

Online Appendix for
“Faster, Wider, More Fragile? The Evolution of
Intraday Shock Propagation in U.S. Equity Markets,
1992–2026”

Contents

1	Extended data audit	2
2	Additional descriptive evidence	3
3	Additional propagation results	6
4	Computational details and reproducibility	8
5	Analyses not executed	8

1 Extended data audit

Provenance and snapshot. The data are one-minute OHLCV bars originally sourced from Finnhub.io and aggregated into 411 monthly files (January 1992–March 2026). The exact byte-level snapshot is pinned: every file’s SHA-256 checksum was verified at download against the archive’s metadata and re-verified after the full download; the manifest (per-file checksum, byte size, row count, column schema, timestamp bounds, symbol count, duplicate count) accompanies the replication package. The archive carries no explicit licence; the minute data are used for research with attribution and are not redistributed.

Calendar and timezone. Observed trading dates match the official XNYS calendar exactly over the full sample: zero missing sessions, zero dates outside the calendar, zero weekend observations. Timestamps are UTC at bar-start; mapping to Eastern Time was verified empirically on both sides of every daylight-saving transition (the regular session opens at 13:30 UTC in summer months and 14:30 UTC in winter months throughout). The regular-hours share of bars by calendar month is flat (96.75%–97.09%), ruling out any DST mishandling: an offset error would displace roughly 15% of session rows in March and November. Early-closing sessions are truncated at the true exchange close; residual post-close prints (7.3% of morning volume on 2025-12-24 versus 60.4% on the adjacent full day) are labelled after-hours.

Integrity. Across 7,686,275,524 bars: 100 OHLC-impossible bars (1.3 per 10^8 , concentrated 1993–1995), zero non-positive prices, zero zero-volume or negative-volume bars, 10,128 duplicate ticker-minute keys (1.3 per million; merged as open=first, high=max, low=min, close=last, volume=sum). Prices are unadjusted for corporate actions; volumes are in shares (verified against public daily volumes for benchmark securities).

Identity and survivorship. The panel spans 81,306 tickers forming 221,957 continuous spells (a gap exceeding 30 calendar days starts a new spell); 29,112 tickers have multiple spells, so all long-horizon statements operate within spells. Delisted securities are present throughout (exits run 800–1,600 per year); the single discrete coverage expansion (2020, over-the-counter names) is absorbed by fixed-rule lagged universes.

Corporate actions. A conservative overnight-ratio screen (canonical ratios 2 through 8, 10, 20, and $3/2$, with their inverses, $\pm 1.5\%$ tolerance, both-day dollar volume $\geq \$50,000$, prior close $\geq \$1$) detects 6,944 events (4,819 splits, 2,125 reverse splits); event days ± 1 trading day are excluded from every return-based statistic.

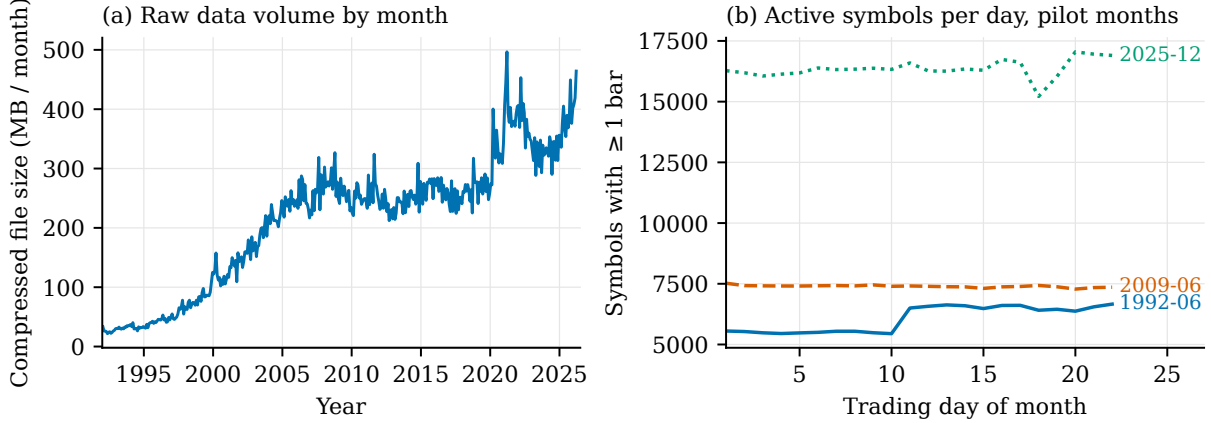


Figure 1: Dataset coverage. Panel (a): compressed size of each monthly source file, 1992–2026 (all 411 files of the pinned snapshot). Panel (b): number of symbols with at least one bar per trading day in the three pilot months.

