Do Published Equity Anomalies Survive Clean Data?

Nine US-equity papers replicated out-of-sample on survivorship-free data: on faithful builds none survive, and the lone apparent survivor does not reproduce when built to the paper

Study	
Author	Vlad Rodeski (publishes as PolyQuant), with review by Radovan Vojtko, CEO and Head of Research at Quantpedia
Date	2026-06-02
Scope	Nine published US-equity strategy papers, replicated and stress-tested on point-in-time data
Data	Core US Equities (daily prices + quarterly fundamentals + point-in-time S&P 500 membership)
History	Prices 1998–2025; delisted names retained, so the universe is free of survivorship bias by construction
Windows	In-sample 2000-01 → 2022-12 (~23y); out-of-sample 2023-01 → 2025-12 (36 months)
Costs	Paper-comparison runs at 5 bps/side; every candidate additionally stress-tested at realistic, turnover-and-liquidity-tiered costs
Benchmark	Long-only strategies vs SPY total-return buy & hold; long/short strategies vs cash, with realized market beta reported

Read this first

We took nine published equity-anomaly papers and asked two separate questions of each:

1. Does the documented effect reproduce on clean, survivorship-free data? The scientific question.
2. If it reproduces, is it worth carrying forward to deeper validation? The practical question.

We keep these apart deliberately. A paper can reproduce its statistics perfectly and still be untradeable; another can fail out-of-sample yet teach a real methodological lesson. Conflating the two is how backtests flatter themselves.

Four takeaways:

- Replication is the exception, not the rule. Of 9 papers, 3 show a statistically significant full-sample alpha ($t > 2$); only one (Bali (MAX / lottery)) clears the stricter $t > 3$ hurdle that the multiple-testing literature now demands; and on faithful, paper-verified builds, none retains a significant alpha out-of-sample. The dominant pattern is **out-of-sample decay**: effects real in-sample have largely faded over a fixed 2023–25 out-of-sample window. We measure decay

against a single common out-of-sample date, not from each paper's own publication year, but it is the same fading the post-publication-decay literature (Hou-Xue-Zhang 2020; McLean-Pontiff 2016) predicts.

- Nothing beat passive in 2023–25. SPY buy-and-hold returned 22.9%/yr (Sharpe 1.43 net of cash), and no long-only strategy beat it on return or risk-adjusted return. The strategies' edge, where it exists, lives in the in-sample years, not the recent regime.
- The lone apparent out-of-sample survivor does not reproduce when built to the paper. Our first build produced one strategy, Zhu (52-week-high reversal) (long/short), with an out-of-sample Sharpe of 1.95. It did not survive a closer look. The paper PDF was inaccessible when we built it; our first build diverged from the paper. When we obtained the paper and rebuilt faithfully (reversal *within* the low-52-week-high quintile, value-weighted, the paper's actual construction), the out-of-sample Sharpe became –0.95. The 1.95 came from a mis-built corner-sort and equal-weighting, not the paper's strategy. On clean data, faithfully built, **none of the nine survive out-of-sample**.
- A clean illustration of look-ahead bias. The Geertsema-Lu (price-level / look-ahead) "low-price anomaly" is worth about three and a half Sharpe points out-of-sample, *entirely* as an artifact of using retroactively split-adjusted prices. On as-traded prices it has no edge. The gap between the two is a textbook demonstration of why point-in-time data matters.

A word on rigour up front: at 36 out-of-sample months, no single result is statistically significant after correcting for the number of strategies tested. Everything below rests on *structural* evidence: independent-engine validation, factor decomposition, mechanism behaviour, and full-sample estimates, not on out-of-sample significance. We say so wherever it matters.

---

1. Executive summary

Each paper carries one verdict: does the documented effect reproduce on clean, survivorship-free data? Reported as the direction it reproduces (or doesn't), with caveats. This is the scientific question; it is not a trading recommendation.

The headline the benchmark column forces: SPY returned 22.9%/yr at a Sharpe net of cash of 1.43 over 2023–25 (the mega-cap-led AI rally). Among the nine, every long-only strategy posted *negative* excess return versus SPY (–9% to –27%/yr), and none matched its Sharpe (see Read-this-first). A faithful replication of a documented effect is not the same as an effect that pays out-of-sample, and over this window, on faithful builds, none of the nine shows a tradeable out-of-sample edge. A per-paper reproduction summary is in §6.

Consolidated scorecard

All returns net of 5 bps/side. Sharpe is annualized arithmetic: long-only in excess of cash, long/short as a zero-cost spread. **α** is the annualized CAPM alpha versus SPY (the part of the return that is *not* market exposure), with its t-statistic. "Overlap" is how much of the paper's original sample our data window covers:

a short overlap means our reproduction is a small-sample read on a sub-period the paper didn't emphasize. The verdict vocabulary is fixed at four values (**Reproduces** / **Reproduces (caveated)** / **Does not reproduce** / **Untestable**); per-paper nuance is in §5.

