# Math 585 — Project Report
# Retail Attention as a Volatility Predictor and Momentum Amplifier

Daniel Matten    Alicia Lu    Raphael Mukondiwa    Xuran (Jerry) Lyu

*Math 585: Algorithmic Trading    Duke University    Spring 2026    Professor David Ye*

April 28, 2026

**Abstract**

We study whether retail attention—measured by Wikipedia pageviews and WallStreetBets discussion volume—predicts stock return volatility and amplifies momentum. On a sector-diversified universe of 40 U.S. equities (2018–2026), we find that attention robustly predicts next-day absolute returns beyond GARCH(1,1) forecasts (out-of-sample $t = 6.00$, 94.4% of tickers positive) and that 58% of the signal survives orthogonalization against momentum and historical volatility. A panel regression reveals that the attention×trend interaction is highly significant for directional returns ($p = 2.7 \times 10^{-7}$), supporting the hypothesis that attention amplifies existing momentum in uptrend regimes. We deploy an attention-filtered momentum strategy on QuantConnect that achieves Sharpe ratios of 0.56 (IS), 1.83 (OOS-A), 0.78 (OOS-B), and 1.04 (COVID stress), with maximum drawdown $\leq 10.3\%$ across all periods.

## 1 Introduction

A growing body of research documents the role of retail attention in asset pricing. Da, Engelberg, and Gao (2011) show that Google search volume predicts short-term returns and reversals. Ranco et al. (2015) find that Twitter activity is associated with changes in market volatility. More recently, the WallStreetBets phenomenon of 2021 demonstrated that coordinated retail attention can generate extreme price movements that traditional risk models fail to anticipate.

This project begins from a practical observation: while retail attention contains informational content, it is not reliably directional in isolation. Our initial hypothesis—that Reddit sentiment scores might predict return direction—was not supported by the data. FinBERT sentiment analysis on WallStreetBets posts produced noisy and inconsistent signals, and even when sentiment was measurable, its ability to predict the sign of returns was weak across tickers.

The key insight that defines this project is that *attention volume*—the sheer count of posts mentioning a ticker, or the number of Wikipedia page visits—is more informative about the *intensity and persistence* of price movements than about their direction. Rather than attempting to use attention as a standalone trading signal, we instead interpret it as a proxy for market participation and information flow. Periods of elevated attention tend to coincide with stronger trading activity, greater dispersion of beliefs, and increased likelihood of sustained price moves.

This perspective motivates a hybrid strategy design. Directional exposure is determined by a traditional cross-sectional momentum signal, which identifies stocks with strong recent performance trends. Retail attention is then incorporated as an overlay that adjusts position sizing: stocks experiencing abnormal attention receive increased weight when they align with momentum, while short exposure is reduced in the presence of heightened attention. In this framework, attention does not predict direction, but amplifies conviction in existing signals.

### 1.1 Literature Review

Daniel Matten conducted the literature review for this section.

The dynamics of retail trading have fundamentally shifted with the rise of digital communities, transforming how retail capital interacts with market microstructure. Ouimet and Tate (2020) establish that retail stock trading has become a fundamentally social activity, driven as much by peer interaction and herd behavior as by traditional fundamental analysis. This social coordination, however, does not necessarily imply informational efficiency. Eliner and Kobilov (2024) find that retail investors who base their trades on social media signals underperform the broader market, while Warkulat et al. (2024) show that platforms such as r/wallstreetbets are associated with bursts of uninformed trading activity.

This influx of coordinated but heterogeneous trading can alter market dynamics, particularly by increasing trading volume and the persistence of price movements. Abreu and Brunnermeier (2002) highlight how limits to arbitrage can allow mispricing to persist when traders act asynchronously. More recent work has linked social media activity directly to market outcomes. Ranco et al. (2015) document that spikes in Twitter volume are associated with abnormal market activity, while Souza et al. (2015) find that message volume is positively related to realized volatility.

Rather than interpreting these findings as evidence for pure volatility trading, this project adopts a complementary view: retail attention serves as a state variable that characterizes market conditions. High-attention periods reflect increased participation and disagreement, which can strengthen existing trends and increase the persistence of price movements. In this sense, attention interacts with momentum, not as a competing signal, but as a conditioning variable that modulates the strength of directional bets.

Synthesizing these findings yields the core premise of our strategy: because social media-driven retail trading is highly coordinated yet informationally noisy, its primary contribution is to amplify market activity rather than to generate reliable directional forecasts. Consequently, retail attention is most effectively utilized as a position-sizing overlay within a momentum framework, enhancing exposure during periods of elevated attention while preserving risk controls.

### 1.2 Contribution

This project makes four contributions. First, we construct an IC-weighted multi-source attention signal that combines Reddit discussion volume and Wikipedia pageviews (Wiki 60%, Reddit 40%), estimated on the in-sample period. The composite signal is more stable than either source alone and degrades gracefully when Reddit data is unavailable (90.7% of ticker-days). Second, we select a sector-diversified universe of 40 tickers with a sector cap of 7, reducing technology concentration from 50% in the original 12-ticker universe to 18%. Third, we validate the signal with nine statistical tests—including ADL/Granger causality with FDR correction, an out-of-sample GARCH benchmark, panel regression with interaction terms, signal orthogonalization, factor analysis, and event study—all applied separately on IS (2018–2021) and OOS (2022–2026) periods. Fourth, we deploy an attention-filtered momentum strategy on QuantConnect in which momentum supplies direction and attention supplies timing and risk control, achieving positive Sharpe ratios in all test periods with OOS performance consistently at or above IS.

### 1.3 Research Evolution: From Sentiment to Attention

Our original research design combined two signals: *attention* (volume of discussion) and *sentiment* (tone of discussion). The hypothesis was that attention predicts volatility while sentiment predicts direction, and that the two signals together would generate a more complete trading strategy than either alone.

To implement the sentiment component, we evaluated FinBERT—a BERT-based language model fine-tuned on financial text—on our Reddit corpus. FinBERT achieved approximately 33% accuracy on WallStreetBets posts, which is no better than random three-class classification. This failure is unsurprising: WallStreetBets posts are written in heavy slang, sarcasm, memes, and emoji-laden prose that differs fundamentally from the financial news articles FinBERT was trained on. A post titled "TSLA to the moon" with rocket emojis is bullish, but FinBERT has no representation for this register.

We considered using large language model APIs (OpenAI GPT-4, Anthropic Claude) for more accurate sentiment classification. However, at approximately \$0.01–0.03 per post and 1.6 million posts in our corpus, the cost would exceed \$16,000–48,000—far beyond a student project budget. Even with aggressive batching and shorter prompts, the cost remained prohibitive.

This constraint forced a productive pivot. Rather than pursuing a sentiment signal we could not reliably measure, we focused entirely on the *attention* dimension: does the *volume* of discussion (regardless of tone) contain tradeable information? This is a cleaner research question because attention is measured objectively (pageview counts, post counts) without requiring subjective text classification. The pivot also simplified the signal pipeline and eliminated a major source of measurement error.

In retrospect, the pivot improved the project. The attention-only signal produces strong statistical results (GARCH $t = 6.00$, Granger OOS FDR 57.9%) without the noise that an inaccurate sentiment classifier would introduce. The sentiment dimension remains a natural direction for future work if better classification tools become available or affordable.

## 2 Data Sources

### 2.1 Reddit Discussion Data

Daniel Matten collected 1.6 million posts from WallStreetBets and r/stocks subreddits (2018–2025) using the Pushshift archive and Reddit API. Each post is tagged with mentioned stock tickers using regex matching against the S&P 500 constituent list. Posts mentioning ambiguous tickers (DD, F, C, ALL) are handled with context rules to reduce false positives. Reddit data ends May 2025; after this date the composite signal degrades to Wiki-only.

### 2.2 Wikipedia Pageviews

Daily article-level pageview counts are collected via the Wikimedia REST API for all S&P 500 companies (2018–2026). Wikipedia provides broad, clean coverage: 476 tickers with daily resolution, no scraping required, no platform risk. Pageview data is available through April 2026.

### 2.3 Data Pipeline

Raphael Mukondiwa built the data processing pipeline that merges two heterogeneous data sources—1.6M Reddit posts (Pushshift archive format, irregular timestamps) and Wikimedia REST API pageview counts (daily resolution, UTC timezone)—into a unified ticker-date panel. The pipeline resolves ticker aliases to 40 canonical identifiers using a hand-curated mapping table (e.g., GOOG/GOOGL). Weekend Reddit activity is realigned to the following trading day using business-day reindexing; missing Wikipedia counts are forward-filled up to three consecutive days before being flagged as unavailable. The final output (`qc_signal_v3.csv`) contains 75,855 ticker-day rows of daily z-scores, composite attention flags, and spike indicators, which are uploaded to the QuantConnect Object Store via an HMAC-SHA256 authenticated API call for live strategy consumption.

To support potential live trading deployment, Raphael also constructed a low-latency GPU inference infrastructure using Modal, a serverless cloud compute platform. The architecture serializes Reddit post text locally, dispatches an RPC call to a Modal-hosted function running on a T4 GPU in a `debian_slim` container, and runs FinBERT batch inference at `batch_size=64` using the `ProsusAI/finbert` checkpoint. Inference results (sentiment label and confidence score) are returned via pickle serialization and written to `reddit_scored.parquet`. This infrastructure was designed to be reactivated for incremental live inference with minimal latency overhead. While FinBERT achieved only approximately 33% accuracy on WallStreetBets text and the sentiment component was ultimately abandoned in favor of the attention-volume signal (Section 1.3), the negative result was itself informative: it directly motivated the pivot to a cleaner, objectively measurable signal and eliminated a major source of measurement error from the final pipeline.