Pilot audit. Before the full download, three pilot months (June 1992, June 2009, December 2025; 58.8 million bars) were audited. Trading-day counts match the exchange calendar in all three, including the December 24, 2025 early close; no OHLC-consistency violation, null value, negative volume, or zero-volume bar occurs; 14 duplicate ticker-minute pairs (all in inactive over-the-counter names) appear in December 2025. Liquid securities are densely covered throughout: the four most active tickers in December 2025 have complete 390-bar regular sessions, and large issuers already average 277–369 bars per day in June 1992. Volume units were verified against public daily volumes for benchmark securities. Figure 1 summarizes raw data volume and per-day symbol counts, Figure 8 the pilot intraday seasonality, and Table 2 pilot summary statistics for the 500 highest dollar-volume securities in each pilot month.

2 Additional descriptive evidence

Figures 2–6 complement the descriptive results of the main text: universe entry/exit dynamics, dollar volume concentration, the cross-section of realized volatility, the stale-price share, and the time series of detected corporate-action events. Table 1 reports decade-level summary statistics for the top-1000 universe.

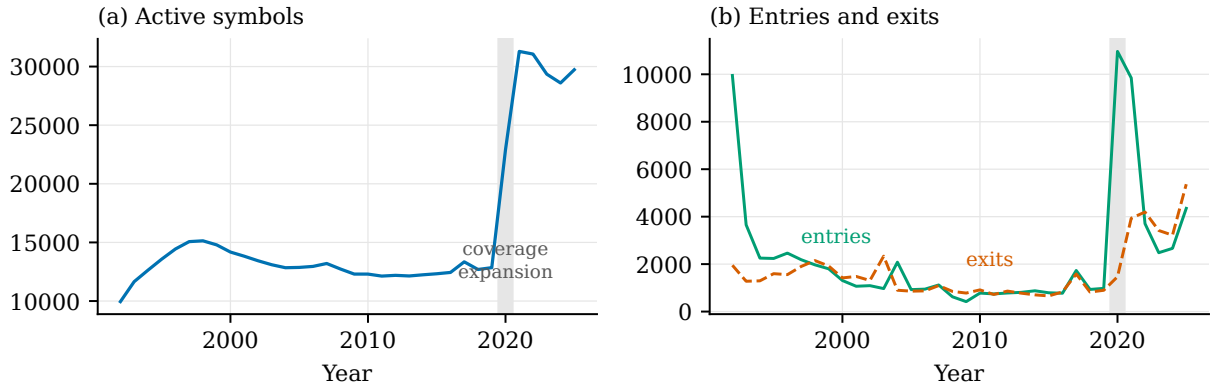


Figure 2: Universe dynamics. (a) Active symbols per year; (b) entries and exits. The shaded band marks the 2020 source-coverage expansion.

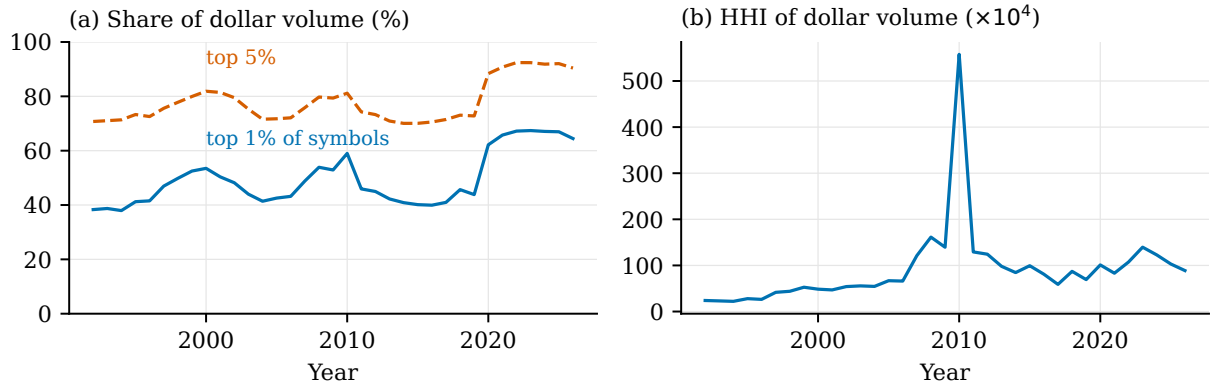


Figure 3: Concentration of trading. (a) Share of aggregate dollar volume of the top 1% and 5% of symbols; (b) Herfindahl index of dollar volume.

