

BUZZ SCORE • ALTERNATIVE DATA RESEARCH

Attention Leads Volatility

The Buzz Score, and a 12-month empirical test across 198 S&P 500 stocks from all 11 GICS sectors. The paper documents how the score is built, then asks one testable question of a rule-selected index sample: when retail attention measured on Reddit rises, does anything happen to the price?

198	~249	49,313	12
S&P 500 stocks, 11 GICS sectors, no ETFs	Average trading sessions per name	Ticker-day observations analysed	Months of daily Buzz history

EXECUTIVE SUMMARY

What a year of attention data reveals

The Buzz Score is a normalized, 0–100 measure of how much a stock is being discussed and how that discussion is moving (construction overleaf). The study asks one testable question across 198 S&P 500 stocks selected by a fixed coverage rule: when attention rises, does anything happen to the price?

+0.24	+0.16	1.9_x	+87%
Correlation of Buzz with trading volume, pooled across names	Correlation of Buzz with same-day volatility (absolute return)	Trading volume on extreme mention-spike days, vs. full-window median	Lift in volatility on a mention-spike session, vs. the pre-spike baseline

- 01 Attention tracks activity.** Across the basket, the Buzz Score is positively correlated with trading volume (mean $r = +0.24$) and realized volatility ($r = +0.16$). The relationships hold for **92% and 88% of names** respectively (cross-sectional $t > 17$).
- 02 Attention leads risk, not direction.** Today's Buzz level forecasts **tomorrow's** volatility ($r = +0.09$, $t = 12.6$), modest in size but surviving the own-history and error-treatment controls in Finding 6. It does not forecast the *sign* of the next move: the Buzz Score is an early-warning signal for magnitude, not a directional alpha source.
- 03 Mention spikes mark short volatility windows.** Around the 1,011 mention spikes entering the event study, volume runs at 1.9_x its median and volatility rises 87%, with the effect persisting into the following session.
- 04 The pattern holds across sectors.** The mean buzz–volatility correlation is positive in 10 of 11 GICS sectors, led by Information Technology (+0.27); only Real Estate (−0.06, $n = 4$) dips below zero.
- 05 Spike days coincided with the year's catalysts.** Fiserv −44%, Trade Desk −39% and Oracle +36% fell on datable events; Paramount Skydance (+37%) moved on attention alone.

THE METRIC

How the Buzz Score is built

The score reads every post and comment across 50+ financial subreddits, filters bots and spam, and combines five components into one 0–100 value per stock and day. Sentiment is an ensemble: finance-tuned RoBERTa (60%) and VADER (40%, plus 150+ finance terms), refined by phrase and emoji signals.

COMPONENT	WEIGHT	CONSTRUCTION
Mention volume	20	$\log_{10}(\text{mentions} + 1) \times 20$; 1,000 mentions → 60 pts
Sentiment	± 20	ensemble score (–1 bearish ... +1 bullish) $\times 20$
Quality	10	$\log_{10}(\text{upvotes per mention} + 1) \times 10$, volume-scaled
Diversity	14	$\log_{10}(\text{effective subreddits} + 1) \times 14$ (HHI-based)
Trend	–10...+20	recent vs. older mentions, asymmetrically capped

Table M. Five additive components (weights are multipliers, not caps), inputs normalized to a 7-day window; the raw sum is scaled asymptotically (linear below 50; raw 100 → 86, raw 150 → 96). Bands: 0–20 minimal, 40–60 active, 80+ exceptional.

Echo-chamber detection

Raw subreddit counts mislead: 95% of mentions in one community is an echo chamber, not breadth. Diversity uses Herfindahl-Hirschman *effective subreddits* ($1/\text{HHI}$): GME at its peak, 18 raw but 1.1 effective; NVDA, 15 raw, 12.5 effective and the higher score.

Thread intelligence

Replies like “this is the way” under a \$NVDA post carry opinion without the ticker. Direct replies inherit the ticker context at 0.1 weight, one level deep; agreement phrases inherit the parent’s sentiment, other replies keep their own.

RELATION TO THIS STUDY’S VARIABLES

Mention counts are the weighted mentions above (explicit + $0.1 \times$ inherited); scores are stored as computed at ingest, so the history has no look-ahead. An author guard caps single-account contributions; non-English subreddits score neutral. Sentiment is both a score component (± 20 points) and the separately exposed field analysed later, so the Buzz and sentiment rows of Table 1 are not independent. Status: deployed implementation V5.5, June 2026.