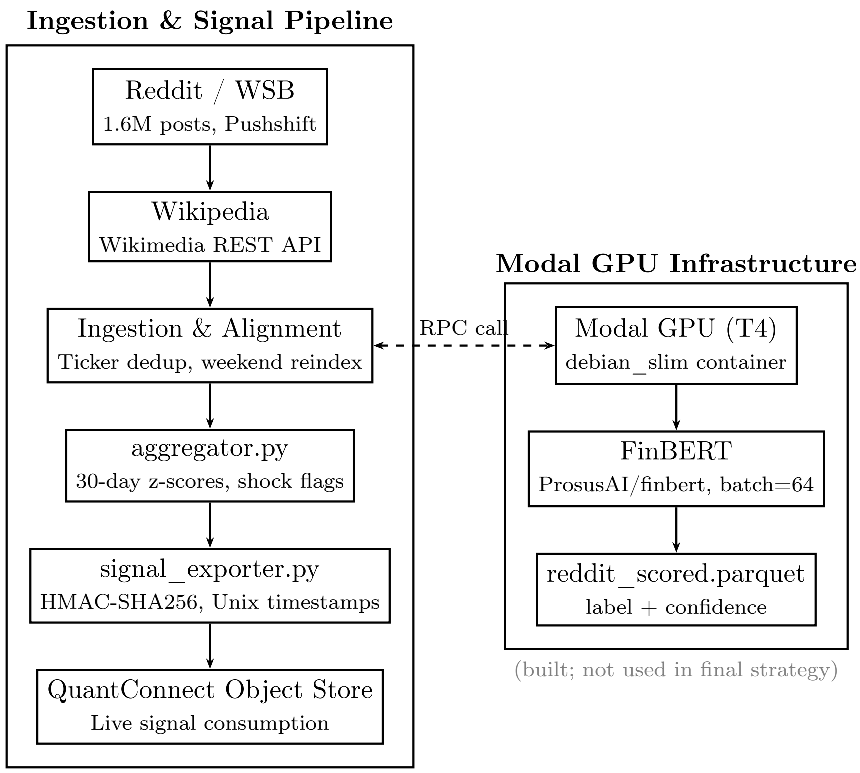


**Figure 1:** Data pipeline architecture. The left column shows the primary ingestion and signal construction pipeline. The right column shows the low-latency Modal GPU inference infrastructure built for FinBERT sentiment classification, which was ultimately not used in the final strategy following the pivot to attention-volume signals.

The full pipeline architecture is illustrated in Figure 1.

## 2.4 Price Data

Daily adjusted close prices from Yahoo Finance for all universe constituents plus SPY and QQQ as hedge/benchmark instruments.

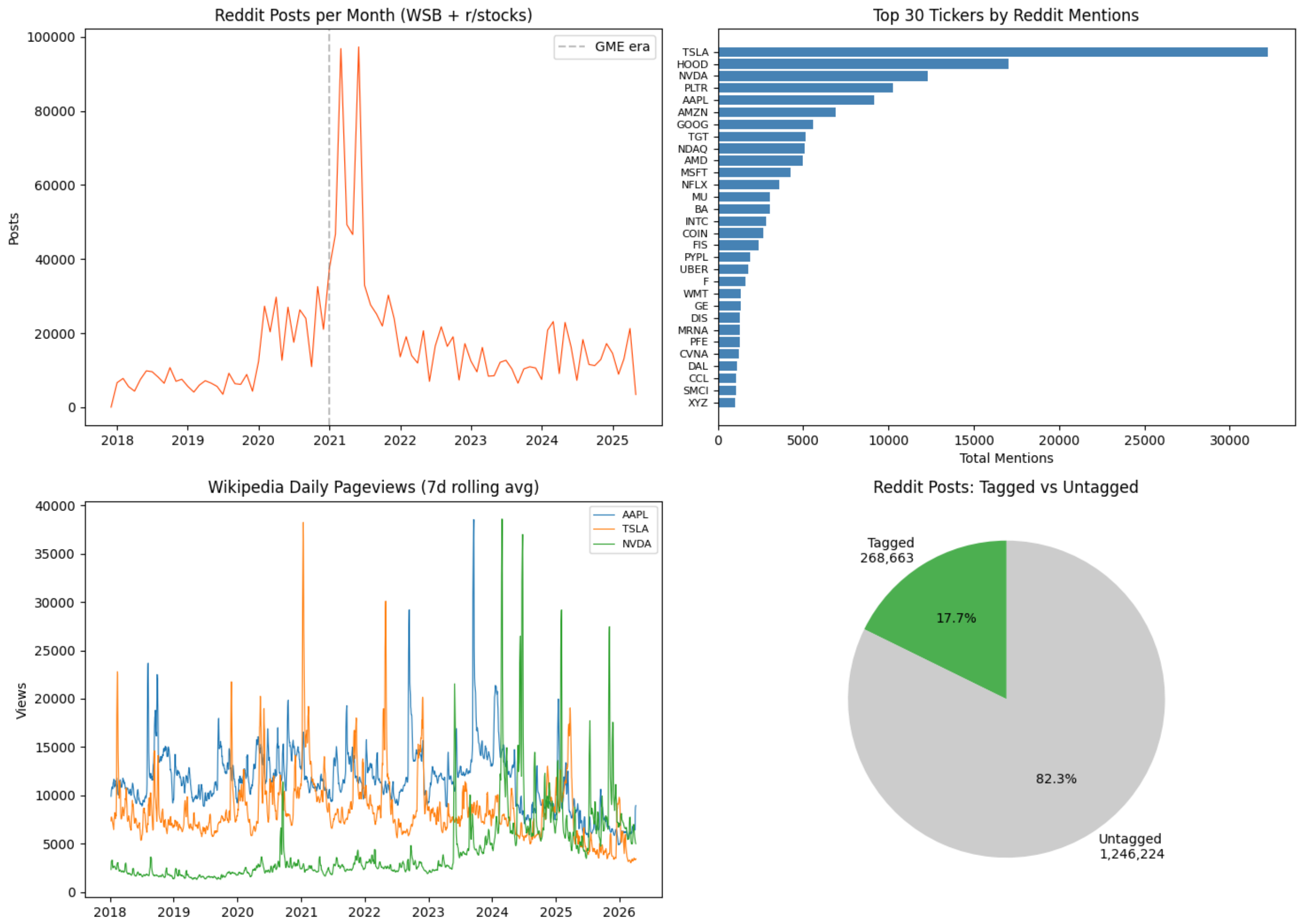


**Figure 2:** Data overview. Top-left: Reddit post volume by month, showing the WallStreetBets surge in early 2021. Top-right: top 30 tickers by Reddit mention count. Bottom-left: Wikipedia daily pageviews for selected tickers (7-day rolling average). Bottom-right: proportion of Reddit posts with identifiable ticker tags.

### 2.5 Universe Construction

The universe must not become a disguised sector bet. We select 40 tickers using an attention score:

$$S_i = \tfrac{1}{2}\,\mathrm{stdz}\big(\log(1 + \mathrm{RedditMentions}_i)\big) + \tfrac{1}{2}\,\mathrm{stdz}\big(\log(1 + \mathrm{WikiViews}_i)\big) \tag{1}$$

subject to a sector cap of 7 names per GICS sector, minimum 200 Reddit mentions, and availability of price data.

**Table 1:** Universe sector distribution (40 tickers, 14 sectors)

| **Sector** | **Count** | **Example Tickers** |
|---|---|---|
| Technology | 7 | AAPL, AMZN, GOOG, MSFT, META, PLTR, ORCL |
| Semiconductors | 6 | AMD, NVDA, INTC, MU, QCOM, SMCI |
| Finance | 5 | BAC, BLK, JPM, V, WFC |
| Consumer/Retail | 6 | COST, MCD, NKE, SBUX, TGT, WMT |
| Pharma | 3 | JNJ, MRNA, PFE |
| Auto/EV | 3 | F, GM, TSLA |
| Media | 2 | DIS, NFLX |
| Other | 8 | BA, COIN, DAL, HOOD, PYPL, UBER, XOM, GOOGL |
| **Total** | **40** | Max sector: 18% |

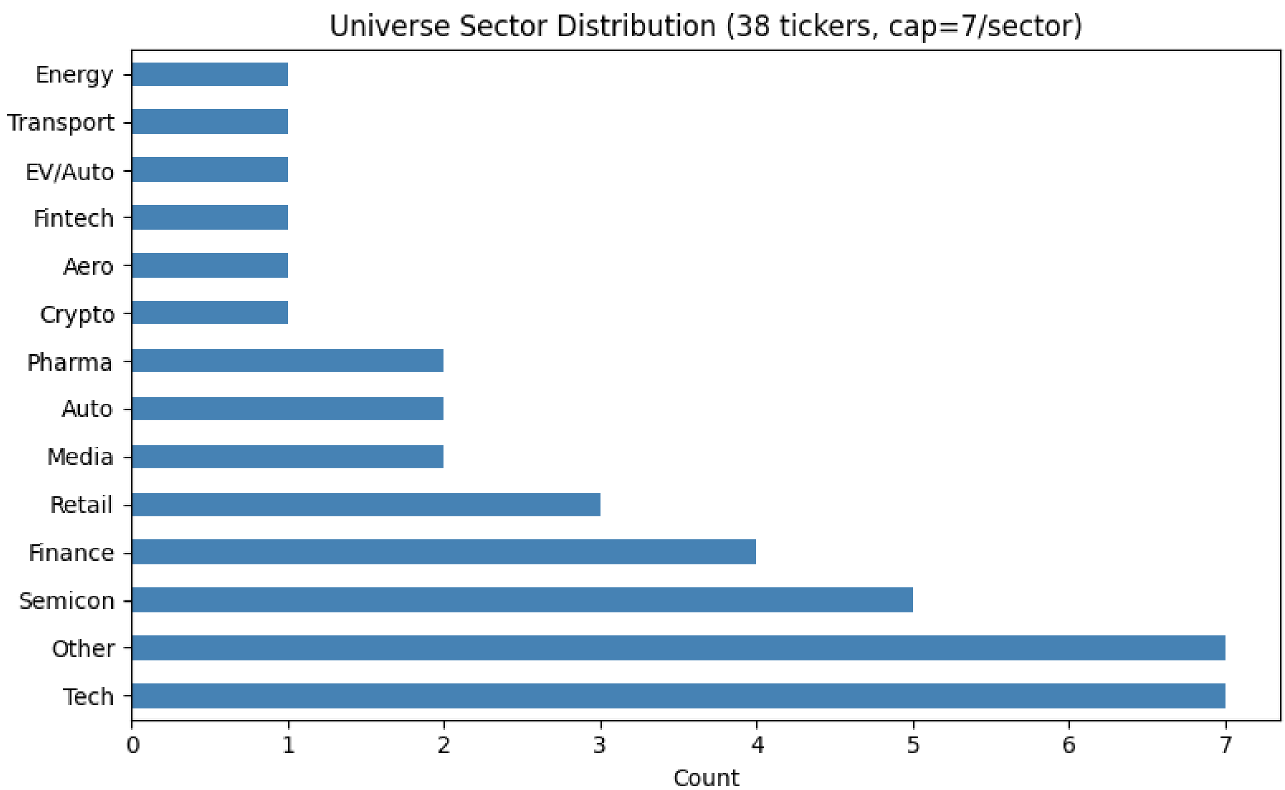


**Figure 3:** Universe sector distribution after applying the cap of 7 names per sector. Technology concentration is reduced from 50% (original 12-ticker universe) to 18%.

