Supplementary Materials

Same Benchmark, Different Leverage Targets: Robinhood Holder-Count Additions after Benchmark Gains and Declines

These Supplementary Materials accompany the article and contain the notes, figure, and tables that the article and its Appendix A cite as Notes S1–S5, Figure S1, and Tables S1–S17. Note, table, and figure numbers with an S prefix refer to this document; all others refer to the article. Unless stated otherwise, estimates, standard errors, and intervals are reported as additions per 1,000 prior holders. The primary and secondary analyses were specified in an analysis plan time-stamped on the Open Science Framework (OSF) on August 31, 2026, at 16:51 UTC, before any outcome-return estimation. An OSF technical archiving failure prevented completion of the registration initiated at 16:52 UTC that day; formal registration was completed on September 23, 2026, after the primary and secondary results were known (https://osf.io/7kyd5). Note S5 documents the time stamps, the file hashes, and the record of the first estimation run, whose output was saved at 16:56 UTC. Every quantity carries the two labels used in the article. Its provenance records how it was specified relative to the analysis plan: primary; secondary; mandatory robustness or mandatory output; exploratory, prespecified in the plan; or exploratory, added after the primary and secondary results were known. Its evidential role records what it can bear on: only the primary test supplies the decision on H-A. Nothing here alters the prespecified estimates or decision rule; no file in the plan package, and no file whose hash its manifest records, has changed since August 31, 2026; every script and output produced after the primary and secondary results were known is archived in dated folders kept apart from the plan package, which has not changed.

Contents. Note S1 records the use of generative AI and the verification steps; Note S2 describes the archive-adjacent series that continue through their splits and the exemplar used to calibrate the gross-integrity screen; Note S3 outlines the mechanics of fixed-effect removal, block resampling, family critical values, and the pooled estimate; Note S4 describes the added segments; and Note S5 documents the registration record. Figure S1 plots the two archive-adjacent series around their splits. Among the tables, S1 to S3 document sample construction, the interior and boundary components of additions and removals, and the inference diagnostics; S4 to S6 report the estimand crosswalk, the correlated-channel specifications, and the prespecified exploratory sample variants; S7 to S11 report influence, the class-level sensitivity exercise, the conditioning and timing variants, and the component and roster-by-window results; S12 reports the stamp audit; S13 to S16 give the full estimand crosswalk, the ledger of departures from the analysis plan, the inference key, and the inventory of intervals for the prespecified analyses; and S17 lists the registration timeline and its sources.

Supplementary Notes

Note S1. Use of Generative AI and Verification Steps

Portions of this study's computational work were prepared with generative-AI assistance. The author used Claude (Anthropic; Claude Opus 5 model, accessed through the Claude desktop application) for data-engineering support (archive extraction, session-clock reconstruction, and integrity-screening code), drafting calibration and estimation code, auditing the reported numbers, and manuscript editing. The author owns all research questions, the prespecified design, the estimand definitions, and every analytical decision.

The verification steps were as follows. First, the author specified each analysis before it was coded, then reviewed the resulting code and its outputs, including those from post-result additions. Second, the outputs

were verified against the archived machine-readable results and calibration values. Third, the deterministic pipeline was re-run from the archived seeds. Fourth, the numbers reported in the text and tables were checked against those outputs. Fifth, for analyses added after the primary and secondary results were known, the scripts first reproduced the primary estimate, its standard error, and its calibrated critical value, and the S6 estimate, to numerical tolerance before any new quantity was computed; the critical values of the components and of $\Delta_2(2\%)$ in Table 4 reuse the original calibration draws, and the reproduced primary critical value and Monte Carlo *p* value are identical to the archived values. Sixth, every row refitted to obtain a block-clustered standard error first reproduced its archived point estimate and standard error; rows whose block-clustered standard errors were already archived were not refitted, and their intervals were built from the archived values. The $\Delta_3(5\%)$ critical value was scored on the original calibration draws after the primary critical value had been reproduced exactly, and the stamp audit was restricted to the exact retained fund-session keys written by the sample builder in the plan package. The checks in the second through sixth steps were scripted and run with generative-AI assistance, and the author reviewed their findings; the audit of the reported numbers ensured that the seven values listed in Table S14 (item 20) are taken directly from the source outputs, and a source-to-table manifest, archived with the post-estimation scripts and outputs, names the output files behind every table and figure. Responsibility for all code, analyses, results, and text rests entirely with the author.

Note S2. Archive-Adjacent Series and the Gross-Integrity Exemplar

Two archive-adjacent series, USO and UCO, continue through their April 2020 reverse splits and illustrate the hazard complementary to the truncation documented in Section 3.4 (Figure S1). Neither fund appears in any estimation sample. USO and UCO exhibit one-day holder-count jumps of 41 to 55 log percent ($100 \times \Delta\log H$) around their splits, and the identified jump days alone account for 51% and 28%, respectively, of each fund's total absolute log holder-count change over April and May 2020. Several of those jumps precede the ledger's split dates. The jumps may reflect participation, reporting changes, or instrument-mapping changes amid the oil-market disruption; the archive cannot distinguish among these explanations, so any treatment of those dates must be explicit.

A deliberately retained bad-series exemplar (STZ, whose records alternate between duplicate values) was used to calibrate the gross-integrity screen applied to every candidate series. STZ's alternation share is 0.726, compared with a median of 0.010 across all series.

Leverage targets in the family map are effective-dated, so a change in target within the sample window cannot contaminate the coding of the leverage target. One issuer batch reduced the targets of ten funds from 3x to 2x, effective March 31, 2020; no fund in the confirmatory sample was affected.

Note S3. Mechanics: Fixed-Effect Removal, Block Resampling, Family Critical Values, and the Pooled Estimate

Fixed-effect removal. The fixed effects of Equation (1) are removed by iterated weighted demeaning to a convergence tolerance of 10^{-10}, with the design rank verified at each estimation.

Block resampling. The prespecified five-session block length is consistent with the serial dependence in the projected outcome (mean first-order autocorrelation 0.13 across cells, maximum 0.32). Across block lengths of 1, 5, 10, and 20 sessions, the critical value lies between 2.09 and 2.17 (Table S3). On the frozen sample, the measured size of a naive 1.96 rule is 6.1% under the prespecified simulation scheme, so uncalibrated inference would be modestly anti-conservative relative to the joint null that the scheme imposes. Applying the algorithm to a different contrast does not, by itself, establish validity for that contrast's composite null.

Family critical values. The family critical values in Table A3 deviate from the normal reference in both directions because each family's resampled null preserves dependence among its cells and cross-family session shocks, whereas the HC1 denominator treats observations as independent. For the 20-year Treasury family, the resampling standard deviation of the contrast under the null is 2.8 per 1,000, compared with a mean analytic standard error of 4.1, which compresses the null $|t|$ distribution and its critical value to 1.32 ($t = -1.72$; $p = .011$). For the 7–10-year Treasury family, the corresponding figures are 24.7 against 12.3, which widens the critical value to 4.02.

Pooled estimate and family average. The unweighted average of the fully interacted family estimates is −0.50 per 1,000, whereas the pooled primary estimate is 0.59; the two differ by 1.09, nearly twice the primary point estimate, and no algebra forces them to agree. The pooled model constrains the limb coefficients to be common across families. Its contrast is therefore a linear functional of the projected outcomes, with the effective absolute-weight mass distributed unevenly across families (from 8% on the S&P 500 family to 19% on the 20-year Treasury family). It is not a convex combination of the fully interacted family estimates, which allow family-specific curvature.

Family-sign cross-check. The prespecified cross-check enumerates all $2^8 = 256$ family-level sign assignments on the fixed-effect-projected data, null-imposed and CR1-studentized, without reprojecting the fixed effects in each assignment. It is reported as the plan specifies ($p = .758$).

Note S4. Descriptives of the Added Segments

Table A4 reports the descriptives for the added segments. The two added pairs with the largest segment-level estimates are associated with the most volatile benchmark proxies in the sample: daily standard deviations of 2.2% for the semiconductor proxy and 2.1% for the S&P biotechnology proxy, compared with 1.5% to 1.8% for the broad equity ladders and 0.4% to 1.0% for the Treasuries. The other added proxies range from 1.7% to 1.9%, at or just above the broad-ladder range. The semiconductor and biotechnology +3x funds post the largest mean absolute own returns in the sample, 4.6% for LABU and SOXL against 3.2% for TQQQ, and have holder bases comparable to the broad-index +3x funds (medians of 2,361 and 3,015 against 2,870 for UPRO), whereas the +2x sector pairs have bases of 12 to 219. These differences in volatility and holder base are unconditional. Conditional on the same 2% benchmark move, every +3x fund targets roughly a 6% own return, so the sector funds do not move more at the prespecified evaluation point; a 2% move is closer to an ordinary session on their benchmarks.

The +3x fund has the largest holder base among the leveraged funds in seven of the eight primary ladders and in the Russell 2000 ladder. TQQQ alone (median 9,458 holders) exceeds every +2x fund's base by an order of magnitude (Table A4). The +1x anchor, however, often has more holders than either leveraged fund. None of these descriptors discriminates among the readings listed in Section 6.1.

Note S5. Registration Record and Verification

The prospectively specified analysis plan and the files it governs were bundled into a single archive, the plan package (OSF submission bundle 2026-08-31.zip; 2,277,665 bytes; SHA-256 902ea76986dfae1e8a753339573c2e889922bd944b69f9e52318b256073c607b). Its manifest, osf_file_manifest_sha256.txt, lists the SHA-256 hash of each of the other 111 files in the package and pins the raw Robintrack archive, which is not attached, by its own hash. Table S17 lists the events from the creation of the package to the public release of the registration, with the source of each time and how a reader can check it.

Two of these records are kept by OSF and can be checked there. The package's only revision in OSF Storage was uploaded at 16:51 UTC on August 31, 2026, and the registration's activity log records that upload and the registration initiated at 16:52 UTC. The registration completed on September 23 archives that same revision, so every file in it can be checked byte for byte against the manifest. These include the analysis plan (SHA-256 14e245c2e215a55a…), the estimator (ef73fab5f0e486d5…), the calibration script (d4d29b7f7ccf0490…), and the S6 segment roster (b9f81f48c86fb4cd…). A third record is OSF's automated notice to the author, sent at 16:52 UTC, which states that the August 31 registration could not be archived because of errors copying files from linked third-party services; the notice is available from the author. The registration's template responses restate the plan's decision rules but are not covered by a hash. Before the author submitted the registration on September 23, OSF listed the draft as last modified at 16:52 UTC on August 31, and the author made no change to it afterward; that listing was visible only to the account holder, so this record rests on the author's report.