— DATA & METHOD

How the study was built

Two independent sources, aligned on trading days, analysed with transparent statistics. Index membership plus a coverage rule defines the sample; no name is hand-picked.

The data

- **Attention.** Daily Buzz Score, mentions and directional sentiment per name from the Adanos Reddit feed.
- **Price.** Split- and dividend-adjusted closes from Alpha Vantage; volume split-adjusted to a constant share basis.
- **Basket.** Every S&P 500 common stock with Reddit mentions on at least half of all window days: 198 names, all 11 GICS sectors, no ETFs, no hand-picking.
- **Window.** 2025-06-12 to 2026-06-10, ~249 sessions per name, 49,313 ticker-days; weekend discussion is not carried into Monday.

The method

- **Returns.** Daily log returns on adjusted closes; volatility proxied by absolute return (ranges in Finding 6).
- **Correlation.** Pearson r per name on pairwise-complete days (the feed omits no-mention days), pooled cross-sectionally; Finding 6 adds date-clustered errors.
- **Lead-lag.** A cross-correlation function relates ΔBuzz to returns at leads and lags of ± 5 sessions.
- **Event study.** Spikes: days with log mentions over $+2\sigma$ above the name's full-window mean (in-sample, descriptive).

TIMING INTEGRITY

Buzz Scores are persisted at ingest (an API design property we rely on rather than audit), so lead-lag results use only attention available before the sessions they predict. The event-study *threshold* is full-window: descriptive, not tradable.

PRIOR WORK

The study extends the attention-and-markets literature: attention-driven retail buying (Barber & Odean 2008; Da, Engelberg & Gao 2011), online talk and news (Antweiler & Frank 2004; Tetlock 2007) and Reddit specifically (Long, Lucey, Xie & Yarovaya 2023).

FINDING 1 • CO-MOVEMENT

Attention, volume and volatility move together

The most robust relationship in the data is the simplest. When a stock is being discussed more, it is also being traded more and moving more. The table reports the mean per-name correlation, its cross-sectional significance, and the share of the 198 names for which it is positive.

RELATIONSHIP	MEAN R	T-STAT	NAMES +	READING
Buzz ↔ log volume	+0.24	18.4	92%	Attention tracks turnover
Buzz ↔ return	+0.16	17.6	88%	Attention tracks volatility
Sentiment ↔ return	+0.08	12.4	80%	Mood aligns with direction, same day (partly talk reacting to the move)
ΔBuzz ↔ return	+0.03	4.9	65%	Rising buzz, rising price (weak)

Table 1. Pooled contemporaneous correlations, 49,313 ticker-days; |return| is the absolute daily return, the size of the move irrespective of sign. All four clear false-discovery-rate (FDR) control at $q = 0.05$ and at the stricter $q = 0.01$ (Methods, Finding 6). Same-day buzz includes after-close discussion, so part of the contemporaneous coupling is attention reacting to the day's move; the predictive results in Findings 2 and 6 are free of that overlap.

