
International	0.87	(1.81)	0.13
Unclassified	1.09	(1.02)	0.29
Market	8.48	(4.35)	0.53

Table 4 Overnight and Intraday Risk Premia

This table presents the decomposition of annualized risk premia. Panel A reports the breakdown into overnight and intraday components, where each is computed by holding the factor during the corresponding time window. Panel B further decomposes the risk premia into three components: the continuous part, overnight jumps, and intraday jumps, based on the real-time Fama–MacBeth regression specified in Equation (11). I report results for two systematic factors: (1) the high-frequency market excess return (Mkt-RF), and (2) the high-frequency GANs-based SDF constructed by Aleti and Bollerslev (2025), following the methodology of Chen et al. (2024). For each factor, I report the annualized risk premia (Ann RP) and their corresponding standard errors (Std Err), both in percentage points. Asterisks *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period spans from September 1997 to May 2020.

Panel A: Risk Premia Decomposition				
	Mkt-RF		SDF	
	Ann RP(%)	Std Err(%)	Ann RP(%)	Std Err(%)
Overnight	7.68***	(2.31)	6.63***	(2.04)
Intraday	-0.25	(3.39)	9.48**	(3.74)

Panel B: Jump Risk Premia Decomposition				
	Mkt-RF		SDF	
	Ann RP(%)	Std Err(%)	Ann RP(%)	Std Err(%)
Continuous	2.11	(3.01)	10.37*	(5.52)
$\text{Jump}_{\text{Overnight}}$	7.96**	(3.41)	9.31**	(4.11)
$\text{Jump}_{\text{Intraday}}$	0.51	(2.64)	4.01	(3.59)

Table 5 Regression against Risk Factors

This table presents the results from regressions of topic-based hedging portfolios on standard risk factors. Two test assets are evaluated: (i) a hedging portfolio constructed to offset exposure to the macroeconomic topic risk, and (ii) a real-time portfolio that dynamically invests in the topic associated with the most significant t-statistic from the Fama-MacBeth regression each month. For each specification, we report the monthly abnormal return (α), expressed in percentage points, in the first row. Regressions incrementally control for the Fama and French (2015) factors—market excess return (Mkt-RF), size (SMB), value (HML), profitability (RMW), investment (CMA), as well as momentum (MOM). Standard errors are reported in parentheses. Asterisks *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample spans January 2007 to May 2020.

	Macro Topic			Real-time Topic		
	(1)	(2)	(3)	(4)	(5)	(6)
Alpha (%)	0.30***	0.29***	0.25***	0.39***	0.31***	0.27**
	(0.10)	(0.10)	(0.09)	(0.12)	(0.11)	(0.10)
Mkt-RF		0.01	0.02		0.11***	0.11***
		(0.04)	(0.03)		(0.04)	(0.03)
SMB			0.15***			0.11***
			(0.02)			(0.04)
HML			-0.11***			-0.09**
			(0.03)			(0.04)
RMW			-0.05			0.04
			(0.07)			(0.10)
CMA			0.17***			-0.10
			(0.06)			(0.11)
MOM			0.04*			-0.01
			(0.02)			(0.05)
Months	161	161	161	161	161	161

Table 6 Counts of Macro News Categories Triggering U.S. Market Jumps

This table presents the category counts of the main macro indicator explaining the market jump within the macroeconomic topic. The macroeconomic topic is identified using DeepSeek-R1 with **Prompt 1** to **Prompt 3**. The sample spans from September 1997 to May 2020.

High-level Categories	Indicator Sub-category	Number of Occurrences
Labor Market	Non-farm Payroll / Employment Situation	34
	Weekly Initial Jobless Claims	13
	ADP Private Payrolls	3
	Subtotal	50
Inflation	Consumer Price Index (CPI)	13
	Producer Price Index (PPI)	7
	Employment Cost Index (ECI)	4
	Import/GDP Price Deflators	3
	Subtotal	27
Growth and Real Activities	Retail Sales	12
	Durable Goods Orders	9
	GDP	8
	Industrial Production / Capacity Utility	4
	Housing Starts / Building Permits	3
	Manufacturing Surveys	2
	Trade Balance	1
	Subtotal	39
Financial Conditions	Credit Spread	5
	Subtotal	5
Sector Specific	Semiconductor Sales	1
	Subtotal	1
	Total	153

Online Appendix for
"Interpretable Systematic Risk around the Clock"

Not for Publication

A Additional Results

Proof of Pure-play Property of Fama-MacBeth Hedging Factors:

Proof. Each Fama-MacBeth cross-section regression regresses the excess returns dR_t on an intercept and lagged βs. Writing the stacked design matrix as $\beta_t = [\mathbf{1}, \beta_t^C, \beta_t^J]$.