The output of the first outcome-return estimation, the primary analysis, is the file estimate_primary_result.json. The author's file system records it as saved to the project folder at 16:56:16 UTC on August 31, and the Google Drive copy of the folder records its creation at 16:56:17 UTC. These two records come from the author's own systems and date the output's arrival in the folder: the run had finished by then, but the records do not show when it began. The author began the run after initiating the registration at 16:52 UTC; that order rests on the author's account. Before that run, the estimator in the package had been executed only in its default mode, on day-permuted outcomes; that output, dry_run_result.json, is labeled as a permuted placebo and is in the package uploaded at 16:51 UTC. Some working records written that afternoon give September 1, 2026, as the date of the run or of the freeze; a dated note filed with the replication materials corrects them, and the records themselves are unchanged.

The plan places its external freeze at OSF submission, and the estimator's header states that the real run is to be invoked once, after the OSF registration timestamp. The author treated initiating registration at 16:52 UTC as meeting both conditions, unaware that OSF's technical archiving failure had prevented completion: the project's status files, saved at 16:54 UTC, record the freeze as in effect and the estimation as authorized, and the author did not see OSF's notice until September 23, 2026, when preparing a view-only link for the reviewers showed that the registration was still a draft. The registration's responses reproduce the draft as last modified at 16:52 UTC, so their statements about what had been computed, such as that no confirmatory quantity had been estimated, describe that time. Part 5 of the responses lists eight price files; the package holds 92, as its README records: files for the eight family anchors, retrieved on August 30, and for the full set of mapped tickers, retrieved on August 31 before the package was written.

Supplementary Figure

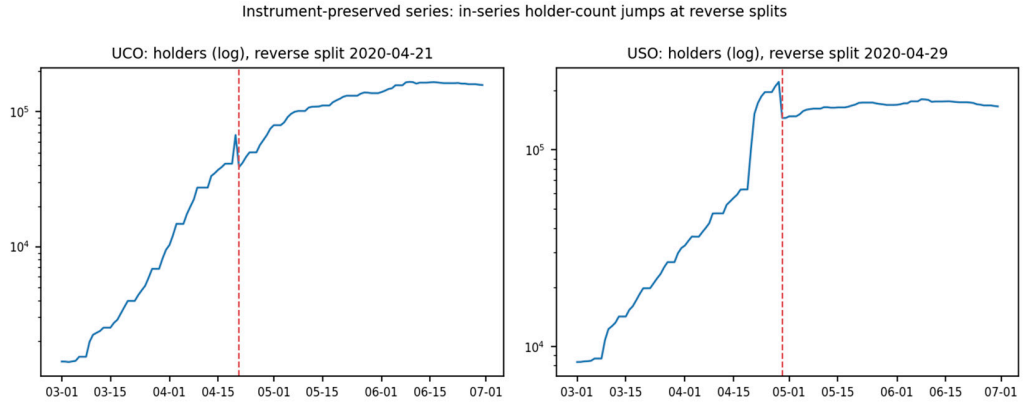


Figure S1. Archive-Adjacent Series That Continue Through Their Splits.

Note. Daily Robinhood holder counts (log scale) for UCO and USO around their April 2020 reverse splits (dashed lines). One-day jumps of 41–55 log percent ($100 \times \Delta \log H$) are recorded in-series around the split dates, several of them before the split markers; their timing flags a measurement hazard without establishing that split mechanics caused the changes.

Supplementary Tables

Table S1. Sample Attrition and Session Reconciliation.

Stage	Count	Session dates
Panel A. Sessions		
Rebuilt NYSE sessions	576	
Less: sessions with no data for any ticker (universal outages)	10	
Less: first archive session (no prior observation)	1	2018-05-02
Less: first session after a collection gap (not consecutive)	2	2019-01-30, 2020-01-16
Less: sessions adjacent to a collection outage	7	2018-08-09, 2018-08-10, 2019-01-23, 2019-01-24, 2019-01-31, 2020-01-06, 2020-01-17
Less: platform-outage sessions and the sessions after them	4	2020-03-02, 2020-03-03, 2020-03-09, 2020-03-10
Analysis sessions (primary sample)	552	
Panel B. Cell-sessions (24 designated cells)		
Candidate consecutive-observation pairs (24 designated cells)	13,559	
After consecutive-session requirement	13,510	
After outage-adjacency exclusion	13,342	
After platform-outage session exclusion	13,246	
After $H \geq 5$ and return-availability requirements	13,246	
After family-session balance (primary sample)	13,242	
S7 hourly-grid sample	13,248	

Note. Panel A reconciles the 576 rebuilt NYSE sessions to the 552 analysis sessions; the itemization is drawn from the prespecified session panel and sample construction, and sums to the totals they record. Panel B gives the attrition path summarized in Section 3.5. The S7 hourly-grid sample adds the six China cell-sessions of 2018-05-03/04 that the hourly grid bridges; eight matched outcome values change (Section 5.2).

Table S2. Interior and Boundary Components of Additions and Removals, and Within-Family Addition Shares (Primary Sample).

Family	Dose	Additions (interior)	Additions (boundary)	Removals (interior)	Removals (boundary)	Family share
FTSE China 50	+1x	4,120	531	2,998	365	0.380
	+2x	152	31	164	25	0.015
	+3x	6,294	1,122	5,978	559	0.605
Dow	+1x	22,114	2,460	5,854	1,204	0.576
	+2x	2,121	344	1,793	283	0.058
	+3x	13,615	2,043	9,382	1,976	0.367
MSCI Emerging Mkts	+1x	4,340	565	3,744	455	0.741
	+2x	68	16	79	24	0.013
	+3x	1,371	262	1,502	193	0.247
S&P MidCap 400	+1x	1,840	205	802	94	0.722
	+2x	287	34	225	32	0.113
	+3x	403	64	291	47	0.165
Nasdaq-100	+1x	43,838	4,545	9,370	1,396	0.390
	+2x	3,102	414	2,629	474	0.028
	+3x	63,339	8,713	46,105	10,242	0.581
S&P 500	+1x	107,223	8,635	10,619	1,337	0.841
	+2x	3,773	518	2,763	650	0.031
	+3x	15,583	2,016	9,278	1,809	0.128
ICE 20+ Yr Treasury	+1x	8,750	1,294	5,357	663	0.725
	+2x	262	37	219	29	0.022
	+3x	2,943	561	2,466	355	0.253
ICE 7–10 Yr Treasury	+1x	2,602	346	1,600	287	0.848
	+2x	158	16	133	22	0.050
	+3x	288	66	271	37	0.102

Note. Counts of measured holder-count additions and removals (sums of positive and of absolute negative inter-poll changes; lower bounds on gross entry and exit events). Interior changes are observed inside the session window; boundary changes span its opening boundary at the prior scheduled close. Family share is the cell's share of its family's measured additions.

Table S3. Inference Diagnostics for the Prespecified Resampling Scheme.

Diagnostic	Value
Block = 1: critical value / Monte Carlo p	2.17 / .711
Block = 5 (prespecified length): critical value / Monte Carlo p	2.09 / .713
Block = 10: critical value / Monte Carlo p	2.09 / .720
Block = 20: critical value / Monte Carlo p	2.15 / .726
Archived block-wild residual diagnostic: critical value	2.10
Projected-outcome AR(1): mean / min / max across cells	0.13 / −0.03 / 0.32
S6 10,000-draw sensitivity: critical value / interval	2.55 / [2.79, 8.12]; zero exceedances, plus-one $p <$ .001 (1/10,001)

Note. Block-length rows use an independent draw stream; the estimation-stage critical value at block = 5, from the archived seed, is 2.10 (Table 4). The archived block-wild diagnostic retains nuisance slopes and applies Rademacher signs to contrast-restricted residuals in five-session blocks at their original dates. It uses HC1 studentization without reprojecting simulated outcomes through the fixed effects; Table A2 adds that reprojection and block-based studentization. With 10,000 calibration draws, the S6 critical value is 2.55 (against 2.41 at the prespecified 2,000 draws), and the interval is 2.79 to 8.12 per 1,000.

Table S4. Estimand Crosswalk by Sample, Weighting, and Scale, With HC1, Session-Clustered, and Block-Clustered Standard Errors.

Estimand	Sample	Scale	Weights	Estimate	*SE* (HC1)	*SE* (session)	*SE* (block)	*t* (block)	95% interval (block)
Primary (confirmatory)	8 families	level rate	equal family	0.59	1.53	1.64	1.65	0.36	[−2.68, 3.86]
Equal observation	8 families	level rate	equal cell-session	0.59	1.53	1.64	1.65	0.36	[−2.68, 3.86]
Holder-base weighted	8 families	level rate	prior holder base	4.75	0.83	1.43	1.41	3.36	[1.94, 7.55]
Inverse-variance (descriptive)	8 families	level rate	1/family resid. var. (estimated)	4.04	1.06	1.26	1.11	3.65	[1.84, 6.23]
Level on PPML-retained sample	8 families (N = 12,963)	level rate	equal family	0.56	1.57	1.69	1.71	0.33	[−2.83, 3.94]
PPML (mandatory robustness)	8 families (N = 12,963)	proportional (log-mean)	model-implied mean	0.378	0.054	0.093	0.106	3.55	[0.167, 0.589]
All-valid segments (secondary S6)	16 segments	level rate	equal segment	5.46	1.05	1.31	1.20	4.54	[3.07, 7.84]

Note. Level-rate rows are in additions per 1,000 prior holders; the PPML row is in log points. The status under the analysis plan is shown by row; the remaining comparisons are exploratory. Each row states the sample, scale, and weights that define its estimand; pooled coefficients need not equal arithmetic averages of unit-level effects. The level-rate rows change the weights or the retained sample one at a time from the primary specification. The PPML row is not a one-dimension change: it jointly replaces the level link, the outcome scale, and the effective weighting, because a pseudo-likelihood mean model concentrates influence where conditional means are large. Its coefficient implies a +3x-versus-+1x conditional-mean addition ratio about 46% larger after a 2% decline than after an equal gain (exp(0.378) = 1.46). The inverse-variance row is descriptive: its weights are estimated from the same sample, and its *SE* treats them as fixed. Block-clustered standard errors use five-session blocks; the calibrated inference for the primary and S6 rows appears in Tables 4 and 5. The last column is the estimate plus or minus the 0.975 quantile of t with $G - 1$ degrees of freedom ($G = 111$ blocks) times the block-clustered *SE*: an individual analytic interval, not the prespecified calibrated construction (Table S15). For PPML it is 0.167 to 0.589 log points, a ratio of 1.18 to 1.80. PPML drops 279 observations in zero-outcome fixed-effect groups: MidCap 180, 7–10-year Treasury 81, Emerging Markets 6, 20-year Treasury 6, Dow 3, and China 3.