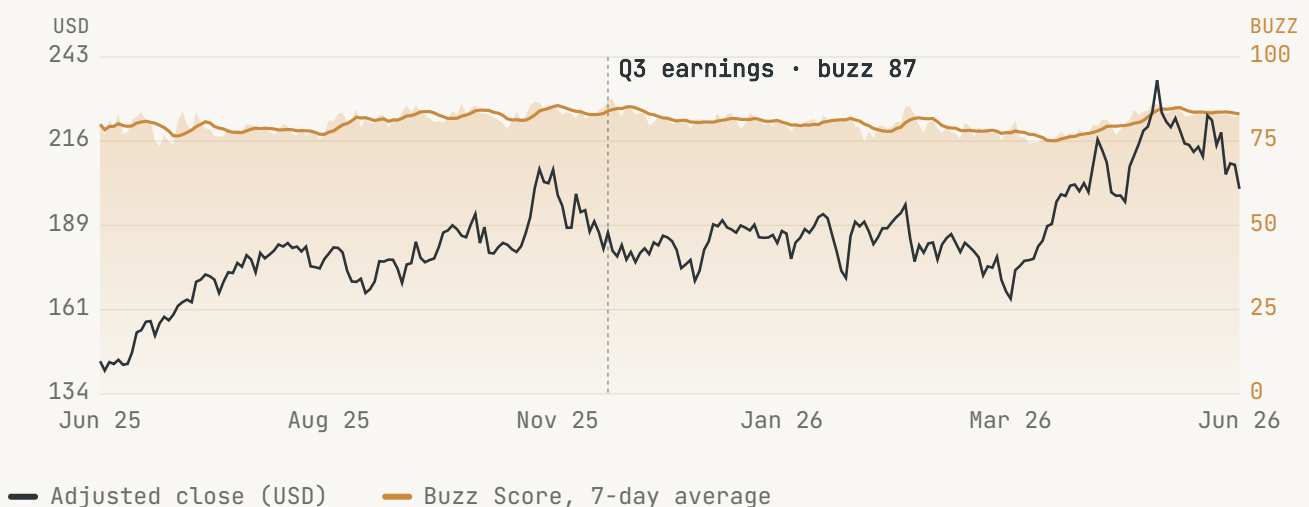


Figure 1 • NVIDIA, 2025-06-12 to 2026-06-10. Buzz and price trace the same regime: attention swells into the November earnings event (spike-day Buzz 87.2) and ebbs through quieter stretches.

FINDING 2 • LEAD-LAG

Buzz leads the size of the move, not its sign

Same-day correlation could just mean both react to the same news. The sharper test: does a change in attention *precede* a change in the tape? The cross-correlation below relates a one-day change in Buzz to absolute returns at leads and lags of up to a week.

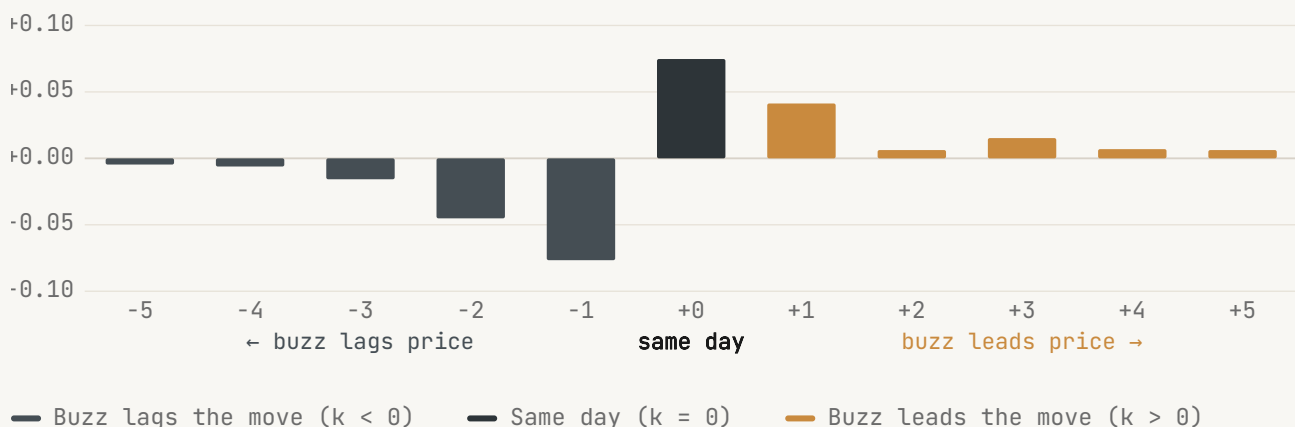


Figure 2. Pooled cross-correlation of ΔBuzz with $|\text{return}|$. The dominant features are reactive: buzz jumps with the move ($k = 0$, $r = +0.07$) and falls back after it ($k = -1$, $t = -11.9$). A smaller lead survives at $k = +1$ ($t = 6.9$); no significant lead remains beyond one session ($k = +2$: $t = 1.2$); an isolated $k = +3$ bar clears FDR, read as noise.

Magnitude is predictable

Today's Buzz *level* forecasts tomorrow's volatility with a pooled r of **+0.09** ($t = 12.6$), positive for **79%** of names: modest in size, statistically robust (Finding 6).

Direction is not

A change in Buzz does not predict the *sign* of the next day's return ($r = -0.00$, $t = -0.4$). Isolated direction bars ($k = +5$) do clear FDR; with no support at adjacent lags we read them as noise. Buzz says how big the next move is, not which way.

INTERPRETATION

The Buzz Score acts as an attention-driven volatility signal: an early estimate of *where the action will be*, plus a directional sentiment field. The lead survives Finding 6's controls; what it cannot separate is scheduled versus spontaneous attention (Limitations).