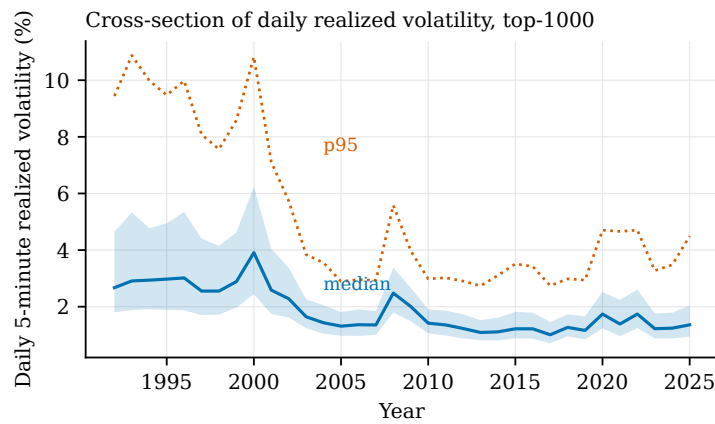


Figure 4: Cross-section of realized volatility. Median, interquartile range, and 95th percentile of daily 5-minute realized volatility, lagged top-1000 universe.

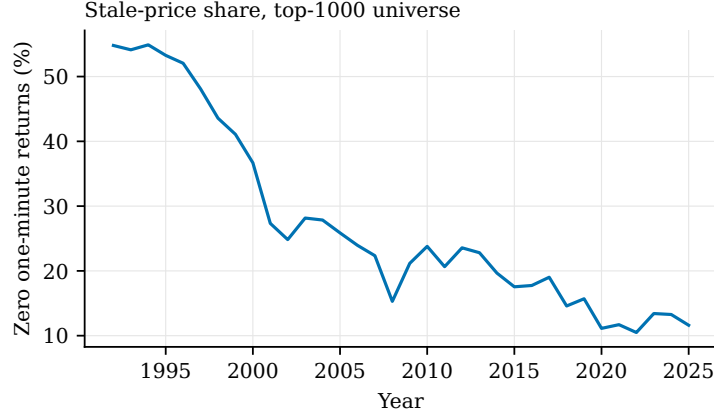


Figure 5: Stale-price share. Percentage of one-minute returns exactly zero, lagged top-1000 universe (52% in 1996, 13% in 2024).

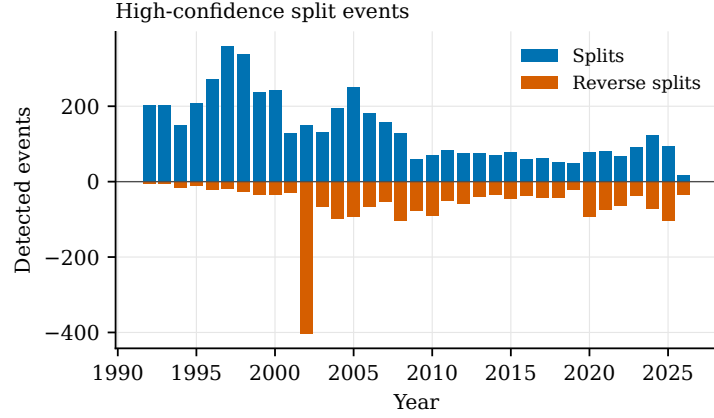


Figure 6: Detected corporate-action events per year. Forward splits (up) and reverse splits (down) from the overnight-ratio screen.