Table A: Performance & significance

Paper	Type	Reproduces?	IS Sharpe	OOS Sharpe	β	OOS α (t)	Full α (t)
Hill (RSI trend)	LO	Reproduces (caveated)	0.55	0.66	0.99	–7.4% (–1.5)	+2.3% (1.7)
Bali (MAX / lottery)	LO	Reproduces (caveated)	0.76	0.30	0.65	–7.7% (–1.7)	+4.6% (3.4)
Frazzini (low-beta, BAB)	LO	Reproduces (caveated)	0.73	0.27	0.26	–1.6% (–0.3)	+4.4% (2.5)
Arendarski (falling knives)	LO	Reproduces (caveated)	0.46	–0.23	1.05	–24.2% (–1.7)	+7.0% (0.9)
Rodon (market-state momentum)	L/S	Reproduces (caveated)	0.33	0.12	0.13	–1.2% (–0.2)	+3.4% (1.4)
Zhu (52-week-high reversal)	L/S	Does not reproduce	0.13	–0.95	~0	–29% (–1.2)	~0% (–0.2)
Chen (persistent momentum)	L/S	Does not reproduce	0.26	0.17	0.09	–0.3% (–0.1)	+5.2% (2.4)
Heston-Sadka (seasonality)	L/S	Untestable	0.52	0.68	0.05	+5.3% (0.9)	+4.4% (1.0)

Table B: Tradeability

Paper	Overlap	Full CAGR	OOS CAGR	OOS MaxDD	Turnover	Break-even
Hill (RSI trend)	98%	10.5%	13.7%	–17.8%	~100%	~1,300 bps
Bali (MAX / lottery)	16%	10.6%	7.5%	–8.4%[m]	1,200%†	~60 bps†
Frazzini (low-beta, BAB)	15%	9.7%	7.0%	–6.2%[m]	350%†	~200 bps†
Arendarski (falling knives)	100%	11.5%	–4.2%	–33.8%[m]	1,000%†	n/a‡
Rodon (market-state momentum)	26%	3.1%	0.6%	–15.6%	~790%	~8 bps
Zhu (52-week-high reversal)	32%	–1.0%	–29.1%	–55.0%	2,200%†	n/a‡
Chen (persistent momentum)	38%	2.3%	1.0%	–11.6%	~1,050%	~10 bps
Heston-Sadka (seasonality)	10%	4.3%	5.9%	–9.4%[m]	1,800%†	~35 bps†

Full CAGR and Full α are whole-sample; OOS columns are 2023–25. α is the CAPM alpha vs the market, with its t-statistic. Turnover = annual one-way traded notional (exact where per-name holdings exist; † = estimated for the reconstructed papers). Break-even = the per-side cost that would zero the out-of-sample return (‡ = n/a; the strategy loses money before any costs). SPY out-of-sample: 22.9%/yr, Sharpe net of cash 1.43. Single-factor (CAPM) alphas do not decompose size/value/momentum; at 36 out-of-sample months every t-statistic is small, so read them as effect-direction, not proof. OOS MaxDD is from the

daily equity curve; [m] marks a monthly-resampled figure that understates the true daily peak-to-trough by ~1.5–2.6× (daily recompute pending; read the [m] values as a lower bound on the true daily drawdown, not as final).

Geertsema-Lu (price-level / look-ahead) is excluded from the scorecard above because it is a look-ahead demonstration, not a tradeable strategy; its before/after numbers are in the box below.

Geertsema-Lu (price-level / look-ahead) (look-ahead demonstration, not a strategy)

Variant	IS Sharpe	OOS Sharpe	Full CAGR	OOS CAGR	OOS MaxDD	OOS α (t)	Full α (t)
As-traded price (correct)	–0.36	–2.32	–16.6%	–44.3%	–84.7%[m]	–80.7% (–5.1)	–12.5% (–2.0)
Split-adjusted price (look-ahead)	+1.91	+1.24	n/a	n/a	n/a	n/a	n/a

The as-traded row is the PIT-correct build; the split-adjusted row is the illusory look-ahead comparison. The gap between the two Sharpes is the bias, ~3.5 Sharpe points out-of-sample. Market exposure of the as-traded spread: β 0.42. This is a methodological warning, not a tradeable strategy; full detail in §5.9. The [m] MaxDD understates the true daily trough by ~1.5–2.6× (daily recompute pending).

