

AI model news and relative equity repricing: Evidence from infrastructure and software stocks

AUTHORS: not stated

SUMMARY: The study investigates whether large returns around AI model news reflect common infrastructure-software repricing or individual constituent-specific movements. Using a frozen external database, the researchers identified 28 language-model records and found that seven two-session spreads exceeded a descriptive volatility threshold. Only two retained the threshold after every single-stock deletion, involving Claude models with opposite performance directions. Two large negative GPT spreads were strongly influenced by software firms reporting earnings.

METHOD: The research employed event-study methods to benchmark abnormal performance against historical return variation. The sample included infrastructure and software stocks, with Yahoo Finance adjusted prices covering 1,399 trading sessions from 2021 to 2026. Compustat provided firm-report records, and the main cohort used Epoch AI's Frontier Models CSV, downloaded and frozen on September 6, 2026.

KEY FINDINGS:

- Seven two-session spreads exceeded a descriptive volatility threshold.
- Only two retained the threshold after every single-stock deletion.
- Claude 3 Opus and Claude 3.5 Sonnet showed opposite directions of relative repricing.
- Two large negative GPT spreads were strongly influenced by software firms reporting earnings.

IMPLICATION FOR TRADING: The findings suggest that interpreting AI equity responses requires careful consideration of timing, portfolio legs, and constituent influence. Practitioners should be aware that large returns around AI model news may not necessarily reflect common infrastructure-software repricing, and should consider the specific impact of individual constituents. This research highlights the importance of thorough analysis in making trading decisions related to AI model news.

AI model news and relative equity repricing: Evidence from infrastructure and software stocks

Abstract

We examine whether large returns around AI model news reflect common infrastructure–software repricing or constituent-specific movements. An external, frozen database supplies 28 frontier-model dates. No database-session spread exceeds 1.96 trailing standard deviations in absolute value, but seven two-session spreads do. Only two remain above the threshold after every single-stock deletion: Claude 3 Opus and Claude 3.5 Sonnet, with opposite directions of relative repricing. Two large negative spreads are strongly influenced by software firms reporting earnings. A broader 176-date cohort yields 24 cumulative flags but weaker evidence of an elevated median movement. The evidence shows why timing, both portfolio legs, and constituent influence matter when interpreting AI equity responses.

Keywords: artificial intelligence; event studies; abnormal returns; earnings; market breadth

JEL classification: G12; G14; O33

1 Introduction

AI advances can increase compute demand, reduce infrastructure scarcity rents, or change software margins. These channels need not imply the same relative valuation response. Recent studies associate DeepSeek with heterogeneous firm-value responses (Yang, 2025; Ge and Yang, 2025) and compute-supply exposure (Zeng, 2026). We examine a broader set of model-news dates and distinguish opposing portfolio movements from spreads dominated by individual constituents.

Event-study methods benchmark abnormal performance against historical return variation (Brown and Warner, 1985; MacKinlay, 1997). Yet a large infrastructure-minus-software spread does not necessarily represent a common transfer of value between the two groups. An earnings-related software jump can generate a negative spread without infrastructure losses, and a single-date window can miss a cumulative movement. Examining these possibilities together changes the financial interpretation of model news.

Our main sample contains 28 language-model records selected mechanically from a frozen external compute-defined database. Seven two-session spreads exceed a descriptive volatility threshold, compared with none on the database session alone. Only two of those seven retain the threshold after every single-stock deletion. Both concern Claude models, but their directions are opposite: infrastructure outperforms around Claude 3 Opus and underperforms around Claude 3.5 Sonnet. Two large negative GPT spreads instead depend strongly on software firms whose earnings windows overlap. The contribution is evidence on the incidence, direction, and constituent dependence of relative repricing across model dates. It does not establish a uniform compute-cost effect or identify announcement effects causally.

2 Data and design

2.1 Portfolios and candidate dates

Infrastructure contains NVDA, AMD, AVGO, TSM, MU, MRVL, and VRT; software/platforms contains MSFT, CRM, NOW, ADBE, SNOW, and DDOG. We retain the original equal-weighted portfolios, denoted U and D. They are narrow, judgment-selected groups with mixed business exposures: Microsoft also develops models and owns infrastructure. They do not represent an exhaustive AI industry or isolate pure upstream and downstream activities. SMCI is added in a sensitivity analysis.