Table 1: Summary statistics by decade, top-1000 universe

Decade	Variable	Mean	SD	p1	p5	p50	p95	p99	N
1992–1999	Minute log return (bps)	-0.17	90.42	-158.31	-65.57	0.00	65.09	157.48	8,844,318
	High–low log range (bps)	17.60	93.48	0.00	0.00	0.00	83.22	190.48	8,890,611
	Dollar volume per bar (\$000)	228.72	910.93	1.60	3.12	42.09	948.24	2,942.87	8,890,611
2000–2009	Minute log return (bps)	-0.04	333.07	-61.16	-23.50	0.00	23.17	60.61	21,554,950
	High–low log range (bps)	14.42	348.96	0.00	0.00	5.26	47.73	113.00	21,619,622
	Dollar volume per bar (\$000)	386.36	23,791.54	2.30	5.35	87.62	1,315.95	4,372.70	21,619,622
2010–2019	Minute log return (bps)	-0.01	45.04	-26.39	-12.54	0.00	12.49	26.21	23,038,817
	High–low log range (bps)	7.26	64.62	0.00	0.00	4.62	22.26	42.64	23,103,541
	Dollar volume per bar (\$000)	586.51	3,504.50	5.11	14.99	173.84	1,939.16	6,120.19	23,103,541
2020–2026	Minute log return (bps)	0.00	13.10	-36.19	-17.14	0.00	17.17	36.24	22,114,997
	High–low log range (bps)	9.47	14.11	0.00	0.00	5.85	30.53	59.57	22,180,658
	Dollar volume per bar (\$000)	1,338.66	7,744.21	15.80	41.87	337.03	4,125.48	16,852.31	22,180,658

Notes: Regular trading hours, lagged top-1000 dollar-volume universe, OHLC-violation and split-window rows excluded. Statistics from a stratified three-months-per-decade sample (Mar/Jun/Sep of an early, middle, and late year); full-decade moments in the Internet Appendix.

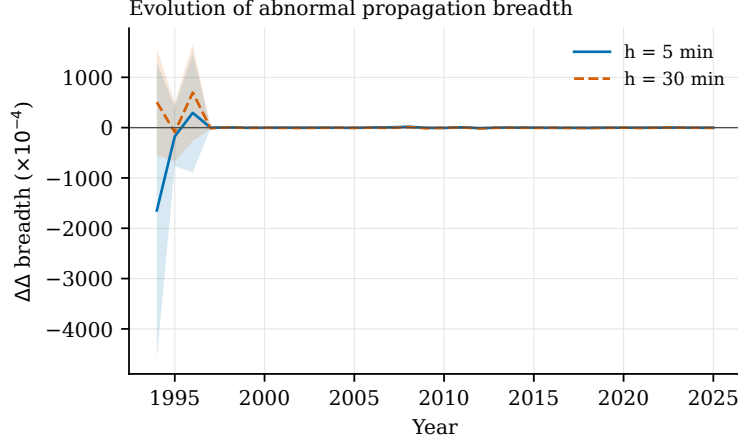


Figure 7: Evolution of abnormal market-wide propagation. $\Delta\Delta$ breadth by year at $h = 5$ and $h = 30$ minutes, with 95% day-block bootstrap intervals.

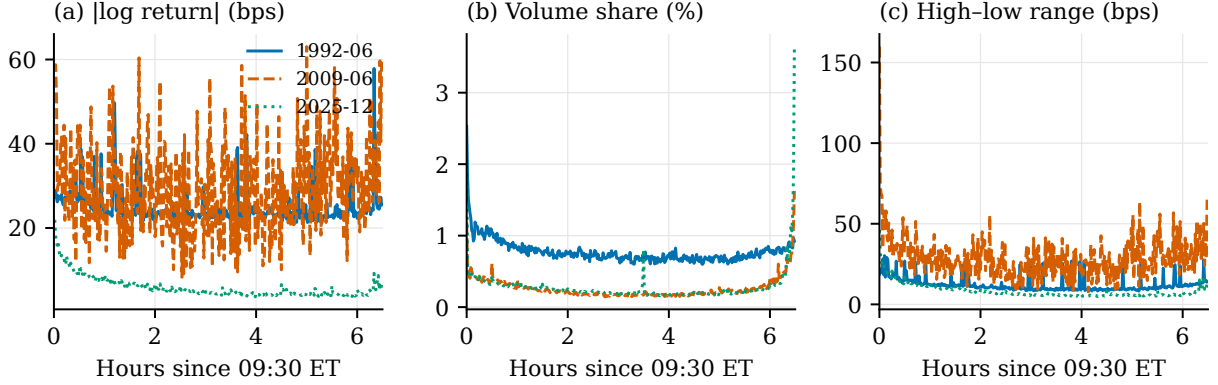


Figure 8: Pilot-sample intraday seasonality (June 1992, June 2009, December 2025; top-500 by within-month dollar volume).