Table S5. Correlated-Channel Specifications, Primary Sample (Near-Zero Baseline; HC1 and Block-Clustered Standard Errors).

Specification	Δ_3(2%)	*SE* (HC1)	*t* (HC1)	*SE* (block)	*t* (block)	95% interval (block)
Baseline (channel subsample)	0.65	1.53	0.43	1.65	0.40	[−2.62, 3.93]
+ own-fund return limbs	−6.55	6.86	−0.95	8.03	−0.82	[−22.45, 9.36]
+ abnormal volume	0.56	1.53	0.36	1.65	0.34	[−2.71, 3.82]
+ overnight-gap limbs	2.25	2.24	1.00	2.72	0.83	[−3.15, 7.64]

Note. Estimates, standard errors, and intervals are in additions per 1,000 prior holders. $N = 13{,}214$, the primary-sample channel subsample with own-price data. With a near-zero baseline, absorption shares are not meaningful; the movement of the own-return row reflects collinearity (its standard error rises more than fourfold). These rows change the estimand and are near-collinear reparameterizations, not mediation tests. The block-clustered columns were computed after the script reproduced each archived estimate; the interval uses the t reference with 110 degrees of freedom (Table S15).

Table S6. The Primary Contrast Under the Prespecified Exploratory Sample Variants (HC1 and Block-Clustered Standard Errors).

Specification	Estimate	*SE* (HC1)	*t* (HC1)	*SE* (block)	Blocks	95% interval (block)
Pre-COVID sessions (through 2020-02-19)	0.40	1.98	0.20	1.79	87	[−3.15, 3.96]
Excluding five largest S&P decline sessions	0.95	1.53	0.62	1.78	110	[−2.57, 4.48]
March 2020 platform-outage sessions retained	0.78	1.57	0.50	1.85	112	[−2.89, 4.45]
Holder-base floor $H \geq 25$	0.12	2.12	0.06	2.19	111	[−4.21, 4.45]
Holder-base floor $H \geq 50$	2.49	2.63	0.94	3.75	111	[−4.94, 9.91]
Leave out FTSE China 50	0.16	1.73	0.09	1.87	111	[−3.55, 3.86]
Leave out Dow Jones Industrial Avg.	0.04	1.59	0.02	1.73	111	[−3.39, 3.47]
Leave out MSCI Emerging Markets	1.17	1.61	0.73	1.78	111	[−2.36, 4.70]

Specification	Estimate	*SE* (HC1)	*t* (HC1)	*SE* (block)	Blocks	95% interval (block)
Leave out S&P MidCap 400	0.77	1.54	0.50	1.48	111	[−2.16, 3.70]
Leave out Nasdaq-100	−0.38	1.62	−0.23	1.77	111	[−3.88, 3.13]
Leave out S&P 500	0.64	1.57	0.41	1.74	111	[−2.81, 4.08]
Leave out ICE 20+ Year Treasury	3.01	1.86	1.62	1.81	111	[−0.57, 6.60]
Leave out ICE 7–10 Year Treasury	0.04	1.61	0.02	1.71	111	[−3.35, 3.42]
High-VIX regime (VIX(t−1) > 15.84)	0.62	1.88	0.33	2.18	56	[−3.75, 5.00]
Low-VIX regime	1.68	2.58	0.65	2.12	56	[−2.57, 5.94]
Before fractional shares (pre 2019-12-12)	0.22	2.15	0.10	1.87	80	[−3.50, 3.94]
After fractional shares	1.75	2.25	0.77	2.33	31	[−3.02, 6.51]
Memo: exploratory contrast added after the primary and secondary results were known (calibrated inference)						
Primary sample without both Treasury ladders (*N* = 9,930; critical value 2.18; *p* = .021)	4.00	1.52	2.63	1.46	111	[0.68, 7.33] (calibrated)

Note. Estimates, standard errors, and intervals are in additions per 1,000 prior holders. The analysis plan lists these sample variants as exploratory. Each row changes the sample and therefore the estimand, and none establishes a zero effect. The block-clustered columns were computed after the script reproduced each archived estimate and HC1 standard error: the interval is the estimate plus or minus the 0.975 quantile of *t* with (blocks − 1) degrees of freedom times the five-session-block clustered *SE* (Table S15); blocks are formed from consecutive retained sessions of each subsample, so their number differs by row. It is an individual analytic interval, and every one of the 17 includes zero; the memo row shows its calibrated interval instead. The five excluded decline sessions are March 11, 12, 16, and 18 and June 11, 2020, removed as both the current and the lagged session. The VIX split uses the prior-session VIXCLS level (median 15.84); a 20-session realized-volatility split classifies 82.6% of sessions identically.

Table S7. The Ten Most Influential Observations for the Primary Contrast (Mandatory Output).

Rank	Family	Leverage target	Session	Addition rate (per 1,000)	Prior holders	Change in $\Delta_3(2\%)$ when dropped
1	S&P MidCap 400	+3x	2020-03-23	217.4	46	−0.29
2	ICE 7–10 Yr Treasury	+2x	2018-10-11	333.3	9	0.29
3	ICE 20+ Yr Treasury	+1x	2020-04-22	3.9	3,894	0.24
4	ICE 20+ Yr Treasury	+2x	2020-04-22	219.5	41	0.20
5	ICE 7–10 Yr Treasury	+3x	2018-10-11	0.0	10	−0.20
6	ICE 20+ Yr Treasury	+3x	2020-03-11	12.4	971	0.19
7	ICE 7–10 Yr Treasury	+1x	2018-10-11	49.8	261	0.19
8	ICE 20+ Yr Treasury	+3x	2020-04-22	13.5	669	−0.18
9	ICE 20+ Yr Treasury	+1x	2019-01-04	9.7	1,443	0.17
10	FTSE China 50	+1x	2020-07-06	24.2	3,015	0.17

Note. Leave-one-observation-out change in the primary contrast, in additions per 1,000 prior holders, with the fixed effects held at their full-sample projection. The largest change is 0.29 per 1,000, about half of the point estimate (0.59) and under one fifth of its standard error (1.53). Archived output of the primary estimation.

Table S8. Class-Level Sensitivity Exercise: Gain and Loss Limbs With Direct Difference Tests (Two-Way Clustered Standard Errors).

Specification	PosRet[−1] (*SE*)	NegRet[−1] (*SE*)	Diff (neg−pos)	*t* (diff)	95% interval (diff)	Funds	*N*
Own returns, ticker FE, no day FE	0.0836 (0.0180)	0.0947 (0.0200)	0.0111	0.49	[−0.036, 0.058]	25	11,089
+ day fixed effects	0.0778 (0.0238)	0.1210 (0.0391)	0.0432	0.90	[−0.056, 0.143]	25	11,089
Benchmark-proxy returns + day FE	0.0720 (0.0527)	0.1817 (0.0771)	0.1097	2.34	[0.013, 0.207]	25	11,089
Inverse funds, own returns, no day FE	0.1259 (0.0687)	0.1586 (0.0632)	0.0327	0.37	[−0.149, 0.214]	26	9,919

Note. Coefficients relate holder-count growth to lagged return limbs in a class-level specification modeled on Bejar-Garcia (2026); they are not in per-1,000 units. Two-way clustered (day, fund) standard errors. The interval for the difference uses the archived two-way clustered *SE* of the difference and the *t* reference with (funds − 1) degrees of freedom; the difference is positive in all four rows, and only the third interval excludes zero. Sample reconciliation for the long-leveraged rungs: the family map contains 33 long +2x/+3x candidates; one

volatility ETP (TVIX) and one fund without a retrievable price series (FNGU) are excluded up front, and six funds (SAA, UBT, UST, TYD, XPP, EET) never reach the specification's 100-holder prior-base floor (maximum holder bases 30–74), contributing zero observations and leaving 25 estimated funds. The inverse-fund rung estimates 26 of 34 mapped candidates. The roster, filters, and conditioning variables all differ from the prior study's, so the rows document sensitivity on this sample and do not test that study's result.

Table S9. Holder-Base Conditioning and Tracking-Deviation Limbs, Primary and 16-Segment Samples (Exploratory; Added After the Primary and Secondary Results).

Specification	Δ_3(2%) / coefficient	*SE* (block)	*t* (block)	MC *p* (calibrated, HC1)
Primary sample (N = 13,242)				
Prespecified model	0.59	1.65	0.36	.726
+ log holder base × lagged limbs: dose contrast at the mean base	2.81	1.45	1.94	.134
Holder-base contrast (per SD of log Ht−1)	2.92	1.86	1.57	.162
+ within-family holder share × lagged limbs: dose contrast	6.53	1.61	4.05	
Share contrast (per SD of share)	3.56	0.99	3.61	
+ tracking-deviation limbs: dose contrast (N = 13,242)	0.88	1.73	0.51	
Signed deviation coefficient (per unit)	−0.122	0.085	−1.44	
Absolute deviation coefficient (per unit)	0.084	0.174	0.48	
All-valid-segments sample (N = 22,466)				
Prespecified model	5.46	1.20	4.54	< .001
+ log holder base × lagged limbs: dose contrast at the mean base	6.51	1.12	5.80	< .001
Holder-base contrast (per SD of log Ht−1)	2.78	1.54	1.80	.018
+ within-family holder share × lagged limbs: dose contrast	7.27	1.18	6.16	
Share contrast (per SD of share)	1.99	0.74	2.69	
+ tracking-deviation limbs: dose contrast (N = 22,466)	5.55	1.23	4.51	
Signed deviation coefficient (per unit)	−0.105	0.056	−1.88	
Absolute deviation coefficient (per unit)	0.022	0.121	0.18	

Note. Contrast rows are in additions per 1,000 prior holders; the tracking-deviation coefficients are per unit of deviation. Standard errors cluster on five-session blocks. The holder-base rows add the standardized log prior holder base (standard deviation 2.33 in the primary sample and 2.12 in the 16-segment sample) interacted with the four lagged limbs; the contrast is then evaluated at the mean log holder base, and the holder-base contrast is the decline-versus-gain difference at 2% per one standard deviation of log holder base. The share rows use the fund's within-family-session share of holders instead (standard deviation 0.34 and 0.34). MC *p* values, where reported, come from the calibrated block-resampling null applied to the HC1 statistic (2,000 draws); the prespecified-model rows repeat Tables 4, 5, and A2. The tracking-deviation rows add the lagged own-fund return minus the leverage multiple times the lagged benchmark-proxy return, signed and in absolute value, computed from each fund's own price series (standard deviation of the deviation 0.0033 and 0.0035). Apart from the repeated prespecified-model rows, all rows are exploratory reparameterizations that change the estimand; none is a mediation test, and none measures prominence.