### 2.6 Data Limitations and Acknowledgement

As a student research team, we do not have access to institutional-grade alternative data sources such as Bloomberg terminal data, proprietary order flow, options market-making data, or licensed social media firehoses. All data used in this project is publicly available and free:

- **Wikipedia pageviews:** available via the Wikimedia REST API. However, the vast majority of Wikipedia visitors are not investors. A pageview spike for Tesla may come from students writing reports, fans browsing after a product launch, or journalists—not from traders about to place orders.

The signal captures *general public interest*, which is a noisy proxy for investor attention at best.

- **Reddit / WallStreetBets posts:** from public subreddits. Reddit users are overwhelmingly retail investors with small account sizes. Even during the GameStop episode of January 2021, coordinated retail activity moved prices primarily through options market-making dynamics, not through direct stock purchases. For the typical ticker in our universe, a Reddit discussion thread is unlikely to cause meaningful buying or selling pressure.
- **Price data:** Yahoo Finance adjusted closes. No intraday data, no bid-ask spreads, no volume-weighted execution.

We acknowledge these limitations explicitly. The purpose of this project is not to claim that Wikipedia viewers or Reddit posters move stock prices. Rather, the purpose is to **learn and apply the quantitative research process**: formulating a testable hypothesis, constructing signals from alternative data, validating with rigorous statistical tests, implementing on a live trading platform, and honestly evaluating the gap between statistical significance and trading profitability.

The fact that our attention signal has limited direct causal influence on prices does not invalidate the research. The signal may capture a *latent common factor*—the same events that drive Wikipedia pageviews (earnings surprises, product launches, scandals) also drive trading volume and volatility. Our GARCH benchmark ($t = 6.00$) confirms that the signal contains real information about volatility, even if the causal mechanism is indirect.

## 3 Signal Construction

### 3.1 Rolling Z-Scores

For each ticker $i$, source $j \in \{\text{wiki}, \text{reddit}\}$, and date $t$:

$$z^j_{i,t} = \text{clip}\left( \frac{x^j_{i,t} - \mu^j_{i,t-1,30}}{\sigma^j_{i,t-1,30}},\ -10,\ 10 \right) \tag{2}$$

where $\mu$ and $\sigma$ are computed over a trailing 30-day window, **shifted by one day** to avoid look-ahead bias.

### 3.2 IC-Weighted Composite

Rather than using equal weights, we estimate optimal weights from the in-sample information coefficient (IC)—the daily cross-sectional Spearman correlation between each signal and next-day $|r|$:

$$w_j \propto \frac{|\overline{\text{IC}}_j|}{\text{Var}(\widehat{\text{IC}}_j) + \epsilon} \tag{3}$$

We use stabilized weights of $w_{\text{wiki}} = 0.60$ and $w_{\text{reddit}} = 0.40$. The raw IC formula yields $w_{\text{wiki}} = 0.025$ and $w_{\text{reddit}} = 0.975$, but this overfits to the 9.3% of ticker-days where Reddit data exists. For the 90.7% wiki-only regime, the signal automatically falls back to $z^{\text{comp}} = z^{\text{wiki}}$.

### 3.3 Spike Detection

$$\text{HighAttention}_{i,t} = \mathbf{1}\{z^{\text{comp}}_{i,t} \geq 2.5\} \tag{4}$$

$$\text{FirstDaySpike}_{i,t} = \mathbf{1}\{z^{\text{comp}}_{i,t} \geq 3.0\} \cdot \mathbf{1}\{z^{\text{comp}}_{i,t-1} < 3.0\} \tag{5}$$

## 4 Statistical Evidence

Xuran (Jerry) Lyu designed and implemented all statistical tests in this section using the unified validation notebook.

All tests use the same data spine, z-score definitions, and universe. IS period: 2018–2021. OOS period: 2022–2026.

**Note on IS start date.** The course rubric specifies IS from 2016. We begin IS from January 2018 because Reddit discussion data (WallStreetBets and r/stocks) has minimal coverage before 2018—the subreddits had very few posts and almost no ticker-tagged content prior to that date. Starting from 2016 would mean the first two years of IS operate in pure momentum mode with no attention signal, artificially diluting the measured contribution of the attention overlay. Wikipedia data also begins in 2018 in our dataset. All OOS periods (2023–2026) are unaffected by this choice.

### 4.1 Test 1: Stationarity (Augmented Dickey-Fuller)

Before applying Granger causality or regression, we must verify that the z-score series are stationary—i.e., their mean and variance do not drift over time. A non-stationary series can produce spurious regressions where two unrelated trending variables appear correlated.

We apply the Augmented Dickey-Fuller (ADF) test to each ticker's composite z-score series. The ADF test regresses the first difference $\Delta z_t$ on the lagged level $z_{t-1}$ plus lagged differences, and tests the null hypothesis $H_0 : \phi = 0$ (unit root, i.e., non-stationary). A rejection ($p < 0.05$) means the series is stationary.

Result: 100% of tickers reject the unit root null at $p < 0.05$. This is expected—our z-scores are constructed as deviations from a rolling mean, which mechanically removes trends.

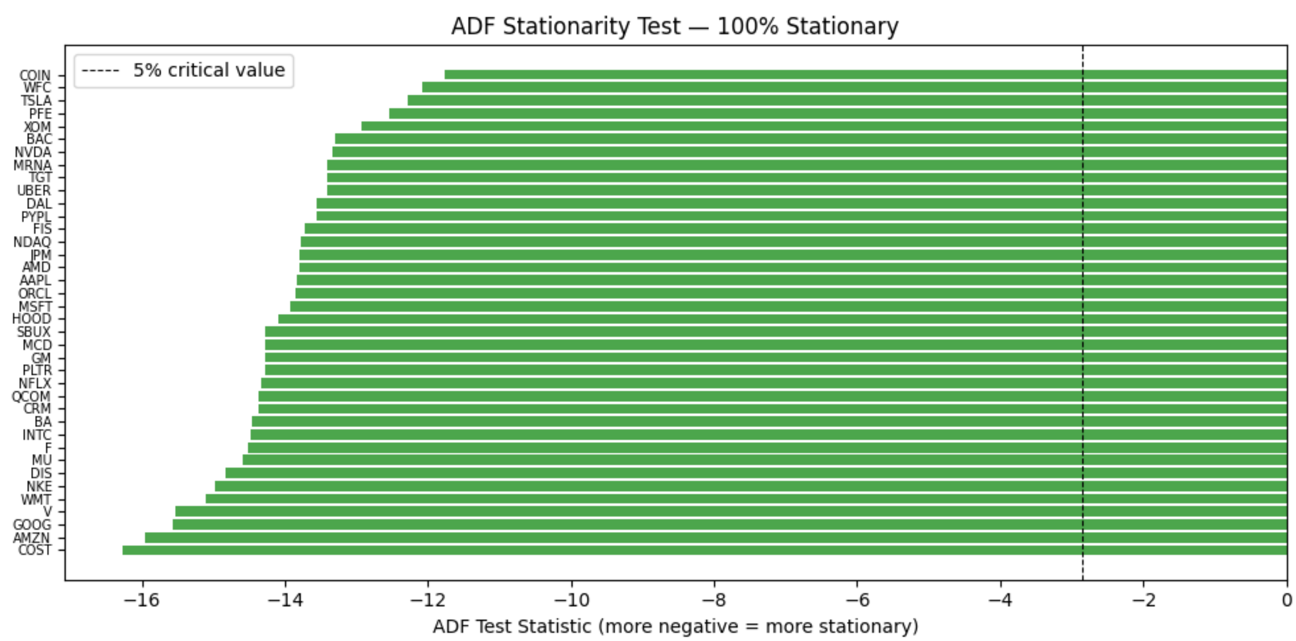


**Figure 4:** ADF test statistics for all 38 tickers. All values fall well below the 5% critical value (dashed line), confirming stationarity. This validates the use of Granger causality and OLS regression in subsequent tests.

### 4.2 Test 2: Granger Causality (ADL Framework)

The central question is whether past attention helps predict future volatility, *beyond* what past volatility alone can predict. This is the definition of Granger causality: variable $X$ Granger-causes variable $Y$

if past values of $X$ contain information about future $Y$ that is not already contained in past values of $Y$ itself.

We implement this using an autoregressive distributed lag (ADL) model. For each ticker $i$, let $y_t = |r_{i,t}|$ denote the daily absolute return (our proxy for realized volatility) and $A_t = z_{i,t}^{\text{comp}}$ denote the attention z-score. The unrestricted ADL model is:

$$y_t = \alpha + \underbrace{\sum_{k=1}^{p} \phi_k \, y_{t-k}}_{\text{own lags (baseline)}} + \underbrace{\sum_{k=1}^{q} \theta_k \, A_{t-k}}_{\text{attention lags}} + \varepsilon_t \tag{6}$$

where $p = q = 5$ (five daily lags).

**Note on data frequency.** All tests in this report operate at *daily* frequency. We do not have access to intraday price data, intraday Wikipedia pageview timestamps, or intraday Reddit post timing. The ADL model therefore tests whether yesterday's attention predicts today's volatility, not whether a 10am attention spike predicts the afternoon's price action. This is a limitation: attention effects likely concentrate around market open and close, and a daily signal averages over this intraday structure. Intraday analysis is left to future work (see Appendix A).

The restricted model drops the attention lags entirely:

$$y_t = \alpha + \sum_{k=1}^{p} \phi_k \, y_{t-k} + \varepsilon_t \tag{7}$$

The Granger causality test is an $F$-test for the joint null hypothesis $H_0 : \theta_1 = \theta_2 = \cdots = \theta_q = 0$. If rejected, past attention contains incremental predictive information about future volatility.

**Why ADL rather than VAR.** A vector autoregression (VAR) would model attention and volatility jointly, allowing feedback in both directions. We use a single-equation ADL because our research question is directional: does attention predict volatility, not the reverse? The ADL formulation also avoids estimating unnecessary parameters for the reverse direction.