The portfolio weighting matrix W_t satisfies:

$$W_t' \beta_t = I_{K+2}.$$

By design, the j-th column of W_t satisfies Equation (4). Therefore, the portfolio $w_j' dR_t$ is the return on a portfolio purely driven by exposure to the corresponding risk and is immunized against all other sources of risk. □

B Hosting and Serving DeepSeek-R1

This section describes the deployment and serving setup for the DeepSeek-R1-0528 language model, which is used throughout this study to analyze high-frequency financial news and attribute market jumps.

Model Overview

DeepSeek-R1-0528 is a 685-billion parameter Mixture-of-Experts (MoE) reasoning large language model (LLM), released by DeepSeek and currently the top-ranked open-weight model in reasoning on the LLM Arena Leaderboard.[8] Its architecture is designed for strong logical reasoning, robust retrieval, and high-capacity multitask learning. In this paper, I utilize the AWQ-quantized version, which offers high inference throughput with reduced memory requirements, making it well-suited for self-hosting at scale. The model supports a native context window of up to 128K tokens, allowing input of all contemporaneous market news at the jump time.

Thinking and Reasoning

Beyond sheer scale, DeepSeek-R1's performance stems from a training pipeline that *explicitly rewards reasoning*. After a brief supervised "cold-start" stage, the model is optimised end-to-end with **Group-Relative Policy Optimisation** (GRPO), a variant of PPO run over large number of RL steps. The reward signal scores *intermediate* chain-of-thought traces on diverse math, code, and logic tasks, encouraging the policy to "think out loud" and converge on verifiably correct solutions.[9]

Because the objective values answer quality more than brevity, the learned policy naturally allocates *more tokens—and therefore more FLOPs*—to harder questions, mirroring the compute-adaptive effects first noted by Wei et al. (2022). In this study we exploit that capability by enabling more thinking tokens for market jump reason attribution. With more thinking and reasoning, DeepSeek-R1 can weigh multiple candidate narratives, evaluate temporal precedence and sentiment consistency, and output the most plausible driver of the market jump.

[8] See `https://lmarena.ai/leaderboard` for live rankings.
[9] See Guo et al. (2025) for details.

Deployment Environment

The model is hosted using vLLM, an open-source, high-throughput inference engine tailored for large transformer models. I use the AWQ version of DeepSeek-R1-0528, which enables efficient quantized inference.

The deployment is carried out on a cluster with eight NVIDIA H100 GPUs, each with 80GB of VRAM, providing a total of 640GB of GPU memory. This hardware is required to load the full DeepSeek-R1-0528-AWQ model and serve concurrent inference requests at low latency.

Model Loading and Serving Details

The hosting workflow consists of the following steps:

1. **Model Download:** The quantized model weights are downloaded from Hugging Face (`https://huggingface.co/cognitivecomputations/DeepSeek-R1-0528-AWQ`) and stored locally.
2. **Environment Setup:** The serving environment uses Python 3.10+, CUDA 12.x, and installs dependencies via `pip install vllm`.
3. **Launching the Server:** The vLLM engine is initialized with

   ```
   python -m vllm.entrypoints.openai.api_server \
     --model /data/models/dsr1-0528-awq \
     --served-model-name deepseek-r1-0528 \
     --tensor-parallel-size 8 \
     --pipeline-parallel-size 1 \
     --max-model-len 131072 \
     --gpu-memory-utilization 0.95 \
     --enable-chunked-prefill \
     --enable-prefix-caching \
     --trust-remote-code \
     --port 8000 --host 0.0.0.0
   ```

   Here, `tensor-parallel-size 8` enables model parallelism across all 8 H100 GPUs.
4. **API Serving:** vLLM serves an OpenAI-compatible API endpoint on `localhost:8000` (by default). This endpoint is used for all downstream inference tasks described in Section 2.3.

This hosting and serving setup ensures high-throughput, low-latency LLM inference for complex, high-frequency market news analysis, and is fully open for academic replication within reasonable hardware constraints.