Table S10. Lead-Return Variants (Exploratory; Calibrated Inference).

Quantity	Estimate	*t*	crit	MC *p*	95% interval	*N*
Primary sample						
Session *t+1* limbs substituted for the lagged limbs	2.33	1.74	2.13	.115	[−0.53, 5.19]	13,242
Session *t+1* limbs added to the prespecified model	2.85	2.13	2.12	.048	[0.02, 5.67]	13,242
Prespecified contrast given the *t+1* limbs	0.30	0.20	2.21	.851	[−3.00, 3.60]	13,242
Session *t+1* limbs added, with *t−2* and *t−3* limbs	2.23	1.68	2.06	.123	[−0.51, 4.97]	13,242
Session *t+2* limbs substituted	1.23	0.92	2.20	.392	[−1.71, 4.17]	13,218
16-segment sample						
Session *t+1* limbs substituted	1.24	1.33	2.57	.303	[−1.15, 3.63]	22,466
Session *t+1* limbs added	1.35	1.48	2.57	.272	[−1.00, 3.71]	22,466
Prespecified contrast given the *t+1* limbs	5.49	5.26	2.62	< .001	[2.75, 8.23]	22,466

Quantity	Estimate	*t*	crit	MC *p*	95% interval	*N*
Session *t+2* limbs substituted	1.54	1.68	2.54	.187	[−0.79, 3.87]	22,427

Note. Estimates are in additions per 1,000 prior holders. Calibrated rows use the prespecified stratified block-resampling algorithm (2,000 draws) applied to each quantity's own estimand and sample; MC *p* is the plus-one Monte Carlo *p* value, and the interval, built from the archived outputs, is the estimate plus or minus the quantity's own critical value times its HC1 *SE*; it inherits the joint-null assumption of Section 4.3 and is not adjusted across rows. Only the first row is the prespecified lead falsification. Substituted lead limbs replace the lagged limbs with the session *t+1* (or *t+2*) benchmark-proxy limbs while retaining the contemporaneous limbs; added lead limbs enter alongside the lagged limbs, and the prespecified contrast is then re-evaluated given them. In the primary sample, the lead contrast with the *t+1* limbs added to the prespecified model (2.85; $p = .048$) is a boundary result in a battery of this size, and it is 2.23 ($p = .123$) once *t−2* and *t−3* limbs are added; because volatility clusters, a response that persists over several sessions would load partly on the leads. In the 16-segment sample the lead contrast is 1.2 to 1.5 under every variant ($p \geq .18$). The prespecified contrast re-evaluated given the *t+1* limbs is 0.30 ($p = .851$) in the primary sample and 5.49 ($p < .001$) in the 16-segment sample.

Table S11. Interior-Only Intervals, S6 Components, and the Roster-by-Window Grid (Exploratory; Added After the Primary and Secondary Results).

Quantity	*N*	Estimate	*SE* (HC1)	*t*	crit	MC *p*	Interval	*SE* (block)	Bootstrap *p*
Panel A. Interior-only and spanning components of the primary and S6 contrasts, under the prespecified interval construction									
Primary sample, interior-only additions	13,242	−0.70	1.32	−0.53	2.07	.615	[−3.44, 2.04]	1.35	.620
Primary sample, spanning additions	13,242	1.29	0.60	2.16	2.27	.064	[−0.07, 2.64]	0.77	.116
16-segment sample, interior-only additions	22,466	2.46	0.93	2.65	2.49	.036	[0.15, 4.77]	0.95	.013
16-segment sample, spanning additions	22,466	2.99	0.38	7.86	2.45	< .001	[2.06, 3.93]	0.53	< .001
Panel B. Loss-side and gain-side components of the S6 contrast									
$\theta_{L3}(2\%)$, loss-side component	22,466	6.95	0.94	7.41	2.63	< .001	[4.48, 9.41]	1.20	—
$\theta_{C3}(2\%)$, gain-side component	22,466	1.49	0.86	1.73	2.69	.206	[−0.82, 3.80]	0.95	—
$\Delta_3(2\%)$, same draws	22,466	5.46	1.05	5.22	2.40	< .001	[2.95, 7.97]	1.20	—
Panel C. Predefined roster-by-window grid for the 16-segment contrast									
Prespecified outcome, full roster	22,466	5.46	—	—	—	—	[3.07, 7.84] (block)	1.20	< .001
Prespecified outcome, without the biotechnology and semiconductor pairs	20,258	2.26	—	—	—	—	[−0.57, 5.08] (block)	1.42	.131
Prespecified outcome, without the four added +3x segments	17,658	1.28	—	—	—	—	[−1.69, 4.26] (block)	1.50	.402
Interior-only outcome, full roster	22,466	2.46	—	—	—	—	[0.58, 4.34] (block)	0.95	.011
Interior-only outcome, without the biotechnology and semiconductor pairs	20,258	0.49	—	—	—	—	[−1.92, 2.90] (block)	1.22	.700
Interior-only outcome, without the four added +3x segments	17,658	−0.15	—	—	—	—	[−2.71, 2.40] (block)	1.29	.915

Note. Estimates, standard errors, and intervals are in additions per 1,000 prior holders. All rows were added after the primary and secondary results were known, are exploratory, and were run under a reporting rule fixed before they were computed: every cell is reported as shown. Panel A holds the prespecified sample, regressors, fixed effects, weights, and contrast fixed and changes only the outcome numerator, to changes observed inside the session window (interior) or to the single inter-poll change that spans the prior scheduled close (spanning); the two coefficients sum to the prespecified coefficient in each sample, exactly before rounding. Panel A and B intervals use the prespecified construction (2,000 stratified five-session block resamples, HC1 *t*, the quantity's own critical value) and inherit its joint-null assumption (Section 4.3). Bootstrap *p* in Panel A is the archived contrast-specific wild-bootstrap value of Table A2. Panel B scores the two halves of the S6 contrast on one set of draws (seed 2026092005), so the Δ_3 critical value (2.40) differs in the second decimal from the archived 2.41; each component is a modeled +3x-minus-+1x response relative to a zero-return session, and the simultaneous critical value across the three contrasts is 3.08. Panel C applies one convention to all six cells: five-session block-clustered standard errors and the contrast-restricted wild bootstrap with 9,999 draws under new seeds, so the full-roster interior cell ($p = .011$) differs in the third decimal from the archived value in Table A2 (.013). Roster deletions refit the model and change the weights. The Panel C intervals are block-clustered *t* intervals (111 blocks; Table S15), not calibrated intervals and not inversions of the bootstrap test. The four point estimates in the first four rows of Panel C reproduce archived values; the two interior-only cells with deletions were computed for this table.

Table S12. Stamp Audit and Boundary-Spanning Shares of Measured Additions (Descriptive; Added After the Primary and Secondary Results).

Panel A. Stamp timing around the prior scheduled close	Fund-sessions	Last stamp before close (min)	Next stamp after close (min)	Spanning interval (min)	Last stamp at :40–:50
16-segment estimation sample (S6): the retained fund-sessions					
All retained fund-sessions	22,466	14.5 (8.8, 17.7)	45.6 (42.7, 51.4)	60.0 (59.5, 60.7)	94%
Regular-close boundary	22,220	14.5 (8.8, 17.7)	45.6 (42.7, 51.4)	60.0 (59.5, 60.8)	94%
of which 2018	6,636	15.6 (13.7, 198.0)	44.5 (42.5, 46.7)	60.0 (59.8, 389.1)	99%
of which 2019	9,922	14.6 (12.0, 16.7)	45.4 (43.4, 48.9)	60.0 (59.8, 60.2)	99%
of which 2020	5,662	11.1 (6.8, 15.5)	48.8 (44.7, 52.6)	60.0 (58.0, 62.6)	80%
Early-close boundary	246	14.7 (9.8, 17.1)	45.3 (42.8, 50.1)	60.0 (59.7, 60.0)	96%
Primary estimation sample (a subset of S6)					
All retained fund-sessions	13,242	14.2 (8.5, 17.4)	45.8 (43.0, 51.7)	60.0 (59.5, 60.8)	93%
Broader archive-roster universe (41 funds; three eligibility checks only)					
Regular-close boundary	22,548	14.5 (8.8, 17.8)	45.6 (42.7, 51.5)	60.0 (59.5, 60.8)	94%
Early-close boundary	246	14.7 (9.8, 17.1)	45.3 (42.8, 50.1)	60.0 (59.7, 60.0)	96%

Note. Medians, with 5th and 95th percentiles in parentheses, from the raw Robintrack ticker files. Each row of a panel is aligned with the boundary that opens its outcome window: for the fund-session of session t, the last stamp at or before the scheduled close of session $t - 1$, the next stamp after it, and the single inter-poll interval that spans that close. Regular sessions close at 4:00 p.m. Eastern; the six early-close sessions close at 1:00 p.m. "Last stamp at :40–:50" is the share of last pre-close stamps whose minute of the hour lies between 40 and 50. The year rows are regular-close boundaries only, whereas Panel B counts all retained fund-sessions of a year, so the two panels' yearly counts differ by the early-close boundaries; the large 95th percentiles for 2018 are the long intervals that Panel B describes. The first two blocks use the exact retained keys written by the sample builder in the plan package (22,466 and 13,242 fund-sessions), and every retained key has a timing record. The third block is a broader universe that applies only the consecutive-session, outage-adjacency, and first-observation checks; it contains the 16-segment sample plus 328 fund-sessions that the full eligibility rules remove (164 in the four platform-outage-related sessions, 84 in family-sessions that are incomplete because another fund of the segment fails a rule, and 80 outside a segment's effective window). The medians are the same to the first decimal in the two universes. The delay between the underlying Robinhood figure and its retrieval is not observed in these records.