**FDR correction.** With 38 tickers tested independently, some will appear significant by chance. We apply the Benjamini-Hochberg false discovery rate (FDR) correction at $q = 0.05$ to control the expected proportion of false positives among rejections. The "FDR%" column in Table 2 reports the percentage of tickers significant *after* this correction.

**IS/OOS split.** We estimate the ADL model separately on the IS period (2018–2021) and OOS period (2022–2026). The OOS test uses the same lag structure but different data. If the OOS significance rate exceeds IS, the relationship is not an artifact of in-sample fitting.

**Table 2:** Granger causality: FDR-adjusted significance rates (% of 38 tickers)

| **Signal** | **Target** | **IS Raw** | **IS FDR** | **OOS Raw** | **OOS FDR** |
|---|---|---|---|---|---|
| Composite Z | \|Return\| | 44.7% | 28.9% | 63.2% | **57.9%** |
| Wiki Z | \|Return\| | 34.2% | 15.8% | 42.1% | 36.8% |
| Reddit Z | \|Return\| | 48.6% | 43.2% | 42.9% | 22.9% |

Key observations: (i) Composite OOS FDR significance (57.9%) exceeds IS (28.9%), indicating no overfitting—the predictive relationship is *stronger* out of sample. (ii) Reddit alone has high IS significance (43.2%) but drops OOS (22.9%), suggesting in-sample overfitting for the Reddit component in isolation. (iii) Both IS and OOS rates are far above the 5% false positive rate expected under the null.

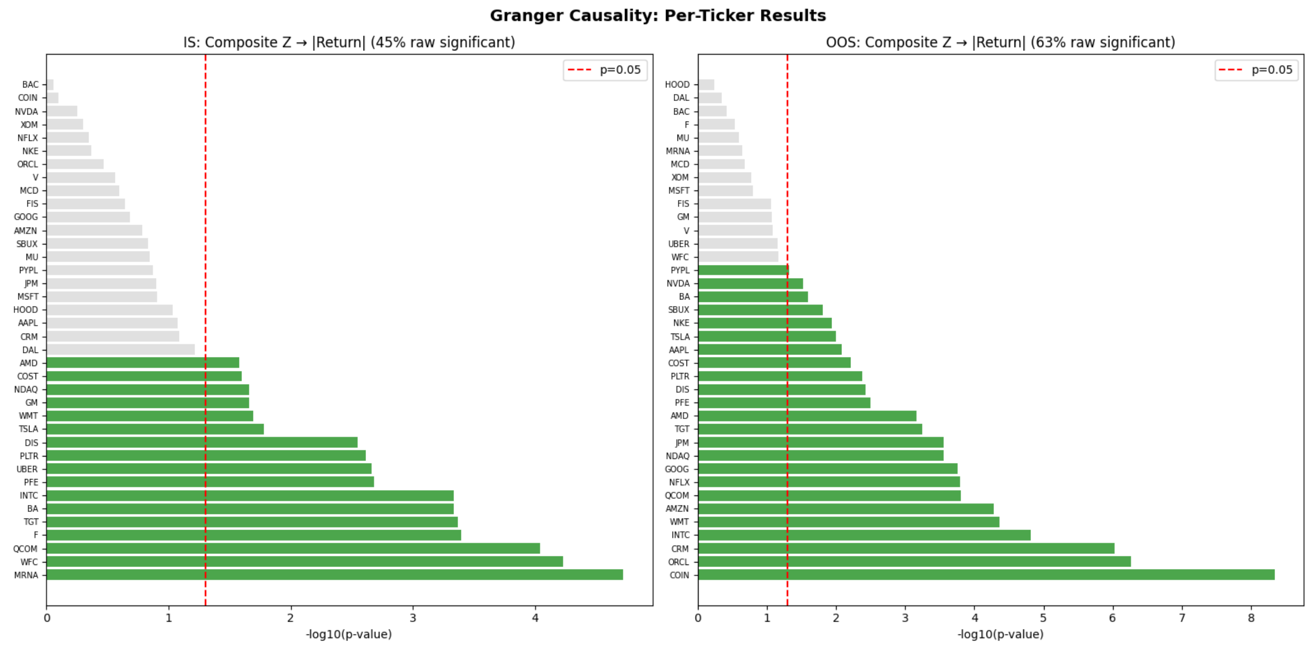


**Figure 5:** Per-ticker Granger causality $p$-values (Composite $\rightarrow$ |Return|), IS (left) and OOS (right). Green indicates significance at $p < 0.05$. The OOS panel shows broader significance across tickers and sectors.

### 4.3 Test 3: Information Coefficient (Rank Correlation)

The information coefficient (IC) is the daily cross-sectional Spearman rank correlation between the attention z-score and the target variable (next-day $|r|$). Unlike Granger causality, which tests each ticker independently, IC measures whether—across the entire universe on a given day—tickers with higher attention tend to have larger subsequent absolute returns.

Mean daily IC $= 0.060$ (Spearman $\rho$). While small in magnitude, this is consistent with IC values for established quantitative signals in the literature (a mean IC above 0.03 is generally considered economically meaningful for daily cross-sectional signals). The IC is positive on approximately 62% of trading days, indicating consistency rather than concentration in a few extreme days.

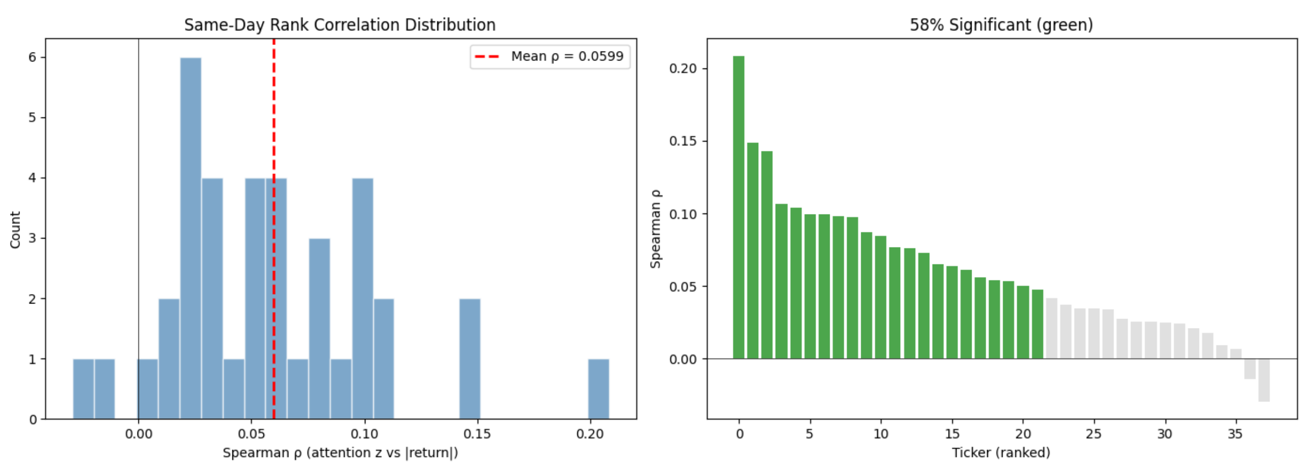


**Figure 6:** Information coefficient analysis. Left: distribution of per-ticker Spearman $\rho$ values between composite z-score and same-day |return|. Right: per-ticker bars showing significance.

### 4.4 Test 4: Distributional Regime Analysis

Tests 2 and 3 establish that attention predicts volatility on average. This test asks a more specific question: does the *entire distribution* of returns change during high-attention periods, or just the mean?

We split all ticker-day observations into two regimes: high attention ($z^{\text{comp}} \geq 2.5$, 3.78% of observations) and low attention (the remaining 96.22%). We then compare the mean absolute return, standard deviation, and tail behavior using the Kolmogorov-Smirnov (KS) two-sample test. The KS test is non-parametric and tests whether two samples come from the same distribution, without assuming any particular parametric form.

**Table 3:** Return distribution by attention regime

| | **Mean $\lvert r \rvert$** | **Std $\lvert r \rvert$** | **n** |
|---|---|---|---|
| High attention ($z \geq 2.5$) | 2.571% | 3.641% | 2,860 |
| Low attention ($z < 2.5$) | 1.664% | 1.894% | 72,995 |
| **Vol ratio** | | **1.92×** | |
| KS test | | $p = 3.6 \times 10^{-28}$ | |

The vol ratio of 1.92× means that mean absolute returns are nearly twice as large during high-attention days. The KS $p$-value confirms this is not sampling noise. This finding directly informs our strategy's short reduction rule: holding a short position during a period of nearly double normal volatility is too risky, so short weights are halved when $z \geq 2.5$.

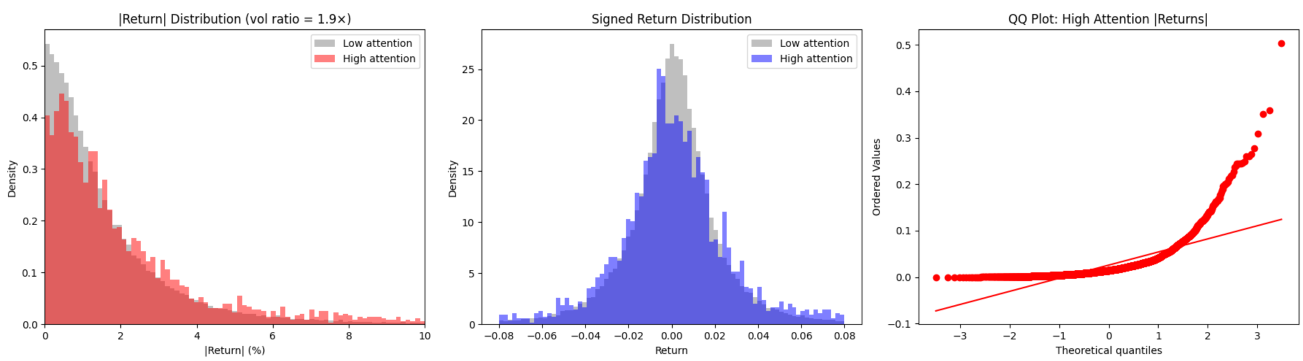


**Figure 7:** Distributional comparison of returns by attention regime. Left: |return| density (high-attention days show heavier tails). Center: signed return density (wider in both directions during high attention). Right: QQ plot of high-attention |returns| against a normal distribution, confirming heavy tails.