C Processing High-frequency Data

C.1 Cleaning of Raw Trade and Quote Data

My intraday sample covers all NYSE, AMEX, and NASDAQ common stocks from 09/01/1997 through 05/31/2020.[10] Because the millisecond data do not begin until late 2003, I use MTAQ for 09/01/1997–09/09/2003 and DTAQ thereafter. The identical cleaning filters are applied to each feed. To construct a 15-minute panel of equity returns, I apply the following steps:

Step 1: Trade & quote filters. Starting from the raw TAQ millisecond files, I follow the procedures of Holden and Jacobsen (2014) and Da and Xiu (2021) to filter the raw trade and quote records.

- *Quote records* – keep only the six "regular-market" condition codes (`Qu_Cond`) `A,B,H,O,R,W`; drop all quotes flagged as cancelled (`Qu_Cancel = B`). For the timing range, we keep 09:00–16:00 ET so that the opening National Best Bid and Offer (NBBO) is available at 09:30 ET.
- *Trade records* – retain original, uncancelled prints (`Tr_Corr = 00`) and the immediately corrected versions (`01`); discard all other correction codes.

Step 2: NBBO construction and trade matching. We reconstruct the official National Best Bid and Offer (NBBO) by combining TAQ's `NBBOM_` and `CQM_` snapshots, ordering by descending sequence number, and removing duplicate micro-second stamps to leave exactly one quote per time-stamp. Quotes with spread $>$ \$5 or bid $>$ ask are likewise deleted. Each trade is then paired with the NBBO that was in force one nanosecond earlier; any trade whose subsequent NBBO is locked or crossed is discarded.

Step 3: Extreme-value filters. Matched trades whose prices fall outside the day's NBBO range are removed.

Step 4: Sampling to 15-minute frequency. Prices are sampled on an equal-spaced 15-minute grid from 09:30 to 16:00 ET using the cleaned trade data. The last trade observed at or before each grid point is carried forward (previous-tick method).

The procedure yields an extensive panel of 15-minute stock prices from high-quality trade prices aligned to the prevailing NBBO with minimal contamination from stale quotes, odd lots,

[10] Daily TAQ ("DTAQ")—time-stamped to the millisecond—first becomes available on WRDS on 09/10/2003, whereas the earlier Monthly TAQ ("MTAQ") product provides second-level stamps beginning 01/01/1993.

corrections, or extreme outliers.

C.2 Linking TAQ to CRSP

The next step is to link this high-frequency panel to CRSP, which provides the overnight returns necessary to construct around-the-clock returns for all stocks. In this section, I describe in detail the procedures used to merge the TAQ database with CRSP and outline the steps taken to verify the quality of the linkage.

I first filter the CRSP stocks to consider only ordinary common shares listed on the three primary U.S. exchanges (exchcd $\in \{1\text{ (NYSE)}, 2\text{ (AMEX)}, 3\text{ (NASDAQ)}\}$, shrcd $\in \{10, 11\}$). After filtering, I merge each security to the TAQ using the CUSIP within the valid date ranges.

Finally, to further clean the merged database, I compare the intraday trade prices against the `ASKHI` and `BIDLO` from CRSP. Any intraday price that falls outside the interval $[\texttt{BIDLO}, \texttt{ASKHI}]$ is set to missing, as such violations almost surely reflect data errors rather than true trades. Missing prices are then *forward-filled* from the most recent valid observation to preserve the regular 15-minute grid.

For daily open and close prices, I use the CRSP values in place of those from TAQ. I also use the CRSP information to calculate the overnight returns for stocks that account for stock mergers and splits, as well as returns from dividend payments.

To verify the quality of the link between the two databases, I follow Aït-Sahalia et al. (2020) to build high-frequency Fama-French factors and compare the return series against the daily frequency factors downloaded from Kenneth French's website at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Specifically, I follow the open-source replication of Fama-French factor construction from Freda Song at https://www.fredasongdrechsler.com/data-crunching/fama-french to first obtain the Fama-French factor portfolio constituents over time.

After this, I use the constructed linking table to merge the portfolio constituents to high-frequency return observations and construct the six Fama-French factors: Mkt-RF, SMB, HML, RMW, CMA, and MOM, using two-by-three sorting with cutting threshold and portfollio rebalancing rules following exactly the same procedure as Fama and French (2015) and Fama and French (2018).