Yahoo Finance adjusted prices cover 1,399 trading sessions from 4 January 2021 to 30 July 2026. The earlier observations provide estimation history; the analysis period remains 3 January 2023–30 July 2026, with 896 valid daily diagnostics and 895 complete two-session windows. Extending estimation history restores thirteen otherwise unmeasurable candidates without changing selection or estimation rules. Compustat supplies 186 in-period firm-report records on 149 calendar dates for the thirteen constituents.

The main cohort uses Epoch AI's Frontier Models CSV, downloaded and frozen on 6 September 2026; its rolling download page was labeled updated 3 September. Frontier status denotes top-ten training compute at release (Epoch AI, 2026). We retain Publication dates from 2 January 2023 through 30 July 2026, the Language domain, and Language modeling or Language modeling/generation tasks. Records labeled Unreleased in the frozen accessibility field are excluded; all confidence categories and updates remain. Same-organization, same-date records are collapsed. Of 29 in-period records, one unreleased Behemoth preview is excluded; all remaining 28 have sufficient estimation history and distinct trading dates.

This is an external retrospective rule adopted after exploratory case analysis, not preregistration or a contemporaneous database vintage. The accessibility filter describes the frozen record, not verified availability on its publication date. Compute disclosure limits coverage, and no selected date follows 9 July 2025. A separately specified Epoch Notable cohort checks broader coverage using the same language, date, and accessibility filters and a nonempty Notability criteria field. Different organizations sharing a trading session count once in its aggregate diagnostics. Notable includes editorial discretion and is not a strict superset of Frontier.

Epoch dates can refer to announcements, technical disclosure, updates, or availability. We therefore call the first US trading session on or after each date the database session. Every candidate receives the same database-session, following-session, and cumulative two-session analyses. Source records link to Supplementary Table S1; documented date conflicts enter separate scenarios without return-based replacement of the baseline. The seven original focal launches remain supplementary cases. The 27 January 2025 DeepSeek selloff is a separately identified delayed-repricing diagnostic.

2.2 Return diagnostics and comparisons

For each stock, an intercept and SPY coefficient are estimated using exactly the preceding 250 returns. The current return is excluded. Equations (1) and (2) define abnormal returns and the equal-weighted portfolio spread:

$$AR_{i,t} = r_{i,t} - (\alpha_{i,t} + \beta_{i,t}r_{m,t}) \quad (1)$$

$$S_t = AR_{U,t} - AR_{D,t}, \quad z_t = S_t / \sigma_{S,t} \quad (2)$$

Daily z divides the spread by its preceding 60-session sample standard deviation; ranks order absolute z across the 896 analysis sessions. Both portfolio legs and their separately standardized returns are retained. The 1.96 cutoff is descriptive, not a calibrated 5% significance test or a multiple-testing correction.

The cumulative diagnostic in Eq. (3) sums the database and following sessions' abnormal returns, retaining each day's rolling coefficients. Its denominator uses 60 historical overlapping two-session sums ending before the database session. The second-day coefficient fit therefore includes the first-day return; this is cumulative aggregation of the original daily diagnostic, not a conventional fixed-parameter event-window estimator.

$$CAR_{S,t} = S_t + S_{t+1}, \quad z_t^{(2)} = CAR_{S,t} / \sigma_{S,t}^{(2)} \quad (3)$$

Threshold breadth is the fraction of portfolio stocks whose absolute abnormal return exceeds twice its own historical standard deviation. Cumulative breadth uses stock-level pair sums and their corresponding

historical-pair volatility. Zero breadth means no threshold exceedances, not no participation. Concentration is the largest absolute constituent return divided by their absolute-return sum. Single-stock deletion reweights the remaining stocks while retaining the original spread denominator. We report whether every deletion retains the threshold, alongside unthresholded constituent signs and both portfolio legs; deletion stability is an influence diagnostic, not a separate significance test.