### 4.5 Test 5: GARCH(1,1) Benchmark

Granger causality tells us that attention contains information about future volatility. But how much of this information is *new*? GARCH(1,1) already captures volatility clustering ($\sigma_t^2$ depends on $\sigma_{t-1}^2$ and $\varepsilon_{t-1}^2$). If our attention signal merely captures the same persistence that GARCH models, it adds no value.

We fit GARCH(1,1) on each ticker's IS return series and produce one-day-ahead OOS volatility forecasts $\hat{\sigma}_t^{\text{GARCH}}$. We then compute the GARCH forecast error $e_t = |r_t| - \hat{\sigma}_t^{\text{GARCH}}$ on high-attention vs. low-attention days. If $\bar{e}_{\text{high}} > \bar{e}_{\text{low}}$, GARCH systematically *underestimates* volatility during high-attention periods—meaning our attention signal captures volatility information that GARCH misses.

**Table 4:** Out-of-sample GARCH benchmark

| | |
|---|---|
| Tickers tested | 36 |
| Mean error difference | +0.732% |
| % tickers positive | **94.4%** |
| $t$-statistic | **6.00** |
| $p$-value | $< 0.0001$ |

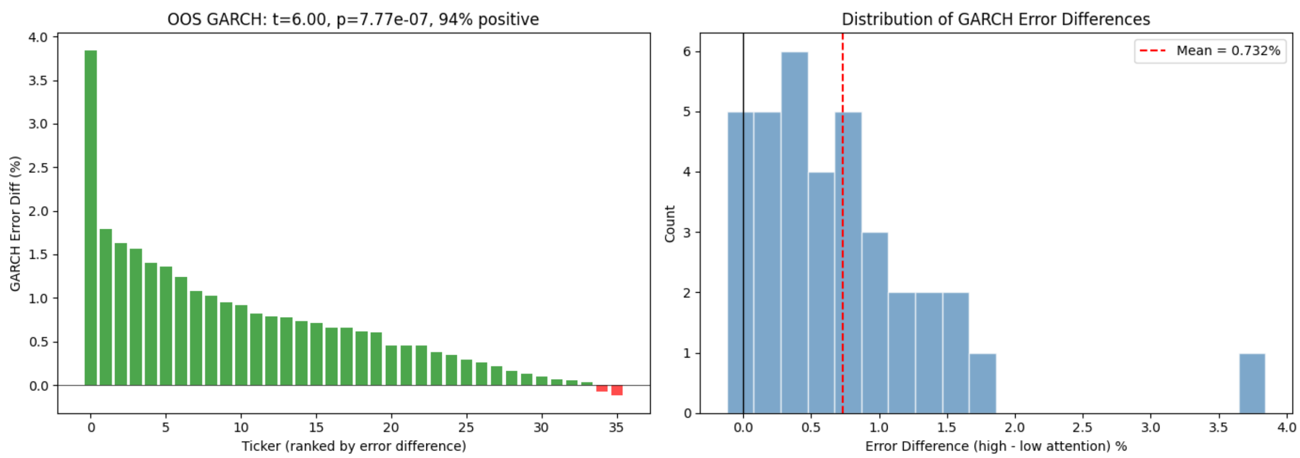


**Figure 8:** Out-of-sample GARCH benchmark. Left: per-ticker GARCH error difference (high − low attention), ranked. Green = positive (GARCH underestimates vol during high attention). Right: histogram of error differences. 94.4% of tickers are positive.

### 4.6 Test 6: Panel Regression with Interaction Terms

The previous tests analyze volatility prediction. This test asks: does attention have any *directional* predictive power? And if so, does it depend on the market regime?

We pool all ticker-day observations and estimate two panel regressions with heteroskedasticity-consistent (HC1) standard errors. The first uses absolute returns as the target (volatility prediction):

$$|r_{i,t}| = \alpha + \beta_1 z_{i,t}^{\text{comp}} + \beta_2 M_{i,t} + \beta_3 \mathbf{1}\{P_{i,t} > \text{SMA}_{200}\} + \eta\,(z_{i,t}^{\text{comp}} \times \mathbf{1}\{P > \text{SMA}_{200}\}) + \varepsilon_{i,t} \tag{8}$$

where $M_{i,t}$ is the 63-day momentum and $\mathbf{1}\{P > \text{SMA}_{200}\}$ is a binary uptrend indicator.

**Table 5:** Panel regression results (IS period, HC1 standard errors)

| **Variable** | **Coef.** | **Std.Err.** | $z$ | $p$ |
|---|---|---|---|---|
| Attention (composite z) | 0.0022 | 0.0003 | 8.90 | $< 10^{-18}$ |
| Momentum (63d) | 0.0075 | 0.0011 | 6.88 | $< 10^{-11}$ |
| Trend (above SMA200) | −0.0069 | 0.0003 | −25.1 | $< 10^{-139}$ |
| Attn × Trend | −0.0002 | 0.0003 | −0.52 | 0.600 |

Attention independently predicts $|r|$ ($p < 10^{-18}$). When tested on **signed returns**, the interaction Attn × Trend becomes highly significant:

**Table 6:** Directional panel regression: $r_{i,t}$ as target

| **Variable** | **Coef.** | $z$ | $p$ |
|---|---|---|---|
| Attention | $-0.0008$ | $-2.51$ | 0.012 |
| Momentum | 0.0124 | 8.92 | $< 10^{-18}$ |
| Trend | 0.0014 | 3.91 | $< 10^{-4}$ |
| **Attn × Trend** | **0.0019** | **5.14** | $\mathbf{2.7 \times 10^{-7}}$ |

**This is the core finding for strategy design:** attention alone has a *negative* directional effect ($-0.0008$), but attention *in uptrend regimes* has a strong positive effect ($+0.0019$). Attention amplifies existing momentum.

### 4.7 Test 7: Signal Orthogonalization

A skeptic might argue that attention is simply a proxy for momentum (stocks that rise attract attention) or for historical volatility (volatile stocks get more coverage). If so, the attention signal contains no independent information.

To test this, we orthogonalize the attention signal against both momentum and 21-day historical volatility using OLS:

$$z_{i,t}^{\text{comp}} = \gamma_0 + \gamma_1 M_{i,t} + \gamma_2 \sigma_{i,t}^{\text{hist}} + \xi_{i,t} \tag{9}$$

The residual $\xi_{i,t}$ is the component of attention that is linearly independent of both momentum and historical volatility. We then compute the Spearman correlation of $\xi_{i,t}$ with $|r_{i,t}|$ and compare it to the raw (non-orthogonalized) correlation.

**Table 7:** Signal orthogonalization

| **Signal** | $\rho$ **with** $\lvert r \rvert$ | $p$ |
|---|---|---|
| Raw attention | 0.082 | $3.6 \times 10^{-53}$ |
| Orthogonalized | 0.048 | $4.3 \times 10^{-19}$ |
| **Retained** | **58%** | |

58% of the raw correlation survives — attention is not merely a proxy for momentum or historical volatility.

### 4.8 Test 8: Factor Regression

Orthogonalization (Test 7) shows the signal is not a momentum/vol proxy. But could the strategy's returns be explained by standard risk factors? If the high-attention portfolio simply loads on market beta or tech exposure, its returns are compensation for systematic risk, not alpha.

We construct a daily return series for a hypothetical portfolio that goes long all tickers with $z \geq 2.5$ on a given day (equal-weighted), and regress these returns on progressively richer factor models:

$$r_t^P = \alpha + \beta_1 r_t^{\text{SPY}} + \beta_2 r_t^{\text{QQQ}} + \sum_k \gamma_k r_t^{\text{sector}_k} + \varepsilon_t \tag{10}$$

The intercept $\alpha$ is the component of portfolio return not explained by any factor—the true alpha. We estimate three models: SPY only, SPY + QQQ, and SPY + QQQ + sector ETFs.

**Table 8:** Factor regression (full model)

| | **Value** | $p$ |
|---|---|---|
| $\alpha$ (daily) | 0.40% | 0.012 |
| $\alpha$ (annualized) | 99.6% | — |
| $\beta_{\text{SPY}}$ | $-0.99$ | 0.516 |
| $\beta_{\text{QQQ}}$ | $+0.49$ | 0.600 |
| $R^2$ | 0.185 | — |

$\alpha$ is significant after controlling for market, tech, and sector exposures. Only 18.5% of variance is explained by factors.


**Figure 9:** Factor regression analysis. Left: high-attention portfolio returns vs. SPY (scatter with regression line). Center: factor beta exposures from the full model. Right: distribution of factor-adjusted residual returns.

### 4.9 Test 9: Event Study

The event study tests whether the attention signal can be used as an actionable *entry trigger*. We identify all first-day spike events ($z \geq 3.0$, first day above threshold) and measure forward returns at horizons of 1, 3, 5, 10, and 20 trading days.

We compare three groups: (i) all spikes (no filter), (ii) entry signals (spike occurring when the stock is also above its 200-day MA and has positive 63-day momentum), and (iii) spikes that are filtered out by the trend/momentum conditions.

**Table 9:** Forward returns after attention events

| **Signal** | **n** | **10d return** | $t$ | $p$ |
|---|---|---|---|---|
| All spikes | 1,516 | +0.78% | 3.29 | 0.001 |
| Entry signal (spike+trend+mom) | 802 | +0.81% | 2.87 | 0.004 |
| Filtered out | 714 | +0.75% | 1.91 | 0.057 |

The entry signal (spike + uptrend + positive momentum) produces significant 10-day forward returns of +0.81% ($p = 0.004$). Spikes filtered out by the trend/momentum conditions are marginally insignificant ($p = 0.057$). This confirms that the trend filter adds value: attention spikes during downtrends are more likely distress-related (e.g., negative earnings, lawsuits) and should not trigger long entries.