Panel B. Long spanning intervals (longer than 75 minutes)	Fund-sessions	Long	Long (%)	Interval p95 (min)	Interval p99 (min)	Median long interval (min)	Of which before the close (median, min)	All additions on these fund-sessions (% of all measured additions)	Spanning additions on these fund-sessions (% of all spanning additions)
16-segment sample, all years	22,466	924	4.1	60.7	978.4	360	197	4.8	11.4
2018	6,759	415	6.1	360.0	1140.0	960	436	5.9	17.6
2019	10,045	278	2.8	60.2	240.0	120	74	2.9	6.4
2020	5,662	231	4.1	62.6	359.6	121	72	5.3	10.9
Primary sample, all years	13,242	548	4.1	60.8	978.4	360	197	4.7	11.9

Note. A long spanning interval is one longer than 75 minutes, a threshold fixed before these summaries were computed; no observation is excluded because of it. Long intervals are 4.1% of the retained 16-segment fund-sessions and fall in 32 sessions, in 20 of which every fund has one; that synchronization is consistent with common collection gaps. Their median length is 360 minutes, with a median of 197 minutes before the close, and in 81% of them most of the interval lies before the close. For these fund-sessions the hourly characterization of Panel A does not apply: the spanning change covers part of the preceding session's trading hours. The last two columns differ in both numerator and denominator. All measured additions on the fund-sessions with a long spanning interval, interior additions included, are 4.8% of all measured additions; the spanning additions on those fund-sessions are 11.4% of all spanning additions. The interior-only numerator excludes every spanning increment, irrespective of its length, but the outcome keeps the same prior holder-count denominator and remains subject to the archive's polling schedule and reporting latency; it is therefore an alternative measured outcome, not a timing-corrected measure of participation. Neither panel shows when a change occurred within an interval.

Panel C. Segment (16-segment sample)	Leverage target	Cell-sessions	Measured additions	Interior	Spanning	Spanning share (%)
Nasdaq-100	+1x	552	48,383	43,838	4,545	9.4
	+2x	552	3,516	3,102	414	11.8
	+3x	552	72,052	63,339	8,713	12.1

Panel C. Segment (16-segment sample)	Leverage target	Cell-sessions	Measured additions	Interior	Spanning	Spanning share (%)
S&P 500	+1x	552	115,858	107,223	8,635	7.5
	+2x	552	4,291	3,773	518	12.1
	+3x	552	17,599	15,583	2,016	11.5
Dow Jones Industrial Average	+1x	552	24,574	22,114	2,460	10.0
	+2x	552	2,465	2,121	344	14.0
	+3x	552	15,658	13,615	2,043	13.0
S&P MidCap 400	+1x	552	2,045	1,840	205	10.0
	+2x	552	321	287	34	10.6
	+3x	552	467	403	64	13.7
ICE 20+ Yr Treasury	+1x	552	10,044	8,750	1,294	12.9
	+2x	552	299	262	37	12.4
	+3x	552	3,504	2,943	561	16.0
ICE 7–10 Yr Treasury	+1x	552	2,948	2,602	346	11.7
	+2x	552	174	158	16	9.2
	+3x	552	354	288	66	18.6
FTSE China 50	+1x	550	4,651	4,120	531	11.4
	+2x	550	183	152	31	16.9
	+3x	550	7,416	6,294	1,122	15.1
MSCI Emerging Markets	+1x	552	4,905	4,340	565	11.5
	+2x	552	84	68	16	19.0
	+3x	552	1,633	1,371	262	16.0
Russell 2000	+1x	552	8,080	7,155	925	11.4
	+2x	552	583	508	75	12.9
	+3x	552	14,698	13,263	1,435	9.8
PHLX Semiconductor	+1x	552	3,102	2,742	360	11.6
	+3x	552	26,665	22,811	3,854	14.5
S&P Biotechnology Select Industry	+1x	552	6,993	6,048	945	13.5
	+3x	552	21,028	18,254	2,774	13.2
Technology Select Sector	+1x	472	11,439	10,452	987	8.6
	+3x	472	7,982	6,963	1,019	12.8
Nasdaq Biotechnology	+1x	552	5,364	4,818	546	10.2
	+2x	552	1,265	1,102	163	12.9
Dow Jones U.S. Financials	+1x	552	443	380	63	14.2
	+2x	552	548	484	64	11.7
Dow Jones U.S. Technology	+1x	552	1,222	1,058	164	13.4
	+2x	552	1,221	1,072	149	12.2
S&P SmallCap 600	+1x	552	3,676	3,306	370	10.1
	+2x	552	138	130	8	5.8

Note. Spanning share is the percentage of measured additions in the single inter-poll change that spans the prior scheduled close, by cell of the 16-segment estimation sample. Shares range from 5.8% to 19.0%, and the pooled share across the +3x cells is 12.7% (23,929 of 189,056). These are shares of measured additions; the 54.9% in Section 5.2 is an arithmetic share of a regression coefficient and is a different quantity. The primary-sample rows reproduce the Boundary % column of Table A1.

Table S13. Full Estimand Crosswalk, Keyed to the Identifiers of Table 3.

ID	Quantity	Provenance	Evidential role	What changes (relative to E1)	Estimate	*SE*	Test statistic and *p*	Interval	Reported in
Primary analysis and its mandatory outputs (8 ladders; 13,242 cell-sessions; equal-family weights; level rate)									
E1	Δ_3(2%), primary contrast	Primary	Tests H-A	—	0.59	1.53	HC1 t = 0.39 vs. calibrated 2.10; MC p = .726	[−2.62, 3.80]	Table 4
E1a	θ_{L3}(2%), loss-side component	Mandatory output; critical value and interval produced from the original calibration draws	H-L gate; descriptive	Component	2.66	1.51	t = 1.76 vs. 2.24; p = .127	[−0.73, 6.06]	Table 4
E1b	θ_{G3}(2%), gain-side component	As E1a	Descriptive	Component	2.07	1.23	t = 1.69 vs. 2.21; p = .131	[−0.64, 4.78]	Table 4
E1c	Δ_2(2%), H-M reporting	As E1a	Descriptive	Leverage target (+2x)	−1.79	1.64	t = −1.09 vs. 2.40; p = .384	[−5.73, 2.15]	Table 4
E1d	θ_{L2}(2%); θ_{G2}(2%)	As E1a	Descriptive	Component (+2x)	−1.45; 0.34	1.58; 1.28	t = −0.92 vs. 2.43 (p = .443); t = 0.26 vs. 2.35 (p = .821)	[−5.29, 2.38]; [−2.68, 3.35]	Section 5.1
E1e	Δ_3(5%)	Exploratory, prespecified; critical value and interval produced from the original calibration draws	Descriptive	Evaluation point	−4.29	3.68	t = −1.17 vs. 2.23; MC p = .294	[−12.49, 3.92]; Treasury support absent	Table 4
E1f	Family-sign cross-check (256 assignments)	Prespecified cross-check	Same estimand	Inference only	0.59	—	CR1 t = 0.33; p = .758	—	Table 4
E2	Same contrast, alternative inference	Exploratory, added after the primary and secondary results	Same estimand; not the decision on H-A	Inference only	0.59	1.65 (block)	Focal-null bootstrap p = .719	HAC [−2.51, 3.69] (no null imposed); block [−2.68, 3.86]	Tables 4, A2, and S4
Secondary family (seven tests; Holm adjustment)									
E3a	S1: Δ_3(1%)	Secondary	Does not test H-A	Evaluation point	0.68	0.86	t = 0.79 vs. 2.19; MC p = .478; Holm p > .999	[−1.22, 2.58]	Table 5
E3b	S2: $\Delta_3(\sigma_I)$, family-volatility-scaled point	Secondary	Does not test H-A	Evaluation point	0.66	1.13	t = 0.58 vs. 2.13; MC p = .595; Holm p > .999	[−1.75, 3.06]	Table 5
E3c	S5: adds the Russell 2000 ladder (N = 14,898)	Secondary	Does not test H-A	Sample	1.09	1.46	t = 0.74 vs. 2.17; MC p = .496; Holm p > .999	[−2.09, 4.26]	Table 5
E3d	S7: common hourly grid (N = 13,248)	Secondary	Does not test H-A	Polling construction; sample (six additional cell-sessions)	0.60	1.53	t = 0.39 vs. 2.19; MC p = .731; Holm p > .999	[−2.74, 3.94]	Table 5
E4	S3: holder-count removal rate, one-sided	Secondary	Does not test H-A	Outcome	−2.70	1.20	t = −2.25 vs. −1.88; MC p = .025; Holm p = .150	Upper bound −0.44	Table 5
E4a	S4: net log holder-count change (log points × 1,000)	Secondary	Does not test H-A	Outcome	3.65	1.82	t = 2.01 vs. 2.17; MC p = .071; Holm p = .355	[−0.29, 7.59]	Table 5