Earnings windows cover the first trading session on or after Compustat RDQ and the next session. Cumulative overlap takes the union across both event sessions. No dates are discarded for overlap. Historical comparisons include 20,000 random-date draws and circular calendar shifts. The latter preserve relative event spacing; for cumulative analyses they shift starts over the valid historical-pair series. These descriptive reference distributions cannot establish random assignment or remove contemporaneous news.

3 Results

3.1 Timing and the incidence of large relative movements

The 28 database sessions have median absolute spread z of 0.55 and maximum 1.82; none enters the twenty largest historical absolute z values (Table 1). Mean threshold breadth is 0.041 for infrastructure and 0.054 for software, compared with historical means of 0.051 and 0.053. Nine dates overlap earnings windows. A modest spread can mask substantial movements in both legs: on the Gemini 1.5 Pro date, their abnormal returns are -2.59 and -2.52 percentage points, respectively.

Seven of 28 absolute cumulative z values exceed 1.96, versus 73 of 895 historical pairs: 25.0% against 8.2%. Their median absolute cumulative z is 0.92, with median tail probabilities of 0.023 under random-date draws and 0.027 under circular shifts. Four flagged windows have no constituent earnings overlap and three do. These results contradict a general conclusion that the selected model dates exhibit little relative movement. They also show why earnings overlap cannot explain every large cumulative spread.

The following-session specification has four flags: GPT-4 Turbo in November 2023, Claude 3 Opus, and the November 2024 and January 2025 GPT-4o updates. Three overlap earnings. Daily median comparison probabilities are 0.649 for draws with replacement, 0.651 without replacement, and 0.664 for circular shifts. QQQ adjustment keeps database-session spreads below 1.96; adding S&P 500 produces one flag, Claude 3 Opus, at $z = 2.56$. Thus, the one-day result also depends on portfolio composition. Source-driven calendar scenarios retain seven cumulative flags; a separate later-availability scenario yields eight. The supplement distinguishes these milestones and reports confidence restrictions.

[Insert Tables 1 and 2 here]

3.2 Direction and constituent influence

Table 2 reports all seven large cumulative spreads. Three pair infrastructure losses with software gains, three pair infrastructure gains with software losses, and one has losses in both legs. Five fall below the descriptive threshold under at least one single-stock deletion. Only Claude 3 Opus and Claude 3.5 Sonnet retain it under every deletion, and their directions differ.

On 4–5 March 2024, the Opus window has infrastructure cumulative abnormal return of $+1.98$ percentage points and software return of -4.04 , producing spread $z = 3.09$. Six of seven infrastructure stocks gain and all six software stocks lose. The minimum absolute z across deletions is 2.65, rounded. On 20–21 June, Sonnet has infrastructure return of -3.72 and software return of $+2.87$, with $z = -2.65$. Six infrastructure stocks lose and all six software stocks gain; every deletion retains absolute z above 2.29. Neither window overlaps sample-firm earnings. These are coherent relative movements in opposite directions, although other news and changing exposures remain possible explanations.

The November GPT episodes illustrate a different pattern. For GPT-4 Turbo on 6–7 November 2023, software gains 5.90 percentage points while infrastructure loses 0.58; cumulative spread z is -4.11 . Datadog accounts for 68.9% of software absolute cumulative abnormal returns, and removing it reduces absolute spread z to 1.76. Its earnings window overlaps. For GPT-4o on 20–21 November 2024, z falls from -3.06 to -1.02 after removing Snowflake, also in an overlapping earnings window. On the following session alone, infrastructure actually gains 0.56 percentage points while software gains 6.83. That following-session negative spread therefore should not be read as infrastructure losses.

Across all 896 sessions, earnings windows cover 271 observations and contain seventeen of the twenty largest absolute standardized spreads. Mean maximum deletion influence is 0.50 spread standard deviations during earnings windows and 0.29 otherwise. Figure 1 and the supplementary concentration results place the selected dates in this historical context. These associations flag competing explanations; they do not prove that earnings caused the returns.

[Insert Figure 1 here]

3.3 Broader coverage and interpretation limits

The Notable rule selects 219 model records, or 195 organization/date events, mapping to 176 different US sessions across 2023–2026. It includes all seven original focal models and 27 of the 28 Frontier records; Nemotron Ultra lacks a notability tag. Database-session flags occur on 11 of 176 dates, following-session flags on fourteen, and cumulative flags on 24 (13.6%). Of the cumulative flags, eleven survive every single-stock deletion, fourteen overlap earnings, and ten do not. Eleven pair infrastructure losses with software gains, eleven show the reverse, and two have same-direction legs.