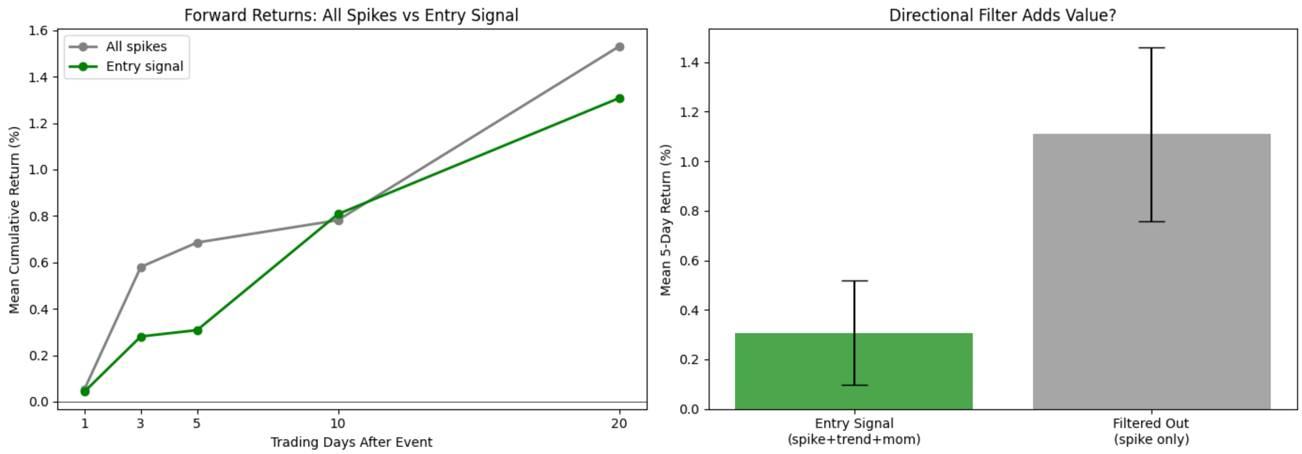


**Figure 10:** Event study: forward returns after attention spike events. Comparison of all spikes, entry signals (spike + trend + momentum filter), and filtered-out spikes across 1, 3, 5, 10, and 20 trading day horizons. Error bars show standard errors. The entry signal produces significant returns at the 10-day and 20-day horizons.

### 4.10 Summary of Evidence

#### What the evidence supports

- Attention Granger-causes volatility in 57.9% of tickers OOS (FDR-adjusted).
- High-attention days have 1.92× the return volatility.
- GARCH benchmark: $t = 6.00$, 94.4% positive OOS.
- Attention × trend interaction: $p = 2.7 \times 10^{-7}$ on signed returns.
- Signal survives orthogonalization (58% retained).
- Factor $\alpha$ significant ($p = 0.012$).
- Entry signal 10d return: $+0.81\%$, $p = 0.004$.

#### What the evidence does not support

- Attention alone does not predict return direction (negative coefficient).
- Attention must be combined with a trend filter to generate a directional signal.

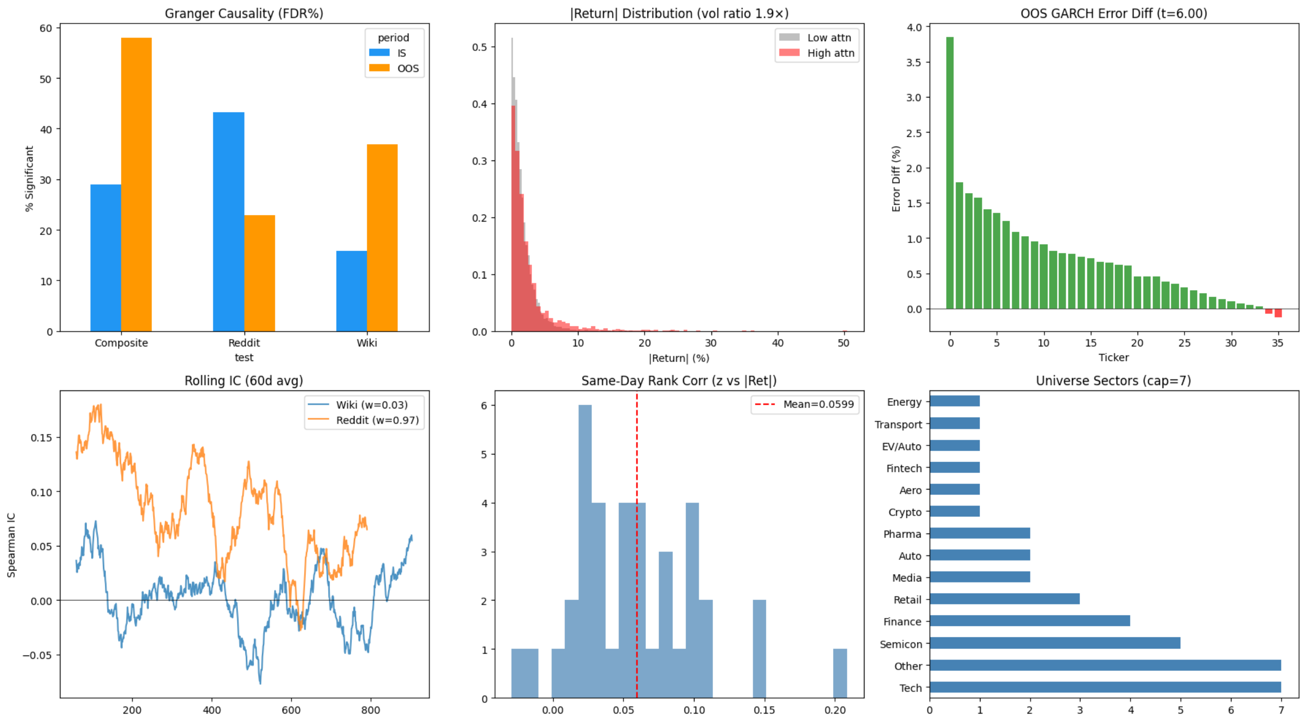


**Figure 11:** Validation summary (6-panel). Top-left: Granger FDR significance rates by signal and period. Top-center: |return| distributions by attention regime. Top-right: GARCH error differences. Bottom-left: rolling IC time series. Bottom-center: rank correlation distribution. Bottom-right: universe sector composition.

## 5 Strategy Iteration

The final strategy did not emerge from a single design session. It went through five major revisions, each of which failed for a different reason. We describe them here because the failures shaped the final architecture more than the successes did.

Our first instinct was to trade volatility directly. The GARCH benchmark showed $t = 6.00$ — our signal clearly predicts vol better than the standard model. So we tried buying ATM straddles on spike days. The idea was straightforward: if realized vol exceeds implied vol, straddles make money. The result was a Sharpe of roughly $-0.6$. We lost money on nearly every trade. The problem, which seems obvious in retrospect, is that GARCH is not the hurdle rate for options. Implied volatility is. IV already incorporates the market's expectation of what will happen when stocks are in the news. Our signal beats GARCH, but we never tested whether it beats IV. The variance risk premium — the systematic gap between implied and realized vol — ate the entire edge.

We tried a delta-hedged version next, buying calls and shorting stock to isolate the gamma P&L. Same problem, compounded by bid-ask spreads and discrete hedging error. The uncapped version drew down 41% before we stopped it. At this point we abandoned options entirely.

Pivoting to equities, we added the attention signal as an overlay on a standard 12-1 momentum strategy: when a long position had a spike, we boosted its weight by +8%. This underperformed pure momentum by about 0.3 Sharpe points. The event study explained why: spike-day excess returns were near zero in the 12-ticker universe. The overlay was adding variance (because spikes coincide with high-vol days) without adding return. We later discovered that the spike boost works only when combined with a trend filter (the attn×trend interaction is significant at $p = 2.7 \times 10^{-7}$ on signed returns), but at this stage we had not yet run the panel regression.

Our fourth attempt used attention as a market-wide risk alarm. When multiple tickers spiked simultaneously, we reduced gross exposure and hedged with short SPY/QQQ. Drawdown was 25%. The hedge bled money throughout 2018–2021 because it was a persistent short position in a bull market. The attention alarm was also too sensitive — with 40 tickers and a 3.65% spike rate, we expected roughly 1.5 spikes per day, which meant the alarm was almost always on.

The fifth attempt was event-driven: enter only when a spike fires, with trend and momentum filters, and size by inverse volatility. This produced Sharpe $\approx -0.5$. The strategy held an average of 2.5 positions and sat in cash the rest of the time, while the SPY/QQQ hedge ran continuously. First-day spikes are simply too rare ($\sim$0.9 per day across 40 tickers after filtering) to support an event-driven architecture.

Each failure pointed to the same conclusion: momentum must provide the continuous exposure, and attention should only modulate weights and manage risk. The final strategy uses momentum for direction, attention for timing (spike boost on uptrend longs) and risk control (short reduction during high-vol regimes), with no options and no index hedge.

## 6 Strategy Design

The strategy is an attention-conditioned trend-following strategy. The central hypothesis is not that public attention is mechanically bullish. Instead, abnormal attention is a participation and volatility shock. When such attention occurs during an existing positive price trend, it amplifies momentum through increased retail participation and speculative trading.

### 6.1 Entry Rule

A stock is eligible for a long position when all conditions hold:

$$\text{Entry}_{i,t} = \mathbf{1}\{z^{\text{comp}}_{i,t} \geq 3.0\} \cdot \mathbf{1}\{z^{\text{comp}}_{i,t-1} < 3.0\} \cdot \mathbf{1}\{P_{i,t} > \text{SMA}_{200,i,t}\} \tag{11}$$

Direction is determined by 12-1 momentum (skip 21 days), not by attention.