ID	Quantity	Provenance	Evidential role	What changes (relative to E1)	Estimate	*SE*	Test statistic and *p*	Interval	Reported in
E5	S6: all valid segments (16 segments; N = 22,466)	Secondary	Does not test H-A; familywise rejection of its own null	Sample	5.46	1.05	t = 5.22 vs. 2.41; MC p < .001; Holm p = .003; bootstrap p < .001	[2.93, 7.98]; HAC [2.96, 7.96]; 10,000 draws [2.79, 8.12]	Table 5; Tables A2 and S3
Outcome window, roster, and components (16-segment sample unless stated)									
E6a	S6, interior-only additions	Exploratory, added after the primary and secondary results; interval produced under the prespecified construction	Descriptive	Sample; outcome window	2.46	0.93; 0.95 (block)	t = 2.65 vs. 2.49; MC p = .036; bootstrap p = .013	[0.15, 4.77]	Table 5 (memo); Tables A2 and S11
E6b	S6, spanning additions	As E6a	Descriptive	Sample; outcome window	2.99	0.38; 0.53 (block)	t = 7.86 vs. 2.45; MC p < .001; bootstrap p < .001	[2.06, 3.93]	Tables A2 and S11
E6c	Primary, interior-only additions	As E6a	Descriptive	Outcome window	−0.70	1.32; 1.35 (block)	t = −0.53 vs. 2.07; MC p = .615; bootstrap p = .620	[−3.44, 2.04]	Section 5.1; Tables A2 and S11
E6d	Primary, spanning additions	As E6a	Descriptive	Outcome window	1.29	0.60; 0.77 (block)	t = 2.16 vs. 2.27; MC p = .064; bootstrap p = .116	[−0.07, 2.64]	Tables A2 and S11
E7a	S6 without the four added +3x segments (N = 17,658)	Exploratory, added after the primary and secondary results	Descriptive	Sample (S6 less four segments)	1.28	1.40; 1.50 (block)	t = 0.92 vs. 2.34; MC p = .439	[−1.99, 4.56]	Table 6
E7b	S6 without the biotechnology and semiconductor pairs (N = 20,258)	Exploratory, added after the primary and secondary results	Descriptive	Sample (S6 less two pairs)	2.26	1.29; 1.42 (block)	t = 1.76 vs. 2.39; MC p = .156	[−0.81, 5.32]	Table 6
E7c	Interior-only, without the two pairs; without the four added +3x segments	Exploratory, added after the primary and secondary results (predefined grid)	Descriptive	Sample and outcome window	0.49; −0.15	1.22; 1.29 (block)	Bootstrap p = .700; .915	Block [−1.92, 2.90]; [−2.71, 2.40]	Section 5.2; Table S11
E8	S6 loss-side; gain-side component	Exploratory, added after the primary and secondary results	Descriptive	Sample; one side of the contrast	6.95; 1.49	0.94; 0.86	t = 7.41 vs. 2.63 (p < .001); t = 1.73 vs. 2.69 (p = .206)	[4.48, 9.41]; [−0.82, 3.80]	Section 5.2; Table S11
Weights, scale, and the leverage-target restriction (primary sample unless stated)									
E9	Holder-base weights	Exploratory, added after the primary and secondary results	Descriptive	Weights	4.75	1.41 (block)	t = 3.36	Block [1.94, 7.55]	Table S4
E9a	Equal-observation weights	Exploratory, added after the primary and secondary results	Descriptive	Weights	0.59	1.65 (block)	t = 0.36	Block [−2.68, 3.86]	Table S4
E9b	Inverse-variance weights	Exploratory, added after the primary and secondary results	Descriptive	Weights (estimated)	4.04	1.11 (block)	t = 3.65	Block [1.84, 6.23]	Table S4

ID	Quantity	Provenance	Evidential role	What changes (relative to E1)	Estimate	*SE*	Test statistic and *p*	Interval	Reported in
E9c	Level model on the PPML-retained sample (N = 12,963)	Exploratory, added after the primary and secondary results	Descriptive	Sample (279 separated observations dropped)	0.56	1.71 (block)	t = 0.33	Block [−2.83, 3.94]	Table S4
E10	PPML (N = 12,963)	Mandatory robustness	Descriptive	Scale, implied weights, and sample (279 separated observations dropped)	0.378 log points	0.106 (block)	t = 3.55	Block [0.167, 0.589] log points	Section 5.3; Table S4
E11	Continuous-dose implied Δ_3: primary; S6 (exploratory extension)	Mandatory robustness; block inference added after the primary and secondary results	Descriptive	Leverage-target restriction (both); sample (S6 row)	1.25; 6.05	1.76; 1.24 (block)	Bootstrap p = .507; < .001	HAC [−1.62, 4.12]; [3.62, 8.48]	Section 5.3; Tables A2 and S16
E11a	$\Delta_3 - 2\Delta_2$: primary; S6	Exploratory, added after the primary and secondary results	Descriptive	Restriction test (both); sample (S6 row)	4.17; 4.22	3.81; 3.01 (block)	Bootstrap p = .300; .179	HAC [−1.48, 9.81]; [−0.86, 9.29]	Section 5.3; Tables A2 and S16
E12	Equity ladders only (N = 9,930)	Exploratory, added after the primary and secondary results	Descriptive	Sample	4.00	1.52	t = 2.63 vs. 2.18; MC p = .021	[0.68, 7.33]	Section 5.1; Table S6
Conditioning sets, unlevered funds, group differences, timing, and sample variants									
E13	S6 with own-return limbs (N = 22,404); baseline on that subsample	Exploratory, prespecified	Descriptive	Sample (S6 channel subsample); conditioning set	0.07; 5.47	5.39; 1.21 (block)	t = 0.01; 4.54	Block [−10.60, 10.75]; [3.08, 7.86]	Table 7
E13a	S6 with abnormal volume; with overnight-gap limbs	Exploratory, prespecified	Descriptive	Sample (S6 channel subsample); conditioning set	5.40; 7.89	1.21; 1.44 (block)	t = 4.48; 5.49	Block [3.01, 7.79]; [5.04, 10.74]	Table 7
E13b	Holder-base conditioning: primary; S6	Exploratory, added after the primary and secondary results	Descriptive	Conditioning set (both); sample (S6 row)	2.81; 6.51	1.45; 1.12 (block)	MC p = .134; < .001	—	Section 5.4; Table S9
E13c	Tracking-deviation limbs: primary; S6	Exploratory, added after the primary and secondary results	Descriptive	Conditioning set (both); sample (S6 row)	0.88; 5.55	1.73; 1.23 (block)	t = 0.51; 4.51	—	Section 5.4; Table S9

ID	Quantity	Provenance	Evidential role	What changes (relative to E1)	Estimate	*SE*	Test statistic and *p*	Interval	Reported in
E14	Unlevered SPY − VOO; IVV − VOO; SPY − IVV (*N* = 1,542)	Exploratory, prespecified (corrected specification)	Descriptive	Sample (three unlevered funds); no leverage difference; fund and session fixed effects	1.48; −0.09; 1.57	0.36; 0.34; 0.45 (block)	Bootstrap $p < .001$ ($t = 4.06$ vs. bootstrap critical value 2.14); .801; .003	HAC [0.73, 2.23]; [−0.76, 0.57]; [0.49, 2.66]	Section 5.4; Tables A2 and S16
E15	Equity average − Treasury average (primary)	Exploratory, added after the primary and secondary results	Descriptive	Group difference	19.4	5.23 (block)	HC1 $t = 2.77$ vs. 3.55; MC $p = .113$; bootstrap $p = .050$	HAC [8.5, 30.4]	Section 5.3; Tables A2 and S16
E15a	Added +3x segments − original eight (S6)	Exploratory, added after the primary and secondary results	Descriptive	Sample (S6); group difference	9.9	1.38 (block)	HC1 $t = 5.81$ vs. 3.51; MC $p = .006$; bootstrap $p < .001$	HAC [7.0, 12.8]	Section 5.2; Tables A2 and S16
E15b	Sector pairs − broad ladders (S6)	Exploratory, added after the primary and secondary results	Descriptive	Sample (S6); group difference	5.4	1.43 (block)	HC1 $t = 5.33$ vs. 2.42; MC $p < .001$; bootstrap $p = .001$	HAC [2.8, 8.1]	Section 5.2; Tables A2 and S16
E16	Session *t+1* lead-return contrast (primary)	Exploratory, prespecified	Descriptive	Regressor timing	2.33	1.34	$t = 1.74$ vs. 2.13; MC $p = .115$	[−0.53, 5.19]	Section 5.4; Table S10
E17	S6 by period: through 2020-02-19; COVID-19 period; difference	Exploratory, prespecified; difference test added after the primary and secondary results	Descriptive	Sample (S6, by period)	5.99; 4.21; −1.78	—; —; 2.73 (block)	MC $p < .001$; .224; bootstrap $p = .534$	[2.99, 8.98]; [−2.24, 10.66]; HAC [−7.51, 3.96]	Table 6; Tables A2 and S16
E18	Return-tail exclusion: primary (*N* = 13,098); S6 (*N* = 22,222)	Exploratory, added after the primary and secondary results	Descriptive	Sample	−0.07; 5.24	1.57; 1.24 (block)	Bootstrap $p = .967$; $< .001$	HAC [−3.02, 2.89]; [2.67, 7.82]	Section 5.4; Tables A2 and S16

Note. Estimates, standard errors, and intervals are in additions per 1,000 prior holders unless stated. Standard errors are HC1 unless marked (block), which denotes clustering on five-session blocks. "vs." gives the quantity's own calibrated critical value; MC *p* is the plus-one Monte Carlo *p* value of the prespecified calibration applied to that quantity; bootstrap *p* is the contrast-specific wild bootstrap of Section 4.3, which tests the zero contrast and yields no interval. Unmarked intervals are those reported under each quantity's own calibration and rest on the assumptions set out in Section 4.3; "HAC" marks a session-score HAC interval with a normal reference, which imposes no null, and "Block" marks a block-clustered *t* interval (Table S15). All intervals are individual, and none is adjusted across rows. "What changes" lists every dimension that differs from E1. Rows that change only the inference keep that estimand; every other row targets a different quantity, so agreement with the primary result is not confirmation and a difference is not a contradiction. Only E1 supplies the decision on H-A; E1f and E2 apply other inference to the same contrast. For E15 the HAC interval excludes zero, whereas the calibrated test does not reject, the procedure sensitivity noted in Section 5.3. Rows E15 to E18 have no counterpart in Table 3. A source-to-table manifest, archived with the post-estimation scripts and outputs, names the output files behind every table and figure.

Table S14. Ledger of Departures From the Analysis Plan and of Additions Made After the Primary and Secondary Results Were Known.