FINDING 3 • EVENT STUDY

The anatomy of a mention spike

Spikes: log mentions over $+2\sigma$ above the name's full-window mean (1,045 de-clustered; 1,011 enter the averages); day 0 is the spike.

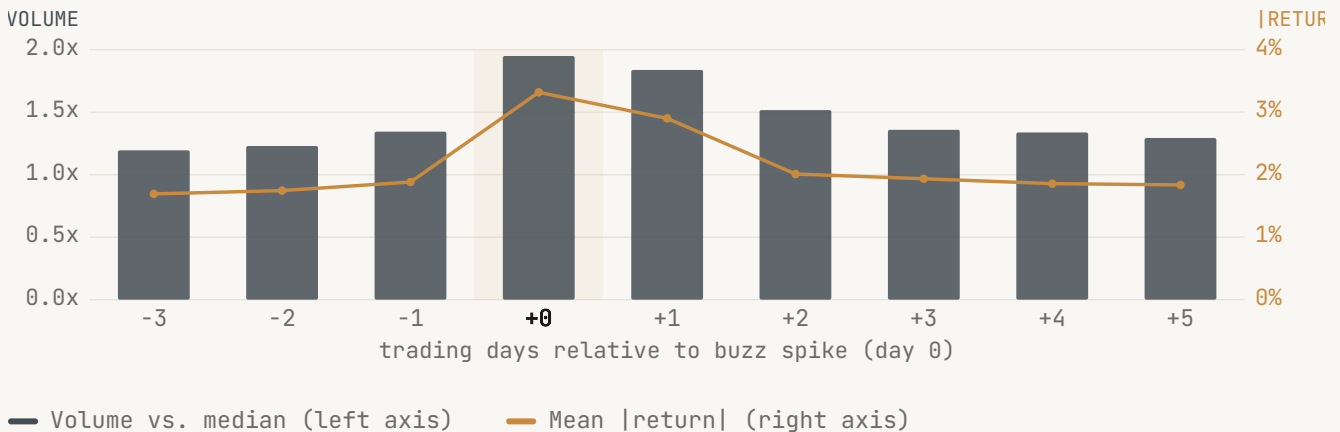


Figure 3. Average response around 1,011 spikes: volume and volatility build ahead of the spike (partly attention reacting to moves under way), peak on day 0, stay high one day. The 1,011 events fall on 229 distinct sessions; not independent, descriptive only.

SESSION	MEAN RETURN	VOLUME VS MEDIAN	MEAN RETURN
-3	1.69%	1.19×	+0.22%
-2	1.74%	1.23×	+0.27%
-1	1.88%	1.34×	+0.29%
+0	3.32%	1.95×	+0.25%
+1	2.90%	1.84×	+0.05%
+2	2.01%	1.52×	-0.00%
+3	1.94%	1.36×	+0.17%
+4	1.86%	1.34×	+0.08%
+5	1.84%	1.29×	-0.22%

Table 2. Spike-session volatility: 3.3% vs a 1.8% baseline (mean of -3 to -1; +87%); volume 1.9× the full-window median (1.7× vs a trailing mean, Finding 6). The mean return (+0.25%) is small: spikes are near-symmetric in direction, asymmetric in size.

FINDING 4 • CASE STUDIES

When the signal met the catalyst

The aggregate statistics describe the average day; the individual events illustrate the pattern at its sharpest. Table 3 lists the four largest reaction-session moves among each name's top mention spike of the year: three fell on datable catalysts, the fourth was pure attention.