Finally, I aggregate the constructed high-frequency factors to daily frequency and compare them against the official version from Kenneth French. If the linking is successful for most stocks

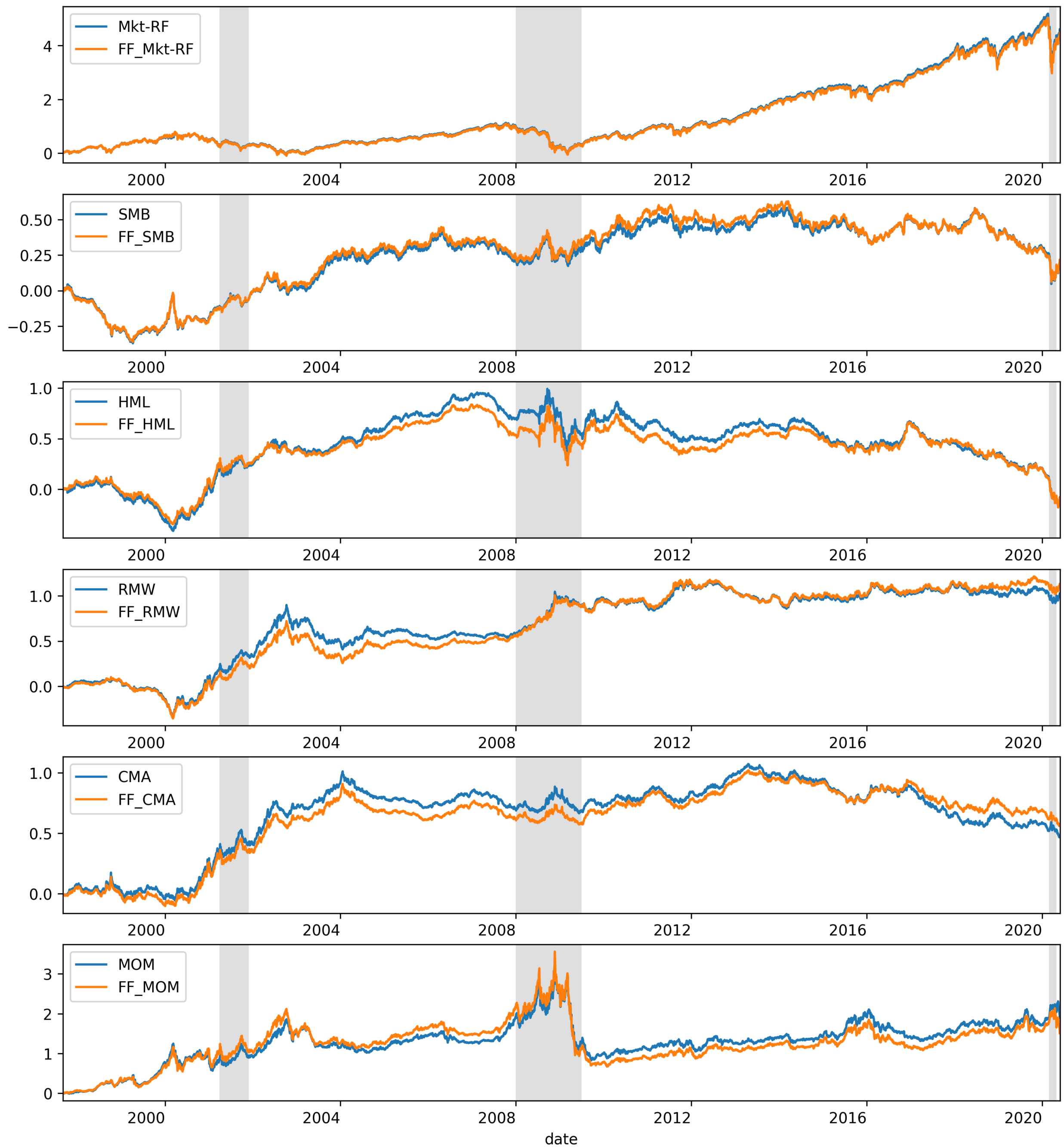

Figure A1 Comparison of High-Frequency Fama-French Factors against Low-Frequency Counterpart

This figure plots the cumulative returns of the high-frequency Fama-French factors against the daily frequency version from Kenneth French's data library. The factors considered include: Mkt-RF, SMB, HML, RMW, CMA, and MOM. The high-frequency factors are aggregated to daily frequency for comparison with the low-frequency counterpart. The sample period is from 09/01/1997 to 05/31/2020.

used to construct the factors, the two versions should closely align with each other. Otherwise, if there are large amounts of mismatch or unmatched stocks, there will exist large discrepancies between the two.

Figure A1 presents the cumulative returns of the high-frequency Fama-French factors versus the daily counterpart. From the figure, I find a close alignment of the two, suggesting that the matching of the two datasets is of high quality.