The broader cohort has median absolute cumulative z of 0.65, with circular-shift tail probability 0.200, compared with 0.92 and 0.027 for Frontier. The broader definition provides weaker evidence of an elevated median against its historical reference; this is not a formal test of the difference between cohorts. Both cohorts contain heterogeneous relative movements; neither supports assigning a common directional effect to all model news. Notable extends coverage through July 2026 but includes 45 model records tagged only Discretionary. Its source links and database dates are retained as metadata, not uniformly authenticated first-announcement timestamps (Supplementary Section S5).

DeepSeek-R1 illustrates a separate timing issue. Its first US trading session, 21 January 2025, has spread $z = 0.12$. Four sessions later, the retrospectively identified selloff has infrastructure return of -13.09 percentage points, software return of $+3.02$, and $z = -8.01$. Six of seven infrastructure firms cross their negative thresholds. This broad delayed movement is not another model release and cannot validate the event-selection rule.

Small diagnostics do not rule out economically meaningful effects. In a historical calibration, adding a common two-percentage-point infrastructure abnormal return produces a daily spread flag in only 16.7% of negative and 18.6% of positive scenarios. This exercise holds pre-event denominators fixed and measures detection frequency, not structural power. The supplement retains the original focal cases and a conditional intraday timing check.

4 Conclusion

Large relative AI equity movements appear more frequently in two-session windows than on database dates alone. In the compute-defined cohort, seven windows exceed the descriptive threshold, but only two retain it after every constituent deletion, with opposite directions of infrastructure–software repricing. Other large spreads depend strongly on individual software firms reporting earnings. A broader model

cohort contains additional episodes but weaker aggregate median evidence. The distinction matters for interpreting relative returns as changes in the distribution of AI-related value.

The study provides a reproducible cross-event comparison rather than a causal estimate of model-release effects. Its narrow portfolios, retrospective database selection, date ambiguity, and sensitivity to the model universe limit generalization. Reporting both legs, cumulative windows, constituent influence, and earnings overlap makes those limits visible alongside the financially meaningful episodes.

References

- Brown, S.J., Warner, J.B., 1985. Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14, 3–31. [https://doi.org/10.1016/0304-405X\(85\)90042-X](https://doi.org/10.1016/0304-405X(85)90042-X)
- Epoch AI, 2026. Data on AI Models. Frontier Models and complete AI Models CSV files and documentation. <https://epoch.ai/data/ai-models-documentation>. Data downloaded and locally frozen 6 September 2026; documentation accessed 14 September 2026.
- Ge, J., Yang, X., 2025. Domestic AI competition: Is DeepSeek a competitor or catalyst to Chinese AI firms? *Finance Research Letters* 85, 108005. <https://doi.org/10.1016/j.frl.2025.108005>
- MacKinlay, A.C., 1997. Event studies in economics and finance. *Journal of Economic Literature* 35, 13–39. <https://www.jstor.org/stable/2729691>
- Yang, X., 2025. AI competition and firm value: Evidence from DeepSeek's disruption. *Finance Research Letters* 80, 107447. <https://doi.org/10.1016/j.frl.2025.107447>
- Zeng, L., 2026. Low-cost AI and the value of compute scarcity: Evidence from the DeepSeek shock. *Economics Letters* 268, 113160. <https://doi.org/10.1016/j.econlet.2026.113160>