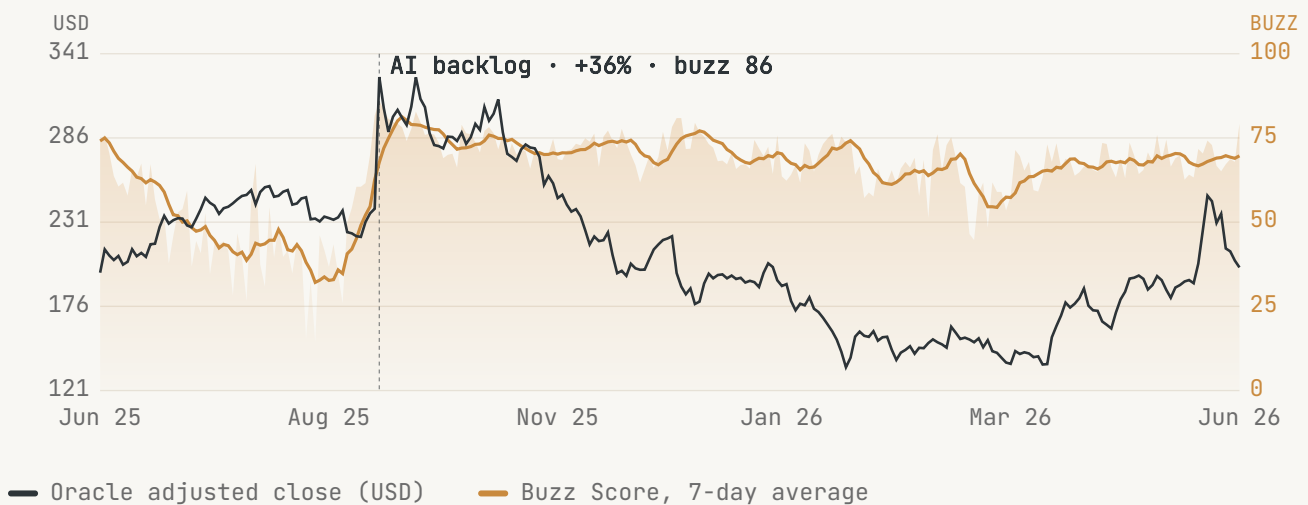


Figure 4 • Oracle. Oracle's fiscal Q1 landed after the prior session's close with a 359% surge in contracted backlog (RPO); the stock closed +36% on 2025-09-10 while the Buzz Score registered 86.1, Oracle's largest mention day of the year ($z = 3.2$, 2,653 mentions).

NAME	DATE	BUZZ	SESSION RETURN	CATALYST
FISV	2025-10-29	77.5	-44.0%	Q3 miss, guidance slashed
TTD	2025-08-08	74.4	-38.6%	Weak Q3 guidance, CFO exit
PSKY	2025-08-13	68.1	+36.7%	Low-float retail squeeze, no news
ORCL	2025-09-10	86.1	+35.9%	AI-cloud backlog, RPO +359%

Table 3. Close-to-close simple returns on the reaction session: Fiserv reported pre-market the same day, Trade Desk and Oracle after the prior close. Paramount Skydance moved on no same-day announcement, a low-float retail squeeze where attention was the event itself. The converse does not hold: the largest spike by z (DUK, $z = 6.6$) moved +0.4%, and 10 of the 20 largest measurable spikes moved under 2%.

FINDING 5 • CROSS-SECTION

Where the signal is strongest, and where it inverts

The attention–volatility link has clear sector structure: means range from -0.06 (Real Estate, $n = 4$) to +0.27 (Information Technology) across the 198 names.