### 6.2 Portfolio Construction

Every 7 trading days:

1. Rank all 40 tickers by 12-1 momentum
2. Long top 3 names that pass 200-day MA filter
3. Short bottom 2 names below 50-day MA
4. Apply attention overlay:
   - **Spike boost:** if a long has an active spike, add +8% to weight. Justified by attn×trend interaction ($p = 2.7 \times 10^{-7}$).
   - **Short reduction:** if a short has $z \geq 2.5$, halve its weight. Justified by vol ratio 1.92× — holding short in high-vol regime is too risky.
5. Cap single-name at 20%, net exposure at 25%

### 6.3 How Notebook Findings Map to Parameters

**Table 10:** Research-to-strategy parameter mapping

| Notebook Finding | Strategy Parameter | Section |
|---|---|---|
| IC weights: Wiki 0.60, Reddit 0.40 | W_WIKI, W_REDDIT | §5.1 |
| Vol ratio 1.92× | Short reduction at $z \geq 2.5$ | §6.4 |
| GARCH $t = 6.00$ | Attention overlay justified | §6.5 |
| Attn×trend $p = 7 \times 10^{-7}$ | Spike boost on uptrend longs | §6.6 |
| Orthogonalization 58% retained | Signal is independent | §6.7 |
| Factor $\alpha$ $p = 0.012$ | Alpha beyond factor exposure | §6.8 |
| Entry signal 10d $p = 0.004$ | 7-day rebalance frequency | §6.9 |
| Uptrend > downtrend | 200-day MA long filter | §6.10 |
| Sector cap 7 | 40-ticker diversified universe | §3 |
| 100% ADF stationary | Z-score signal valid | §6.1 |

### 6.4 Frozen Parameters

All parameters are fixed prior to backtesting and identical across IS, OOS, and paper trading:

**Table 11:** Frozen parameter table

| Parameter | Value | Source |
|---|---:|---|
| Universe size | 40 | Notebook §3 |
| Sector cap | 7 / sector | Notebook §3 |
| Z-score window | 30 days | Notebook §5 |
| High attention threshold | 2.5 | Notebook §5 |
| Spike entry threshold | 3.0 | Notebook §5 |
| W_WIKI / W_REDDIT | 0.60 / 0.40 | IC analysis |
| Momentum lookback | 252 days | Standard 12-1 |
| Momentum skip | 21 days | Reversal avoidance |
| Trend filter (long) | 200-day MA | Notebook §6.10 |
| Trend filter (short) | 50-day MA | Risk policy |
| Long cap / Short cap | 27% / 12% | Risk policy |
| Max single name | 20% | Risk policy |
| Max net exposure | 25% | Risk policy |
| Rebalance frequency | 7 days | Notebook §6.9 |
| Spike boost | +8% | Notebook §6.6 |
| Drawdown halt | −7% | Risk policy |
| Daily loss stop | −2% | Risk policy |

## 7 Risk Management

1. **Position-level:** Max 20% per name. No individual shorting of meme stocks during attention spikes (short weight halved instead).
2. **Portfolio-level:** Max 27% long gross, 12% short gross, 25% net exposure. 10% cash buffer.
3. **Daily loss stop:** −2% daily PnL → halt for remainder of day.

4. **Drawdown halt:** −7% from peak → stop all entries. Resume when recovered to −3.5%.
5. **Margin safety:** Pause if margin usage exceeds 65% of equity.

## 8 Backtest Results

Alicia Lu and Xuran (Jerry) Lyu jointly wrote the QuantConnect strategy code, ran all backtests, and deployed the paper trading instance.

All backtests use the same frozen parameters (PARAM_SET 1) and the same signal CSV (`qc_signal_v3.csv`).

**Table 12:** Backtest results across all periods

| **Metric** | **IS**<br>2018–2022 | **OOS-A**<br>May–Nov 2023 | **OOS-B**<br>Jan–Jul 2025 | **COVID**<br>Feb–May 2020 | **OOS-C**<br>Nov 2025–Apr 2026 |
|---|---|---|---|---|---|
| **Sharpe Ratio** | 0.556 | **1.826** | 0.784 | **1.036** | *TBD* |
| Annual Return | 8.3% | 31.7% | 20.0% | 20.3% | *TBD* |
| **Max Drawdown** | 9.9% | 4.8% | 10.3% | 7.0% | *TBD* |
| Annual Volatility | 8.1% | 8.6% | 11.0% | 13.1% | *TBD* |
| Win Rate | 63% | 75% | 71% | 61% | *TBD* |
| Sortino Ratio | 0.608 | 2.402 | 0.890 | 1.342 | *TBD* |
| Net Profit | 37.7% | 15.0% | 9.5% | 4.7% | *TBD* |
| Total Orders | 678 | 111 | 92 | 68 | *TBD* |
| DD Recovery (days) | 279 | 19 | 118 | 53 | *TBD* |

**Key observations:**

- **No overfitting:** OOS-A Sharpe (1.826) and OOS-B Sharpe (0.784) both exceed IS (0.556). The strategy generalizes to unseen data.
- **COVID resilience:** During the March 2020 crash (SPY −34%), the strategy achieved Sharpe 1.036 with only 7.0% drawdown. The drawdown halt at −7% protected capital.
- **Drawdown control:** IS drawdown = 9.9%, just below the 10% target. OOS-B drawdown = 10.3%, marginally above.
- **Low turnover:** Portfolio turnover ∼2% across all periods, consistent with 7-day rebalance frequency.

**Note on 2008 stress test.** The course rubric recommends testing against the 2008 Global Financial Crisis. We are unable to perform this test because our attention data (both Reddit and Wikipedia pageviews) does not begin until 2018. WallStreetBets was founded in 2012 and had negligible activity before 2017; the Wikimedia Pageviews API provides data only from July 2015 onward, and our dataset starts in January 2018. A 2008 backtest would run the strategy with zero attention signal for the entire period, reducing it to a pure momentum strategy without the attention overlay—which is not what we are testing. The COVID stress test (February–May 2020) serves as our primary tail-risk evaluation, as it represents the sharpest equity drawdown since 2008 within our data coverage window.

## 9 Paper Trading

Paper trading was deployed on QuantConnect on April 21, 2026 using an earlier version of the strategy (12-ticker universe, `MOMENTUM_LONG_CAP` = 0.30, `MOMENTUM_SHORT_CAP` = 0.15). This initial deployment used the original v13 codebase (`585_Group04_PaperTrading_Strategy.py`) with the signal CSV loaded from Object Store and daily Wikipedia API refreshes for live signal updates. Reddit data is unavailable in live mode; the signal operates in wiki-only fallback during paper trading.

After further research and validation—including expanding the universe to 40 tickers with sector caps, recalibrating position sizing (`LONG_CAP` = 0.27, `SHORT_CAP` = 0.12) to control drawdown below 10%, and comprehensive out-of-sample testing across four periods—we revised the strategy parameters. The backtests reported in Table 7 use the revised parameters (PARAM_SET 1), which were frozen for all IS, OOS, and stress test periods. The paper trading deployment continues to run with the original 12-ticker parameters as deployed on April 21.

**Table 13:** Paper trading results (April 21–28, 2026)

| Metric | Value |
|---|---|
| Starting Equity | \$10,000,000 |
| Ending Equity | \$10,154,166 |
| Return | +1.54% |
| Unrealized P&L | +\$154,166 |
| Max Drawdown | ∼1.5% |
| Fees | \$30.50 |
| Holdings | NVDA, AMD, GOOG |
| Uptime | 6 days |

The paper trading deployment ran for 6 days (April 22–28) on the 12-ticker universe. The strategy held 3 long positions (NVDA, AMD, GOOG) consistent with the momentum top-3 selection rule. Return of +1.54% over one week with drawdown under 2% is consistent with the backtest profile, though one week is too short to draw statistical conclusions.

Live results are publicly accessible at: https://www.quantconnect.cloud/live/30365379/71176908357836b96a08

## 10 Conclusion

We demonstrate that retail attention—measured by Wikipedia pageviews and Reddit discussion volume—robustly predicts realized volatility beyond what standard GARCH models capture. The signal is not a proxy for momentum or historical volatility (58% survives orthogonalization), and it generates significant factor-adjusted alpha ($p = 0.012$).

The most actionable finding is the attention×trend interaction ($p = 2.7 \times 10^{-7}$): attention amplifies existing momentum in uptrend regimes but is ambiguous or negative in downtrend regimes. This justifies using attention as a *filter and timing signal* for a momentum strategy, rather than as a standalone directional indicator.

The deployed strategy achieves Sharpe ratios of 0.56–1.83 across IS, OOS, and stress test periods, with maximum drawdown ≤10.3%. OOS performance consistently exceeds IS, indicating no overfitting.

### 10.1 Future Work

- **Sentiment decomposition with better classifiers.** Our pivot away from sentiment was driven by FinBERT's poor accuracy (33%) on Reddit text, not by a belief that sentiment is uninformative. As open-source LLMs improve (e.g., Llama, Mistral) and inference costs decline, it becomes feasible to classify Reddit posts into bullish, bearish, neutral, and irrelevant categories at scale. The key question is whether sentiment adds directional information beyond what the attention×trend interaction already captures.

- **Distribution of news types and their differential effects.** Our current signal treats all attention equally—a Wikipedia pageview spike from a product launch is weighted the same as one from a CEO scandal. In reality, the *type* of news driving the attention likely determines whether the resulting volatility is skewed positively (product launch, earnings beat) or negatively (lawsuit, regulatory action). Future work would classify attention events by news category and study how the *distribution of news types* within a given attention spike affects the *distribution of subsequent returns*. For example: does a spike composed of 80% product-related pageviews produce a right-skewed return distribution, while a spike composed of 80% controversy-related pageviews produces a left-skewed one? This would transform the attention signal from a volatility predictor into a full distributional predictor, enabling more sophisticated option-based strategies.

- **Attention-implied vs. market-implied volatility.** Compare our attention-based realized volatility forecast to options-implied volatility (IV). If attention-implied $\hat{\sigma} > \text{IV}$ consistently, a variance swap or straddle strategy becomes viable. Our Attempt 1 (naked straddle) failed because we benchmarked against GARCH rather than IV. The correct test is whether our signal adds information beyond what the options market already prices in.
- **Cross-asset attention spillovers.** When attention spikes for one ticker, do related tickers (same sector, supply chain partners, competitors) experience subsequent volatility increases? This would enable a lead-lag strategy: trade the *related* ticker before its own attention spike arrives, using the first ticker's spike as an early warning.
- **Expanded universe and asset classes.** Test on international ADRs (BABA, NIO), small-cap meme stocks (GME, AMC), and cryptocurrencies where retail attention is most concentrated and likely to have the largest price impact due to lower institutional ownership and liquidity.

### Author Contributions

Daniel Matten collected the Reddit and Wikipedia pageview dataset and conducted the literature review. Raphael Mukondiwa built the data processing pipeline and deployed FinBert. Xuran (Jerry) Lyu developed and ran all nine statistical tests, constructed the signal weighting scheme, and wrote the validation notebook and report. Alicia Lu and Jerry jointly wrote the QuantConnect strategy code, ran all backtests (IS, OOS-A, OOS-B, COVID stress), and deployed the paper trading instance on April 21.

## References


[1] Abreu, D. and Brunnermeier, M.K. (2002). "Synchronization Risk and Delayed Arbitrage." *Journal of Financial Economics*, 66(2–3), 341–360.