Table 1 Frontier database sessions and daily diagnostics

Model and source	Database session	AR U	AR D	z	Rank	B U	B D
GPT-4 (Mar 2023) [S01]	2023-03-15	-0.12	1.62	-1.13	208	0.00	0.17
SenseChat [S02]	2023-04-10	2.19	-0.27	1.54	120	0.14	0.00
PaLM 2 [S03]	2023-05-10	-0.25	2.56	-1.74	87	0.00	0.00
GPT-4 (Jun 2023) [S04]	2023-06-13	-0.08	-0.76	0.28	658	0.00	0.17
Claude 2 [S05]	2023-07-11	-0.50	0.67	-0.50	509	0.00	0.00
Falcon-180B [S06]	2023-09-06	-0.54	0.89	-0.98	261	0.00	0.00
Amazon Titan [S07]	2023-09-28	0.25	0.07	0.13	778	0.14	0.00
GPT-4 Turbo (Nov 2023) [S08]	2023-11-06	-0.53	-1.32	0.72	386	0.00	0.17
Inflection-2 [S09]	2023-11-22	-0.31	-0.02	-0.21	709	0.14	0.00
Gemini 1.0 Ultra [S10]	2023-12-06	-0.44	-0.70	0.19	729	0.00	0.00
GLM-4 (0116) [S11]	2024-01-17	0.43	1.15	-0.48	523	0.00	0.00
Gemini 1.5 Pro [S12]	2024-02-15	-2.59	-2.52	-0.06	845	0.14	0.50
Claude 3 Opus [S13]	2024-03-04	1.85	-0.81	1.82	78	0.00	0.00
GPT-4 Turbo Apr 2024 [S14]	2024-04-09	-0.55	0.15	-0.36	601	0.00	0.00
GPT-4o [S15]	2024-05-13	-0.24	0.23	-0.22	700	0.00	0.00
Nemotron-4 340B [S16]	2024-06-14	-0.26	3.04	-1.59	109	0.00	0.17
Claude 3.5 Sonnet [S17]	2024-06-20	-1.90	1.15	-1.48	134	0.29	0.00
Llama 3.1-405B [S18]	2024-07-23	0.47	0.92	-0.18	736	0.00	0.00
GPT-4o Aug 2024 [S19]	2024-08-06	-0.89	-0.34	-0.21	706	0.14	0.00
Grok-2 [S20]	2024-08-13	0.74	-0.58	0.50	506	0.00	0.00
GLM-4-Plus [S21]	2024-08-29	-0.88	1.30	-0.93	281	0.14	0.00
GPT-4o Nov 2024 [S22]	2024-11-20	-0.84	0.19	-0.59	451	0.00	0.00
GPT-4o Jan 2025 [S23]	2025-01-29	1.92	-1.32	1.12	211	0.00	0.00
Grok 3 [S24]	2025-02-18	0.52	0.61	-0.03	869	0.00	0.00
GPT-4.5 [S25]	2025-02-27	-2.73	1.53	-1.50	130	0.00	0.17
Nemotron Ultra 253B [S26]	2025-03-18	0.07	0.21	-0.05	852	0.00	0.00
GPT-4o Mar 2025 [S27]	2025-03-27	-2.55	-0.80	-0.61	441	0.00	0.00
Grok 4 [S28]	2025-07-09	-0.26	-1.66	1.05	234	0.00	0.17

Notes: U contains seven infrastructure stocks (NVDA, AMD, AVGO, TSM, MU, MRVL, VRT); D contains six software/platform stocks (MSFT, CRM, NOW, ADBE, SNOW, DDOG). AR is the SPY market-model abnormal return in percentage points, with coefficients estimated using the preceding 250 returns. z divides AR U minus AR D by its prior 60-session standard deviation; rank orders absolute z across 896 analysis sessions. B is the fraction exceeding twice individual trailing volatility in absolute value. The first US session on or after the Epoch date is a database convention, not verified first tradability. All 28 candidates are measured. Source identifiers link to a source; S01–S28 identify the full evidence and limitations in Supplementary Table S1.