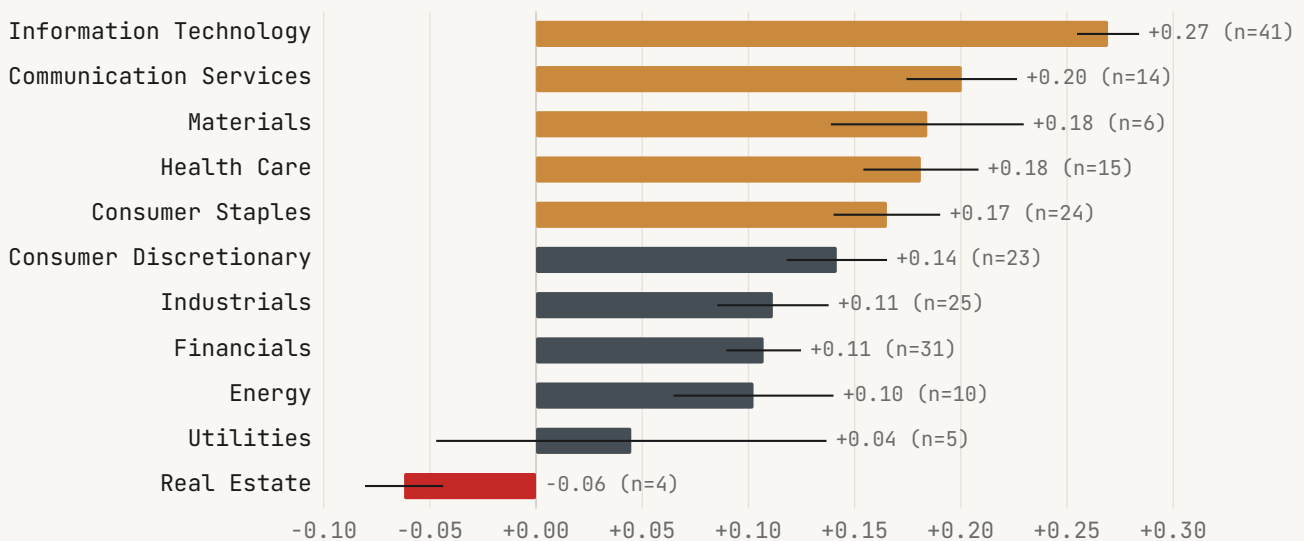


Figure 5. Mean per-name correlation of Buzz with $|return|$ by GICS sector, whiskers ± 1 cross-sectional SE: a descriptive sort, no selection-adjusted inference. Gold bars: sector mean $\geq +0.15$.

Strongest where events live

Information Technology and Communication Services lead the sector table; the strongest single names are **KVUE** (+0.49), **QCOM** (+0.48) and **PSKY** (+0.41): names whose discussion clusters around discrete, datable events.

Negatives are scattered and shallow

88% of names are positive; the 24 negatives are mostly shallow (median -0.04; deepest **DUK** at -0.21). Whether the few negatives reflect anything systematic is untested. Sector dispersion is real but secondary to the common pattern.

FINDING 6 • ROBUSTNESS

Stress-testing the lead

Four objections: a crude volatility proxy, raw volume, clustering as the real driver, names treated as independent. The data answers each.

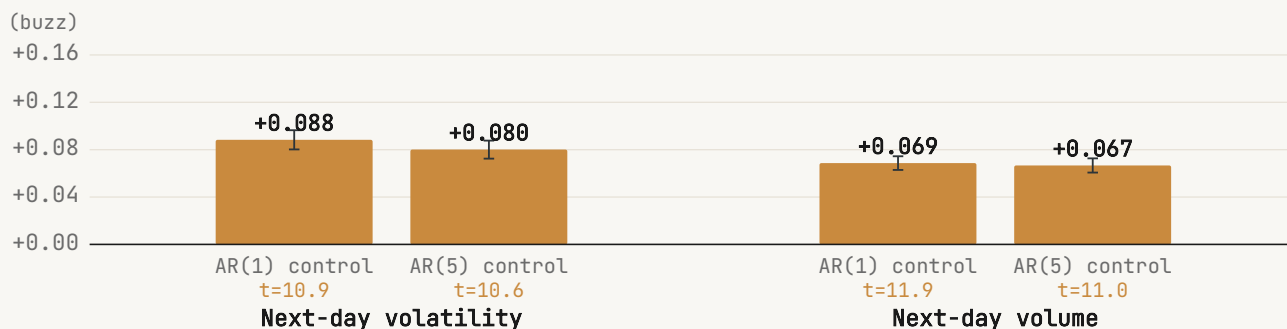


Figure 6. Standardized Buzz($t-1$) coefficient predicting next-day volatility and volume, **controlling for the target's own history** (one and five lags); whiskers ± 1 cross-sectional SE. Positive and significant under every specification and error treatment.

Sharper measures, same result

Normalized true range (Wilder 1978, adjusted OHLC) lifts the Buzz–volatility correlation to **+0.19** (t 17.8). Against **abnormal** volume (vs. a trailing 20-day mean) Buzz correlates **+0.21** (t 21.6, positive for 92% of names). Around spikes, range volatility rises 59% above its pre-event level, to 5.3%, and abnormal volume reaches 1.7 $\times$.

Neither clustering nor co-movement

Volatility is autocorrelated (Engle 1982; Bollerslev 1986), so the lead is tested *net of* the target's own past. With one and five own-lags controlled, Buzz($t-1$) keeps a positive coefficient (mean-group +0.08 to +0.09; pooled +0.08 to +0.09), significant across names (t 10.6–10.9), under within-name Newey–West errors (one-lag, t 13.9) and under calendar-date clusters (t 8.0–8.9; volume $\approx$ 6.1).