3 Additional propagation results

Figure 7 shows the year-by-year evolution of abnormal market-wide breadth at $h = 5$ and $h = 30$ with 95% day-block bootstrap bands: the market-wide difference-in-differences is within two standard errors of zero in essentially every year, not merely on average. The design diagnostics (pre-event levels, co-arrival classes, pre-level-stratified estimates, alternative clustering, signed responses) and the decomposition of the network differential by decade are reported in the main text.

Table 2: Summary statistics, pilot months

Month	Variable	Mean	SD	p1	p5	p50	p95	p99	N
June 1992	Minute log return (bps)	-0.17	255.04	-145.99	-66.01	0.00	63.90	145.28	1,374,563
	High-low log range (bps)	12.57	230.25	0.00	0.00	0.00	63.09	155.85	1,385,486
	Dollar volume per bar (\$000)	140.01	824.08	1.86	3.06	31.32	579.15	1,561.88	1,385,486
	Valid RTH bars per stock-day	126.84	79.58	19.00	35.00	107.00	303.00	369.00	10,923
June 2009	Minute log return (bps)	-0.16	1,172.84	-38.30	-19.51	0.00	19.18	37.47	4,190,657
	High-low log range (bps)	28.67	945.13	0.00	0.00	9.49	36.04	65.47	4,201,656
	Dollar volume per bar (\$000)	886.22	19,849.91	9.21	28.84	285.18	2,807.64	8,382.52	4,201,656
	Valid RTH bars per stock-day	382.00	31.56	270.00	357.00	389.00	390.00	390.00	10,999
December 2025	Minute log return (bps)	-0.02	10.72	-32.15	-14.12	0.00	14.02	31.78	3,904,759
	High-low log range (bps)	8.33	12.25	0.00	0.00	4.66	28.94	57.64	3,915,747
	Dollar volume per bar (\$000)	2,833.57	13,640.57	58.50	142.61	846.13	8,849.93	37,017.92	3,915,747
	Valid RTH bars per stock-day	356.37	53.42	166.00	223.00	383.00	390.00	390.00	10,988

Notes: Regular trading hours (09:30–16:00 ET, exchange calendar) for the top 500 securities by within-month dollar volume in each pilot month of the pinned Finnhub.io minute-bar aggregation. Minute log returns use consecutive same-day closes; the high–low range is $\log(H) - \log(L)$ per one-minute bar; bars exist only for minutes with at least one reported trade. Prices are unadjusted for splits; the pilot is descriptive and imposes no corporate-action screens.

4 Computational details and reproducibility

The pipeline is deterministic end-to-end. All parameters live in versioned configuration files; the global seed (20260713) governs placebo draws and bootstrap resampling; every table and figure is generated by a numbered script from machine-readable intermediate results, and every numerical claim in the manuscript resolves to a macro generated from those results—no number is typed by hand. The full run comprises: download and checksum verification (411 files); per-month integrity audit and daily metadata construction; clean-panel construction (duplicate resolution, session labelling with true exchange closes, integrity and corporate-action flags); universe construction with strictly lagged screens; descriptive aggregation; shock detection and outcome measurement (1,854,384 events, 1,766,058 placebos); network and co-jump measurement; and estimation. Every stage operates month-by-month on partitioned columnar files with four parallel workers on an Apple-silicon workstation; wall-clock times for each stage are recorded in the run logs of the replication package. The revision diagnostics (pre-event levels, co-arrival classes, stratified estimators, alternative clustering) are recomputed from the stored event-level files in a few minutes without access to the raw minute panel.

5 Analyses not executed

The following checks are referred to in the main text as not executed, so that no claim rests on them: source thresholds $|z| \geq 4$ and $|z| \geq 5$; receiver threshold $|z| > 3$; the top-1000 and top-3000 shock universes; neighbor sets estimated on 5- or 30-minute residual correlations or with a transaction-time covariance estimator; the pre-event ($t - 5$) network contrast; permuted-network placebos preserving industry and liquidity composition; propagation measured inside scheduled-announcement windows; a Hawkes-type model of the co-jump minute series; and a reconciliation of bar volumes against consolidated-tape volumes to establish the venue scope of the source. Each requires re-running the detection stage on the raw minute panel or external data.