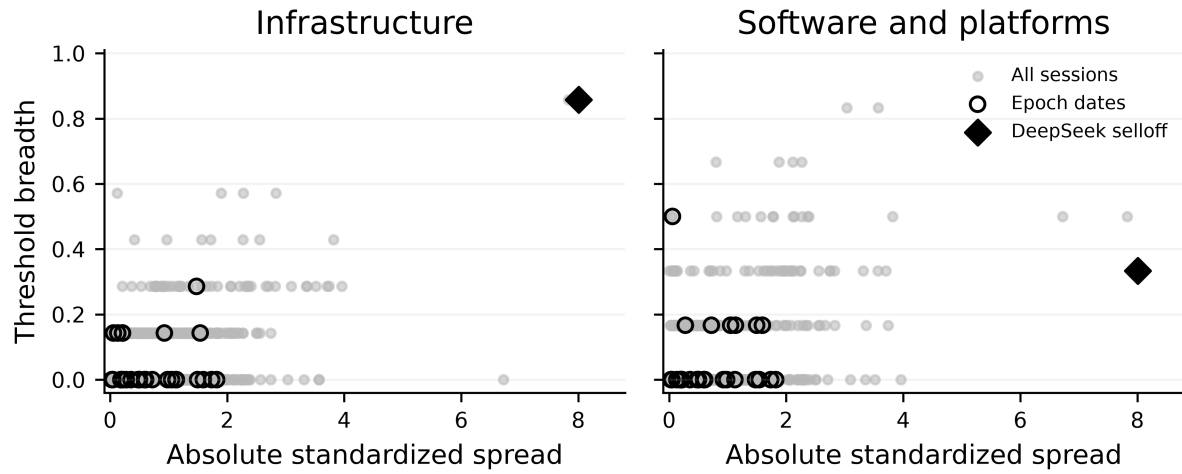
Notes: Earnings overlap: GPT-4 (Mar 2023) (ADBE); PaLM 2 (TSM); Amazon Titan (MU); Inflection-2 (NVDA); Nemotron-4 340B (ADBE); GLM-4-Plus (CRM, MRVL, NVDA); GPT-4o Nov 2024 (NVDA, SNOW); GPT-4o Jan 2025 (MSFT, NOW); GPT-4.5 (CRM, NVDA, SNOW). A report affects the first trading session on or after Compustat RDQ and the following session. Other displayed dates have no sample-firm overlap.

Table 2 Cumulative Frontier windows exceeding the threshold

Model	Two sessions	CAR U	CAR D	z_2	B_2 U	B_2 D	Min $ z_2 $ LOO	Earnings
SenseChat	2023-04-10 to 2023-04-11	1.55	-2.53	1.98	0.14	0.00	1.43	None
GPT-4 Turbo (Nov 2023)	2023-11-06 to 2023-11-07	-0.58	5.90	-4.11	0.00	0.33	1.76	DDOG
Claude 3 Opus	2024-03-04 to 2024-03-05	1.98	-4.04	3.09	0.14	0.33	2.65	None
Claude 3.5 Sonnet	2024-06-20 to 2024-06-21	-3.72	2.87	-2.65	0.29	0.00	2.29	None
GPT-4o Nov 2024	2024-11-20 to 2024-11-21	-0.29	7.02	-3.06	0.00	0.33	1.02	NVDA, SNOW
GPT-4o Jan 2025	2025-01-29 to 2025-01-30	3.53	-5.94	2.13	0.00	0.33	1.77	MSFT, NOW
Grok 4	2025-07-09 to 2025-07-10	-0.93	-4.50	1.98	0.00	0.50	1.60	None

Notes: CAR is the sum of two daily equal-weighted market-model abnormal returns, in percentage points. U and D are the seven-stock infrastructure and six-stock software portfolios defined in Table 1. z_2 divides their cumulative difference by the sample standard deviation of 60 historical overlapping pair sums ending before the database session. Daily rolling coefficient fits are retained. B_2 counts individual two-standard-deviation exceedances using historical pair volatility. Min $|z_2|$ LOO is the minimum over all thirteen single-stock deletions, reweighting remaining stocks and retaining the original denominator. Earnings covers any sample-firm report window intersecting either session. Seven of 28 candidate windows exceed $|z_2| > 1.96$, versus 73 of 895 historical pairs. The cutoff and deletion retention are descriptive, not significance tests. All candidate cumulative results are in Supplementary Table S5.

Figure 1 Daily spread magnitude and threshold breadth



Notes: Grey points represent all 896 daily analysis sessions; open circles mark 28 Frontier database dates; the diamond marks the separately identified DeepSeek selloff on 27 January 2025. The horizontal axis is the absolute standardized infrastructure-minus-software abnormal-return spread. Threshold breadth is the fraction exceeding twice individual trailing volatility in absolute value, with seven infrastructure and six software stocks. Points may overlap. This figure concerns daily diagnostics; cumulative results appear in Table 2.