METHODS & INFERENCE

Range volatility: normalized true range (Wilder 1978) on adjusted OHLC; own-lag controls are motivated by ARCH/GARCH (Engle 1982; Bollerslev 1986). Per-name coefficients: mean-group t -test (Pesaran & Smith 1995); panel errors: Newey–West (1987) within name (Bartlett, 10 lags) and calendar-date clusters. Every effect reported as significant clears Benjamini–Hochberg (1995) FDR control at $q = 0.05$ across all 40 tests, the Table 1 effects also at $q = 0.01$ (exact Student- t p -values); per-name correlations and sector means are descriptive.

— WHAT IT MEANS

Reading the signal

What Buzz is good for

- A leading estimate of **volatility and trading activity** for sizing and risk overlays.
- **Event flagging** at scale: the sample's largest spike-day moves coincided with real catalysts.
- A **cross-sectional attention factor**: positive sector means in 10 of 11 GICS sectors across 198 names.

What it is not

- A standalone **directional** signal. Buzz forecasts magnitude, not sign.
- A claim of **causation**: these are observational relationships over one year.
- Uniform in strength: sector means span -0.06 to +0.27 (Figure 5).

LIMITATIONS & NEXT STEPS

The basket is coverage-conditioned (it is also the feed's deployable universe); within it the volatility effect is nearly flat in coverage ($r = +0.07$), the volume effect tilts toward covered names ($+0.15$). Constituents: a June 2026 snapshot; intra-window leavers are absent, GOOG/GOOGL both enter. Cross-sectional t-statistics treat names as independent (Finding 6); sentiment–return correlation is partly mechanical (Table 1 note); the window is one news-rich year. Value over lagged volume or an event calendar is untested; those controls and an out-of-sample test come next.

Work with the data

The Buzz Score, directional sentiment and raw mentions are available per ticker and platform via the Adanos API, with historical depth for backtesting.

WEB

adanos.org

CONTACT

contact@adanos.org

COMPANY

Adanos Software GmbH • Berlin

REFERENCES

Sources

ATTENTION AND MARKETS

Barber, B. M., & Odean, T. (2008). *All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors*. *Review of Financial Studies* 21(2), 785–818. doi.org/10.1093/rfs/hhm079

Da, Z., Engelberg, J., & Gao, P. (2011). *In Search of Attention*. *Journal of Finance* 66(5), 1461–1499. doi.org/10.1111/j.1540-6261.2011.01679.x

SOCIAL MEDIA, NEWS AND MARKETS

Antweiler, W., & Frank, M. Z. (2004). *Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards*. *Journal of Finance* 59(3), 1259–1294. doi.org/10.1111/j.1540-6261.2004.00662.x

Tetlock, P. C. (2007). *Giving Content to Investor Sentiment: The Role of Media in the Stock Market*. *Journal of Finance* 62(3), 1139–1168. doi.org/10.1111/j.1540-6261.2007.01232.x

Long, S., Lucey, B., Xie, Y., & Yarovaya, L. (2023). *“I just like the stock”: The role of Reddit sentiment in the GameStop share rally*. *Financial Review* 58(1), 19–37. doi.org/10.1111/fire.12328

VOLATILITY MEASUREMENT AND ECONOMETRIC METHOD

Wilder, J. W. Jr. (1978). *New Concepts in Technical Trading Systems*. Trend Research, Greensboro, NC. openlibrary.org/books/OL4745184M

Engle, R. F. (1982). *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. *Econometrica* 50(4), 987–1007. doi.org/10.2307/1912773

Bollerslev, T. (1986). *Generalized Autoregressive Conditional Heteroskedasticity*. *Journal of Econometrics* 31(3), 307–327. [doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)

Pesaran, M. H., & Smith, R. (1995). *Estimating Long-Run Relationships from Dynamic Heterogeneous Panels*. *Journal of Econometrics* 68(1), 79–113. [doi.org/10.1016/0304-4076\(94\)01644-F](https://doi.org/10.1016/0304-4076(94)01644-F)

Newey, W. K., & West, K. D. (1987). *A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*. *Econometrica* 55(3), 703–708. doi.org/10.2307/1913610

Benjamini, Y., & Hochberg, Y. (1995). *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. *J. Royal Statistical Society B* 57(1), 289–300. doi.org/10.1111/j.2517-6161.1995.tb02031.x

DATA & REPRODUCIBILITY

Attention: Adanos Market Sentiment API. Price and volume: Alpha Vantage daily adjusted series (volume split-adjusted, ranges on adjusted OHLC). Universe: S&P 500 constituents, June 2026 snapshot. The full pipeline is available on request; the metric page documents the deployed scoring implementation (V5.5).