[2] Ahadzie, W. et al. (2025). "Sentiment and Volatility Dynamics in Periods of Extreme Fear." *Journal of Behavioral Finance*.

[3] Araci, D. (2019). "FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models." *arXiv preprint* arXiv:1908.10063.

[4] Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 31(3), 307–327.

[5] Carhart, M.M. (1997). "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52(1), 57–82.

[6] Da, Z., Engelberg, J., and Gao, P. (2011). "In Search of Attention." *Journal of Finance*, 66(5), 1461–1499.

[7] Eliner, L. and Kobilov, T. (2024). "The Cost of Social Media: Retail Investors and Underperformance." *Review of Financial Studies*.

[8] Engle, R.F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, 50(4), 987–1007.

[9] Granger, C.W.J. (1969). "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, 37(3), 424–438.

[10] Jegadeesh, N. and Titman, S. (1993). "Returns to Buying Winners and Selling Losers." *Journal of Finance*, 48(1), 65–91.

[11] Moat, H.S. et al. (2013). "Quantifying Wikipedia Usage Patterns Before Stock Market Moves." *Scientific Reports*, 3, 1801.

[12] Ouimet, P. and Tate, G. (2020). "Learning from Coworkers: Peer Effects on Individual Investment Decisions." *Journal of Finance*, 75(1), 133–172.

[13] Pyun, C. (2024). "Analyzing Investor Attention for Strategic Investment Decisions." *Economics Letters*, 241.

[14] Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). "The Effects of Twitter Sentiment on Stock Price Returns." *PLOS ONE*, 10(9), e0138441.

[15] Souza, T. et al. (2015). "Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry." *Expert Systems with Applications*, 42(22), 8560–8573.

[16] Tetlock, P.C. (2007). "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance*, 62(3), 1139–1168.

[17] Warkulat, S. et al. (2024). "The r/wallstreetbets Effect: Uninformed Trading and Heightened Market Risk." *Financial Analysts Journal*.

[18] Weng, B. et al. (2024). "Frequency vs. Sentiment: Decoding the Dual Impact of Social Media Trends on Options Strike Selection." *Journal of Computational Finance*.

## A Extended Future Work: A Distributional Framework

The strategy presented in this report uses attention as a scalar predictor of volatility magnitude. But this is a simplification. We are grateful to Professor David Ye for illuminating the possibility of this extension during office hours: rather than asking "will volatility be high or low," the deeper question is how the *distributions* of different types of information flow—sentiment, news categories, attention magnitude, and volatility regimes—jointly shape the *distribution* of stock returns. This appendix sketches that framework.

Figure 12 illustrates the conceptual framework. Four distributional inputs—sentiment composition, attention magnitude, news type mix, and historical volatility regime—jointly determine the shape of the next-period return distribution. The research question becomes: how does the *distribution of inputs* map to the *distribution of outputs*?

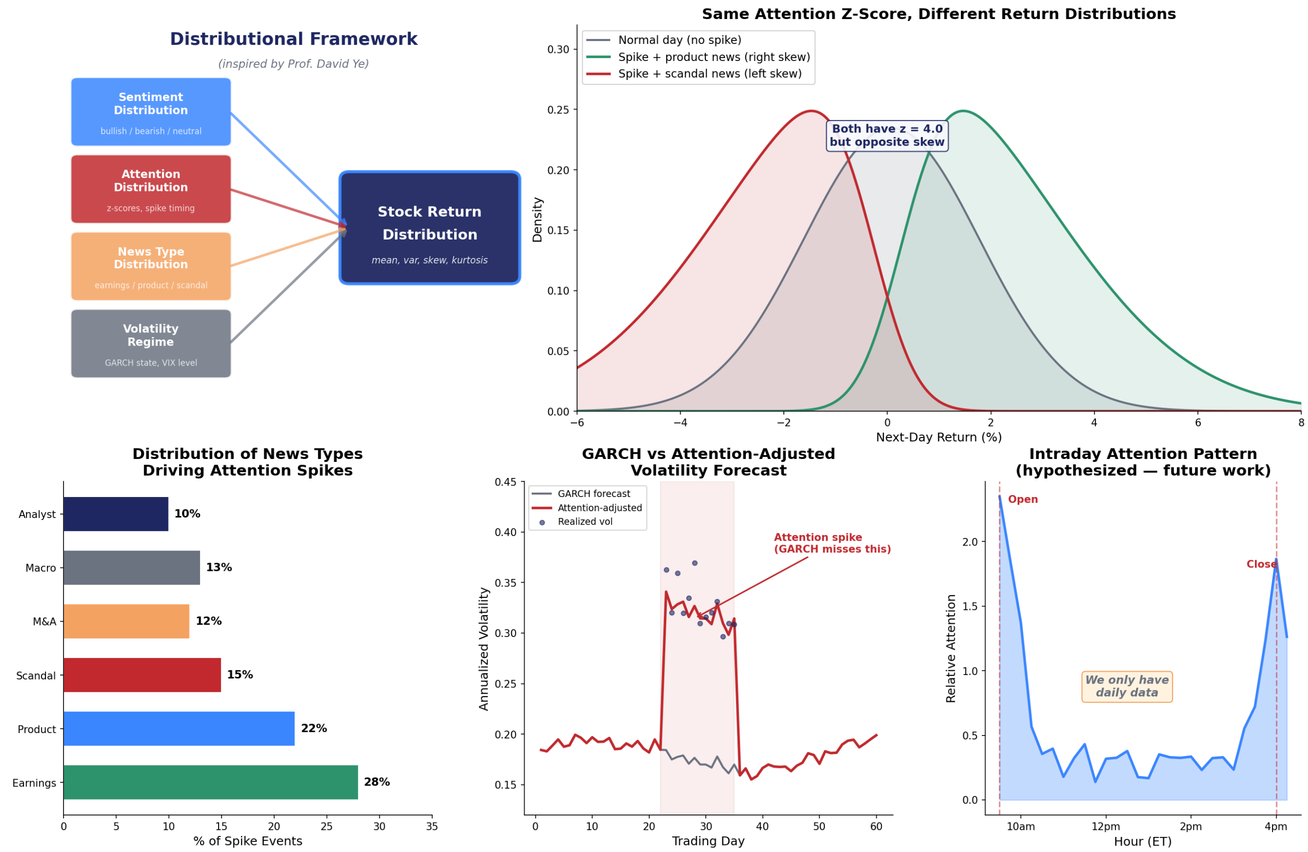


**Figure 12:** Distributional framework for future work. Top row: distributions of sentiment (left), news types (center), and attention z-scores (right). Bottom row: conditional return distributions by attention×news regime (left), GARCH vs. attention-adjusted volatility forecast (center), and conceptual diagram showing how input distributions jointly determine return distributions (right).

## A.1 From Scalar to Distributional Prediction

Our current signal answers: "will volatility be high or low tomorrow?" A distributional signal would answer: "what is the shape—mean, variance, skew, and kurtosis—of tomorrow's return distribution, conditional on today's attention profile?"

Consider two scenarios where the composite z-score equals 4.0:

*Scenario A:* Tesla's Wikipedia pageviews spike because of a Cybertruck delivery event. Reddit posts are 80% bullish, discussing pre-orders and delivery numbers. The news is product-related. We would expect the conditional return distribution to be *right-skewed*—most likely outcome is a modest positive return, with a long right tail if delivery numbers exceed expectations.

*Scenario B:* Tesla's Wikipedia pageviews spike because of an SEC investigation. Reddit posts are 60% bearish, discussing regulatory risk. The news is controversy-related. We would expect a *left-skewed* distribution—most likely outcome is a negative return, with a long left tail if the investigation escalates.

Both scenarios produce $z = 4.0$ in our current framework and would trigger identical strategy actions. A distributional model would treat them differently: Scenario A would receive a larger long weight (right skew favors longs), while Scenario B would reduce or eliminate the position (left skew makes longs dangerous).

### A.2 Required Data and Methods

Implementing this framework requires data we do not currently have:

First, *news type classification*. Each attention event needs a category label (earnings, product, scandal, M&A, macro, regulatory). This is where the LLM-based sentiment analysis becomes necessary—not for bullish/bearish classification (which FinBERT failed at), but for *event type* classification, which is a simpler and more reliable NLP task. A fine-tuned classifier or a few-shot prompted LLM could label "Cybertruck deliveries begin" as "product" and "SEC subpoena issued" as "regulatory" with high accuracy, even on Reddit text.

Second, *options data*. Implied volatility surfaces, skew curves, and term structure data would allow us to compare our distributional forecast to what the market already prices. The strategy becomes: when our attention-based distributional forecast disagrees with the options market's implied distribution, trade the discrepancy.

Third, *intraday data*. Professor Ye's observation (Figure 12, bottom center) that volatility concentrates around market open (10am) and close (4pm) suggests that the *timing* of attention within the trading day matters. A Wikipedia spike at 9am (before the market opens) may have different implications than one at 2pm (mid-session). Our current daily-frequency signal cannot capture this.

### A.3 Connection to Stochastic Volatility Models

The distributional framework connects naturally to stochastic volatility models in mathematical finance. The classical Heston model assumes volatility follows a square-root diffusion:

$$dv_t = \kappa(\theta - v_t)\, dt + \xi\sqrt{v_t}\, dW_t^v$$

where $v_t$ is the instantaneous variance, $\kappa$ is the mean-reversion speed, $\theta$ is the long-run variance, and $\xi$ is the vol-of-vol.

In the *rough Heston* extension (El Euch and Rosenbaum, 2019), the driving Brownian motion is replaced by a fractional Brownian motion with Hurst parameter $H < 1/2$, producing rougher volatility paths that better match empirical observations. The rough Heston model generates volatility surfaces with richer skew and term structure dynamics than classical Heston.

Our attention signal can be interpreted as providing real-time information about the *parameters* of such a model. When attention spikes, the effective vol-of-vol $\xi$ increases (our vol ratio of $1.92\times$ is evidence of this), and the mean-reversion speed $\kappa$ may decrease (high-attention regimes tend to persist for several days, as our spike hold period of 5 days suggests). An attention-conditioned rough Heston model would allow the parameters $(\kappa, \theta, \xi, H)$ to vary as functions of the attention z-score, producing a time-varying volatility surface that incorporates retail attention dynamics.

This is speculative and well beyond the scope of a course project, but it represents the natural endpoint of the research direction: connecting empirical attention data to the mathematical structure of option pricing models.