Furthermore, the paired return series also exhibit nearly perfect correlations. I find the correlation coefficients between the high-frequency factor and daily counterpart for the six factors (Mkt-RF, SMB, HML, RMW, CMA, and MOM) are: 0.9996, 0.9942, 0.9648, 0.9783, 0.9624, and 0.9593, respectively.

C.3 Cleaning and Rollover of the S&P 500 E-mini Futures

My analysis of overnight market jumps is based on high-frequency data from the S&P 500 E-mini futures, which trade nearly 24 hours a day and thus provide a natural proxy for the market's pricing of overnight risk. To construct a continuous time series of futures prices, I roll over the futures with different time to maturities.

Rather than rolling on a fixed calendar date, the front contract is rolled on the first trading day after the second-nearest contract becomes more liquid than the current front contract, where liquidity is measured by daily trading volume (or by trade count when early-sample volume is missing). Once the book is rolled forward, contracts never move "backwards", preserving chronological ordering of front, second, and third positions.

Because the second contract is typically priced above or below the expiring front contract, we scale position size on the roll date to keep notional exposure unchanged. If n_t front contracts priced at f_t^1 are replaced by the second contract priced at f_t^2, the new position size is:

$$n_{t+1} = n_t \frac{f_t^1}{f_t^2}.$$

This guarantees that the strategy is self-financing and the overnight roll return from t to $t+1$ is simply $\frac{f_{t+1}^1}{f_t^2} - 1$.

The resulting high-frequency futures factor tracks the S&P 500 cash index extremely closely: the contemporaneous correlation in overlapping intraday windows is 0.9589. Such a high correlation confirms that the rolled E-mini series captures the same aggregate risk as the underlying index—including overnight price discovery—and is therefore a reliable instrument for studying market-wide jumps.

Figure A2 compares the cumulative returns of the E-mini futures with those of the spot market. The left panel displays total cumulative returns, while the right panel restricts to intraday periods with overlapping high-frequency observations. In both cases, the futures closely mirror the returns of the aggregate equity market. Notably, the cumulative return over the overnight period is nearly flat, underscoring the role of overnight risk in shaping equity risk premia.

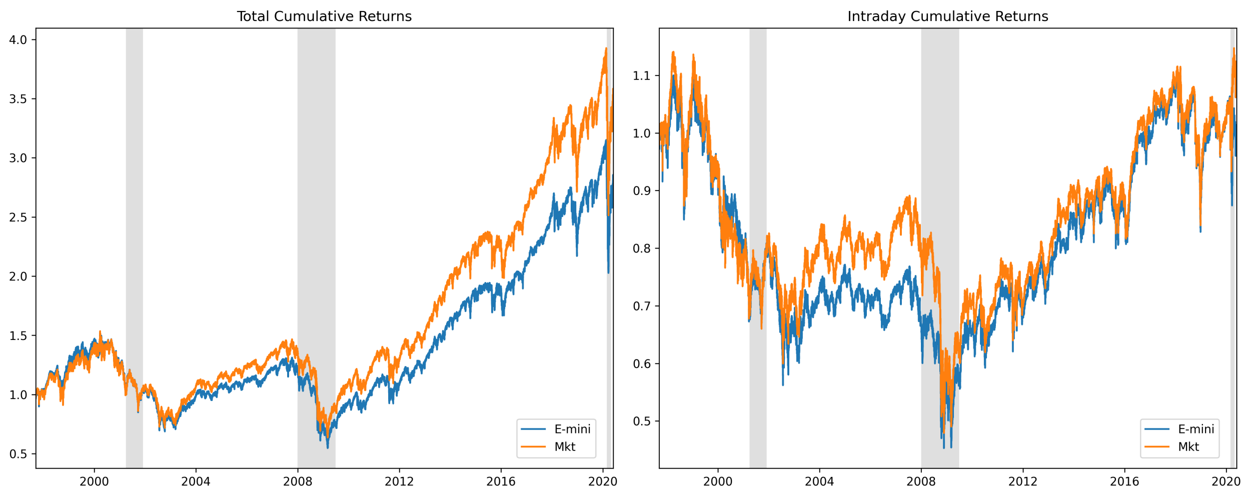

Figure A2 Comparison of Cumulative Returns of E-mini Futures against the Spot
This figure compares the cumulative returns of S&P E-mini futures and the spot market. The left panel plots the total cumulative returns using all available observations. The right panel shows cumulative returns using only intraday observations where both series have overlapping high-frequency data. The shaded gray areas indicate NBER-dated recessions. The sample period spans from September 1997 to May 2020.