#	Item	Kind	What was done	Bearing on the prespecified results	Reported in
1	Name of the interval	Reporting	The analysis plan calls the construction (estimate plus or minus the calibrated critical value times the HC1 standard error) a percentile-t interval; the article calls it a symmetric calibrated interval. The computation is unchanged.	None	Section 4.3
2	Title	Reporting	The article's title is a plainer variant of the plan's title, "Benchmark Returns and the Leverage Gradient in Robinhood ETF Holder Additions," of the kind the plan preauthorizes as cosmetic.	None	Section 4.4
3	Unlevered comparison: first implementation withdrawn	Result-informed	The first implementation carried the leveraged-status-by-session effects into a single-family sample; the projected design had rank 8 for 16 columns, so the SPY − VOO contrast was not identified. The corrected specification uses fund and session effects and has full rank. A dated note records the withdrawal beside the archived output; no archived file was altered.	None; exploratory	Section 5.4; Table A2
4	Contrast-specific wild bootstrap and session-score HAC intervals	Result-informed	Two separate computations for the primary contrast, S6, and the other rows of Table A2: a wild-bootstrap test that imposes only the focal restriction, and a session-score HAC interval computed from the unrestricted model, which imposes no null.	Same estimand, different inference; primary and secondary decisions unchanged	Section 4.3; Tables 4 and A2
5	Interior and boundary decomposition; group-difference tests; period-difference test; return-tail check; $\Delta_3 - \Delta_2$ and $\Delta_3 - 2\Delta_2$ tests	Result-informed	Exploratory checks reported with block-clustered standard errors and the contrast-specific bootstrap.	None; exploratory	Table A2
6	Equity-ladders-only contrast	Result-informed	Pooled contrast on the six equity ladders, with calibrated inference.	None; no confirmatory weight	Section 5.1; Table S6
7	Holder-base conditioning and tracking-deviation limbs	Result-informed	Reparameterizations that change the estimand.	None; exploratory	Section 5.4; Table S9
8	Lead-return variants beyond the prespecified *t+1* check	Result-informed	Added leads, deeper lags, and *t+2* leads.	None; exploratory	Section 5.4; Table S10
9	S6 roster deletions and leave-one-segment-out fits	Result-informed	Each deletion refits the model on a different sample.	None; exploratory	Sections 5.2 and 5.4; Table 6
10	Component and Δ_2 critical values and intervals	Prespecified output, produced for display	Prespecified outputs whose contrast-specific critical values were not among the archived outputs. Computed with the prespecified algorithm on the original calibration draws (seed 20260832); the reproduced primary critical value and Monte Carlo *p* value are identical to the archived values.	Primary estimate and decision unchanged	Table 4
11	Intervals for the seven secondary tests (S3 as a one-sided bound); ten most influential observations	Prespecified output, produced for display	Display of archived outputs of the prespecified estimation.	None	Table 5; Table S7
12	Stamp audit and spanning shares of measured additions	Result-informed	Descriptive; no estimation. The audit is restricted to the exact retained fund-session keys of the estimation samples, reconciled with a broader universe of 22,794 fund-sessions, of which 22,466 are retained, and extended with the long-interval tail.	None	Section 3.3; Table S12
13	Interval of the prespecified construction for the interior-only and spanning estimates	Result-informed	The outcome numerator changes; the sample, regressors, fixed effects, weights, and contrast are the prespecified ones.	None; exploratory	Section 5.2; Table 5 (memo row); Table S11
14	S6 loss-side and gain-side components	Result-informed	The two halves of the S6 contrast, scored on one set of draws.	None; descriptive	Section 5.2; Table S11

#	Item	Kind	What was done	Bearing on the prespecified results	Reported in
15	Roster-by-window grid for the 16-segment contrast	Result-informed	Six cells defined before they were computed and all reported, under one inferential convention.	None; exploratory	Section 5.2; Table S11
16	Return support for the added segments; itemized session reconciliation	Result-informed	Descriptive counts under the rule of Table 1, Panel A, and the prespecified sample construction.	None	Table 1; Table S1
17	Figure 3B marker	Reporting	The marker for the 7–10-year Treasury family sits at its estimate of −22.8, and the interval, which extends beyond the plotted range, is marked with arrows.	Presentation only	Figure 3
18	Units	Reporting	Holder-count rates are reported as additions or removals per 1,000 prior holders, as applicable, and quantities on other scales (log points for the net log change and for PPML, and the class-level coefficients) are labeled with their own units. The rescaling from the archived outputs, which are per prior holder, is a change of units only; no estimate, decision, or interval is affected by it.	Presentation only	All tables
19	Labels	Reporting	The sample variants of Table S6 are labeled prespecified exploratory, as the analysis plan lists them; the class-level item that the plan calls a descriptive replication is reported as a sensitivity exercise.	Labeling only	Sections 5.4 and 5.5; Tables S6 and S8
20	Last-digit rounding	Reporting	Seven displayed values are taken from the source outputs rather than carried through an intermediate rounding: the Holm-adjusted p of S6 (.003); the Monte Carlo p of the added-segment group difference (.006), of the fit without the biotechnology pair in Table 6 (.009), and of the China family in Table A3, Panel B (.005); two addition rates in Table A1 (3.5 and 8.8, which agree with Table A4); and one standard error in Note S3 (12.3). No estimate, decision, or interval is affected.	None	Section 5.2; Tables 3, 5, 6, A1, A3, and S13; Note S3
21	Table order	Reporting	Tables are numbered in order of first citation. The table of sample, period, and roster variants of the 16-segment contrast (Table 6) sits in Section 5.2, where it is first cited, and the contrast-specific inference table (Table A2) follows Table A1.	Presentation only	Tables 6 and A2
22	Intervals for prespecified quantities	Prespecified output, produced for display	The analysis plan promises an interval with every prespecified estimate. Every such interval is reported, with one exception: the prespecified family-sign cross-check is implemented as an enumeration test that reports a statistic and a p value, and no interval is constructed for it (Table S16). No prespecified estimate, decision, sample, or seed is altered: the Δ_3(5%) critical value and interval were scored on the original calibration draws; archived calibrated intervals are shown for the lead-return variants and the rows of Table 6; archived HAC intervals are shown for every row of Table A2; block-clustered t intervals are reported in Tables S4 to S6 and S11 (Panel C) and in the note of Table 7, with every refitted row first reproducing its archived estimate; and two-way clustered intervals are reported for the class-level differences of Table S8. No bootstrap test was inverted into an interval, and no critical value was borrowed across quantities.	None; primary and secondary decisions unchanged	Table 4; Tables S4 to S6, S8, S10, S11, S13, S15, and S16
23	Descriptions of the inference procedures	Reporting	Section 4.3, Table 4, Table 3, and the note to Table A2 separate the focal-null bootstrap test from the HAC interval, which imposes no null, and state which sets of contrasts carry a simultaneous critical value. Table S15 is the inference key.	Wording only	Sections 4.3 and 6.2; Tables 3, 4, A2, S13, and S15

#	Item	Kind	What was done	Bearing on the prespecified results	Reported in
24	Figure 3, Panel A	Reporting	The panel shows the estimated curves, the return support, and the calibrated component intervals of Table 4 at the prespecified 2% evaluation point. Pointwise normal-reference HC1 bands are not drawn, because they are not calibrated for the time dependence in the data. Panel B is titled as showing individual, not simultaneous, intervals.	Presentation only	Figure 3

Note. The prespecified sample, estimator, weights, evaluation point, decision rule, and Holm family are unchanged, and the primary and secondary decisions are reported as the plan specifies. Under the plan's deviation rules, every result-informed addition is labeled exploratory. Scripts and outputs produced after the primary and secondary results were known are archived in dated folders kept apart from the plan package, which, like every file whose hash its manifest records, is unchanged.

Table S15. Inference Key: Statistic, Reference Distribution, Null Imposed, Interval Construction, and Multiplicity Family of Each Procedure.

Procedure	Statistic and *SE*	Reference distribution	Null imposed	Interval	Multiplicity	Where used
Prespecified calibration (decision rule for H-A)	HC1 *t* of the contrast	95th percentile of |*t*| over 2,000 stratified five-session moving-block resamples of the fixed-effect-projected outcome (archived seed)	Joint: every return slope is zero, which is stronger than the focal contrast	Estimate ± own critical value × HC1 *SE*; nominal coverage for the focal contrast is assumed, not demonstrated (Section 4.3)	None: one primary test	Table 4, rows 1 to 5; only row 1 decides H-A
Secondaries S1 to S7	As above, each on its own sample and outcome; S3 one-sided	Own calibration, as above	As above	Individual intervals (S3: a one-sided bound)	Holm across the seven decisions; the intervals are not Holm-adjusted	Table 5
Own calibration of the primary-sample family estimates (a mandatory output) and of exploratory quantities: segments, group differences, lead-return variants, 16-segment sample variants, interior and spanning components, S6 components, equity ladders only, holder-base conditioning	HC1 *t*	Own calibrated critical value (2,000 draws)	Joint, as above	Estimate ± own critical value × HC1 *SE*; individual	A simultaneous max-|*t*| critical value for three sets only: eight family contrasts (4.21), 16 segment contrasts (4.62), three S6 component contrasts (3.08). Otherwise none	Tables 6, A3, S6 (memo row), S9, S10, S11, and S16
Prespecified family-sign cross-check	CR1 *t*	Exact enumeration of the 256 family-level sign assignments	Null imposed	None: a test only	None	Table 4
Archived block-wild residual diagnostic	HC1 *t*; simulated outcomes are not reprojected through the fixed effects	Rademacher signs on contrast-restricted residuals in five-session blocks	Focal restriction only; nuisance slopes retained	None: a critical value only	None	Table S3
Contrast-specific wild bootstrap	*t* with the five-session-block clustered *SE*	9,999 Rademacher block draws; all fixed effects and slopes refit in each draw	Focal restriction only (the displayed zero contrast); nuisance slopes retained	None: a test of the zero contrast, not inverted into an interval	None; *p* values are unadjusted across the battery	Table 4 (last row); Tables A2 and S11

Procedure	Statistic and *SE*	Reference distribution	Null imposed	Interval	Multiplicity	Where used
Session-score HAC	Estimate divided by the HAC *SE* (Bartlett weights, ten lags, session-summed scores of the unrestricted fitted model)	Normal	None	Estimate ± 1.96 × HAC *SE*; asymptotic	None	Tables 3, 4, A2, S13, and S16
Block-clustered analytic	Estimate divided by the five-session-block clustered *SE*	*t* with *G* − 1 degrees of freedom, *G* the number of blocks in the fit	None	Estimate ± *t* quantile × block *SE*; individual	None	Table 7 (note); Tables 3, S4 to S6, S9, S11 (Panel C), and S13
Two-way clustered (class-level exercise)	Coefficient divided by the two-way (day, fund) clustered *SE*	*t* with (funds − 1) degrees of freedom for the interval	None	Difference ± *t* quantile × *SE*	None	Table S8

Note. The decision on H-A comes from the first procedure alone. The three procedures beneath the cross-check are of different kinds and are kept distinct: the block-wild diagnostic and the contrast-specific bootstrap impose the focal restriction and yield a critical value and a test, and the HAC computation imposes no null and yields an interval. Agreement among them for the primary contrast and S6 is a limited sensitivity check and does not establish nominal coverage for any of them. No interval in this article or these materials is an inverted bootstrap confidence set, and no critical value is borrowed from one quantity for another. Apart from the Holm adjustment of the seven secondary decisions and the three simultaneous critical values listed, nothing is adjusted across the exploratory battery.

Table S16. Interval Inventory for the Prespecified Analyses, Keyed to the Analysis Plan, With the Archived Intervals for Tables A2 and 6.

Panel A. Plan item	Quantity	Interval	Construction (Table S15)	Where the estimate is reported
C4	$\Delta_3(2\%)$, primary	[−2.62, 3.80]	Prespecified calibration	Table 4
C1, C3	Mandatory components $\theta_{L3}(2\%)$; $\theta_{C3}(2\%)$	[−0.73, 6.06]; [−0.64, 4.78]	Prespecified calibration; produced from the original draws	Table 4
C1 (H-M)	$\Delta_2(2\%)$; $\theta_{L2}(2\%)$; $\theta_{C2}(2\%)$	[−5.73, 2.15]; [−5.29, 2.38]; [−2.68, 3.35]	As above	Table 4; Table S13 (E1c, E1d)
C3	Family-specific $\Delta_3(2\%)$ and their unweighted average	One per family, and the unweighted average, [−7.6, 6.6]	Own calibration (critical value 3.55 for the average); simultaneous critical value 4.21 across the eight family contrasts	Table A3
C3	PPML (mandatory robustness), log points	[0.167, 0.589]	Block-clustered *t*	Table S4
C3	Continuous-dose implied Δ_3 (mandatory robustness)	[−1.62, 4.12]	HAC	Table A2; Panel B below
C4	Family-sign cross-check	None	An enumeration test of the zero contrast; the prespecified implementation reports a statistic and a *p* value, and no interval is constructed	Table 4
C6	Secondaries S1 to S7	Seven intervals (S3: upper bound −0.44)	Own calibration; Holm applies to the decisions	Table 5
C7	$\Delta_3(5\%)$	[−12.49, 3.92]	Prespecified calibration on the original draws (critical value 2.23; $p = .294$); fails the plan's return-support standard	Table 4
C7	Continuous-dose variant beyond its mandatory reporting (16-segment sample)	[3.62, 8.48]	HAC	Table A2; Panel B below

Panel A. Plan item	Quantity	Interval	Construction (Table S15)	Where the estimate is reported
C7	Decompositions that change the estimand (own return; abnormal volume; pre-move (overnight) gap): 16-segment sample	[−10.60, 10.75]; [3.01, 7.79]; [5.04, 10.74]	Block-clustered *t*	Table 7 (note)
C7	The same decompositions, primary sample	[−22.45, 9.36]; [−2.71, 3.82]; [−3.15, 7.64]	Block-clustered *t*	Table S5
C7	Fractional-shares split: before; after	[−3.50, 3.94]; [−3.02, 6.51]	Block-clustered *t*	Table S6
C7	Unlevered exercise (corrected specification): SPY − VOO; IVV − VOO; SPY − IVV	[0.73, 2.23]; [−0.76, 0.57]; [0.49, 2.66]	HAC	Table A2; Panel B below
C7	Session *t+1* lead falsification	[−0.53, 5.19]	Own calibration	Table S10
C7	Pre-COVID; five-decline exclusion; outage sessions retained; $H \geq 25$ and 50; family leave-one-out; VIX split	Fifteen intervals, all including zero	Block-clustered *t*	Table S6
C7	Class-level exercise, including the inverse-fund panel	Four difference intervals	Two-way clustered; *t* reference	Table S8
C7, C8	Data-integrity map and mandatory descriptive outputs	Descriptive; no estimate	—	Table 2; Tables A1, S1, S2, and S7

Note. The analysis plan promises a point estimate and an interval for every prespecified test (its Section C9) and open estimates with intervals for its exploratory list (C7). Panel A lists each prespecified item with its interval, the construction behind it, and the table that reports the estimate. Estimates and intervals are in additions per 1,000 prior holders unless stated. No prespecified estimate, decision, sample, or seed is altered. For this inventory, new computation was limited to the $\Delta_3(5\%)$ critical value, scored on the original calibration draws after the primary critical value was reproduced exactly, and to block-clustered standard errors for the rows of Tables S5 and S6; each of those rows first reproduced its archived estimate and HC1 standard error. The block-clustered intervals of Tables S4 and S11 (Panel C) and of Table 7 are built from archived estimates and archived block-clustered standard errors. Every other interval is an archived value. All intervals are individual; none is adjusted across the exploratory battery.

Panel B. Contrast and sample (rows of Table A2)	Estimate	Block *SE*	Bootstrap *p*	HAC 95% interval
Primary $\Delta_3(2\%)$	0.59	1.65	.719	[−2.51, 3.69]
S6 $\Delta_3(2\%)$	5.46	1.20	< .001	[2.96, 7.96]
Primary continuous-dose implied Δ_3	1.25	1.76	.507	[−1.62, 4.12]
S6 continuous-dose implied Δ_3	6.05	1.24	< .001	[3.62, 8.48]
Primary $\Delta_3 - \Delta_2$	2.38	2.05	.281	[−0.34, 5.10]
S6 $\Delta_3 - \Delta_2$	4.84	1.62	.006	[2.22, 7.45]
Primary $\Delta_3 - 2\Delta_2$	4.17	3.81	.300	[−1.48, 9.81]
S6 $\Delta_3 - 2\Delta_2$	4.22	3.01	.179	[−0.86, 9.29]
Primary interior-only additions	−0.70	1.35	.620	[−3.25, 1.85]
S6 interior-only additions	2.46	0.95	.013	[0.50, 4.42]
Unlevered SPY − VOO	1.48	0.36	< .001	[0.73, 2.23]
Unlevered IVV − VOO	−0.09	0.34	.801	[−0.76, 0.57]
Unlevered SPY − IVV	1.57	0.45	.003	[0.49, 2.66]
Primary equity average − Treasury average	19.43	5.23	.050	[8.47, 30.39]
S6 added +3x segments − original eight	9.90	1.38	< .001	[6.97, 12.83]
S6 sector pairs − broad ladders	5.42	1.43	.001	[2.79, 8.05]
Primary boundary additions	1.29	0.77	.116	[−0.34, 2.92]
S6 boundary additions	2.99	0.53	< .001	[1.92, 4.07]

Panel B. Contrast and sample (rows of Table A2)	Estimate	Block *SE*	Bootstrap *p*	HAC 95% interval
Primary COVID − pre-COVID Δ_3	1.57	3.49	.678	[−5.28, 8.42]
S6 COVID − pre-COVID Δ_3	−1.78	2.73	.534	[−7.51, 3.96]
Primary Δ_3, return-tail exclusion	−0.07	1.57	.967	[−3.02, 2.89]
S6 Δ_3, return-tail exclusion	5.24	1.24	< .001	[2.67, 7.82]

Note. Archived session-score HAC intervals (Bartlett weights, ten lags, normal reference) for every row of Table A2. They are computed from the unrestricted fitted model and impose no null; the bootstrap *p* values beside them come from a separate test of the zero contrast and are repeated from Table A2. The intervals are individual and unadjusted. For the equity-minus-Treasury difference the HAC interval excludes zero, whereas the calibrated test does not reject ($p = .113$; Section 5.3), so that comparison remains sensitive to the procedure. The archive also holds HAC intervals for the holder-count removal-rate contrast: −4.96 to −0.43 in the primary sample and −6.52 to −3.20 in the 16-segment sample; they are two-sided and unadjusted, and the prespecified one-sided, Holm-adjusted decision for S3 is unchanged. Definitions for Table A2: period tests split at February 20, 2020, interact all return limbs and cell intercepts with period, and weight families equally within period; tail checks remove whole family-sessions above each family's baseline 99th percentile of absolute lagged return and recompute equal-family weights, and the retained samples span 551 sessions; no return or outcome is winsorized. The fund order of the unlevered exercise (VOO, IVV, SPY) matches the funds' ranking by mean daily dollar volume (SPY $25.3 billion, IVV $1.4 billion, VOO $1.0 billion) but not by Robinhood holder base, where VOO exceeds SPY.

Panel C. Rows of Table 6 (16-segment contrast)	Estimate	crit	MC *p*	Calibrated 95% interval
Secondary test S6 (Table 5)	5.46	2.41	< .001	[2.93, 7.98]
Sessions through 2020-02-19	5.99	2.49	< .001	[2.99, 8.98]
Sessions from 2020-02-20 (COVID-19 period)	4.21	3.49	.224	[−2.24, 10.66]
Excluding the five largest S&P 500 decline sessions	5.46	2.42	< .001	[2.97, 7.95]
Without the biotechnology pair (XBI/LABU)	3.59	2.42	.009	[0.79, 6.39]
Without the semiconductor pair (SOXX/SOXL)	4.69	2.43	< .001	[1.92, 7.46]
Without both pairs	2.26	2.39	.156	[−0.81, 5.32]
Without all four added +3x segments	1.28	2.34	.439	[−1.99, 4.56]

Note. Archived intervals under each row's own calibration (2,000 draws), which inherit the joint-null assumption of Section 4.3. Each row below the first changes the sample, so none is an interval for the S6 estimand.

Table S17. Registration Timeline and Its Sources.

Time (UTC, 2026)	Event	Record	How to check
Aug 31, 16:40–16:41	Manifest and plan package written	Google Drive modification times (16:40:49; 16:41:03)	Times available from the author; package SHA-256 902ea769…c607b
Aug 31, 16:51	Package uploaded to OSF Storage	OSF Storage revision 1; activity log	Registration: Files, Revisions; Recent Activity
Aug 31, 16:52	Registration initiated from the draft	OSF activity log	Recent Activity
Aug 31, 16:52	Draft last modified, as OSF listed it before submission on Sep 23	OSF draft listing, visible only to the author	Not publicly visible; reported by the author
Aug 31, 16:52	OSF notice of technical archiving failure (errors copying linked third-party files)	Email from OSF to the author	Available from the author
Aug 31, 16:54	Project status files record the freeze as in effect and estimation as authorized	Author's file system	Available from the author
Aug 31, 16:56	Output of the first outcome-return estimation (primary analysis) saved to the project folder	Author's file system (16:56:16); Google Drive (created 16:56:17)	Available from the author
Aug 31, 17:13	Output of the secondary tests saved to the project folder	Author's file system	Available from the author
Sep 23, 20:16 and 20:20	Registration attempt initiated, then cancelled by the author	OSF activity log	Recent Activity

Time (UTC, 2026)	Event	Record	How to check
Sep 23, 20:22	Registration submitted after the primary and secondary results were known; date of formal registration	OSF registration record	Registration page
Sep 24, 16:42	View-only link created for the editors and reviewers	OSF activity log	Recent Activity
Oct 5	Embargo ended; registration public	OSF registration record	Registration page

Note. Times are in UTC. OSF may display times in the viewer's time zone; in Pacific Daylight Time, the August 31 entries read 9:51 and 9:52 a.m. Records marked "available from the author" are held on the author's own systems, and the draft listing was visible only to the author; the OSF records can be checked on the registration page (https://osf.io/7kyd5).