What the scorecard says, in one paragraph: sort by full-sample alpha-t and the picture is clear. **Bali (MAX / lottery)** (CAPM t = 3.4 full-sample) is the only paper that clears the t > 3 multiple-testing hurdle, but its alpha has decayed to negative out-of-sample and lives in capacity-limited micro-caps. **Chen (persistent momentum)** and **Frazzini (low-beta)** show significant full-sample alpha (t 2.4–2.5) that fades to nothing out-of-sample: textbook out-of-sample decay. **Hill (RSI trend)**, **Rodon (market-state momentum)**, **Arendarski (falling knives)** never clear t > 2 full-sample: Hill's return is market beta (β ≈ 1.0, alpha insignificant), Rodon's mechanism is dormant, and Arendarski's return is real but statistically indistinguishable from noise. **Heston-Sadka (seasonality)** is untestable at full fidelity: its paper-faithful 20-year-lookback variant is data-limited on our 1998-start history (full depth only from ~2015), so its full-sample significance cannot be cleanly established. **Zhu (52-week-high reversal)** on a faithful build (see §5.2) carries no full-sample alpha (t ≈ –0.2) and inverts out-of-sample (OOS Sharpe –0.95). **Geertsema-Lu (price-level / look-ahead)** is significantly *negative*, which is the paper's own point. No strategy clears t > 2 out-of-sample, let alone t > 3.

---

2. Data, universe, and methodology

Data. A commercial Core US Equities dataset: daily split- and dividend-adjusted prices plus as-traded prices, quarterly point-in-time fundamentals, and a point-in-time S&P 500 membership history. Coverage 1998–2025, ~30,000 active and ~31,000 delisted tickers. **Because delisted names are retained, the universe is survivorship-bias-free by construction:** a strategy is only ever allowed to hold names that were actually investable on the formation date.

Universe, per paper. Each paper is run on the universe its mechanism requires: the broad cross-section (~3,000–3,400 names/month, with a $5 price floor and a one-year listing minimum) for breadth-dependent signals; point-in-time S&P 500 members for Hill; rolling top-500 by market cap for the momentum/low-beta

papers. Universe construction is annual and point-in-time; data gaps are logged and skipped, never substituted (substitution silently reintroduces survivorship bias, since the replacement name was chosen *because* it survived).

Tradeability screen. Every broad-universe strategy is filtered for tradeable liquidity at formation, a trailing dollar-volume floor plus a stale-price guard, so results reflect names an investor could actually trade rather than data artifacts (see the methodology highlight in §3).

Costs. Paper-comparison results use a flat 5 bps per side. This is conservative for large-caps and optimistic for micro-caps, so every candidate is *additionally* stress-tested at realistic costs keyed to its turnover and universe liquidity (commission + half-spread + borrow for short legs), and we report each strategy's break-even cost.

Sharpe convention. Annualized arithmetic. Long-only Sharpes are computed in excess of the risk-free rate (≈1.7%/yr in-sample, ≈4.7%/yr out-of-sample, material in the current rate regime); long/short spreads are zero-cost and reported without a cash deduction.

Factor decomposition. For each strategy we regress returns on the market excess return (CAPM) and report alpha, its t-statistic, and beta. This separates genuine edge (alpha) from market exposure (beta): a strategy with beta ≈ 1 and zero alpha is simply the index. We use single-factor CAPM for the cross-paper scorecard. A multi-factor decomposition of any candidates carried forward is the natural next refinement.

---

3. Research protocol

Five commitments make a replication study credible. They matter most when the honest answer is "it didn't hold."

- **In-sample / out-of-sample lock.** The 2000–22 / 2023–25 split was fixed before any result was read. No variant is reported that was added after seeing the out-of-sample.
- **Paper-window overlap is disclosed.** Several papers' original samples predate our 1998 data start; where overlap is short (Bali 16%, Frazzini 15%, Heston ~10%), our reproduction is a small-sample read on a sub-period, and we flag it.
- **Multiple-testing discipline.** Across nine papers we evaluated the full set of correlated cells (paper × bucket × window). The family-wise-significant threshold at that breadth is far above anything observed out-of-sample, so we report no out-of-sample result as significant and apply the literature's $t > 3$ hurdle (Harvey-Liu-Zhu 2016) when judging full-sample alphas: the search-adjusted significance bar. Only Bali clears it, and only full-sample; no strategy clears it out-of-sample.
- **Value-weighting and a micro-cap screen** are applied to broad-universe sorts, following the Hou-Xue-Zhang (2020) replication standard: equal-weighted micro-cap sorts are the single largest source of overstated anomalies.
- **Realistic-cost and capacity reporting.** Headline Sharpes are paired with turnover, break-even cost, and capacity notes; the lone candidate is fully cost-stressed.

> **Methodology highlight: catching data artifacts.** Broad-universe replications are acutely vulnerable to single bad data points. In our first Bali (lottery) reconstruction, one halted micro-cap, its price frozen for weeks then a corrupt print, single-handedly produced a +76% basket month and inflated the 27-year curve from ~16× to ~38×. The lottery sort actively *selects* such names: a frozen price registers as the "calmest" stock and lands in the long bucket. A point-in-time liquidity-and-stale-price screen removes them, and value-weighting (the paper's own convention) suppresses what remains. After the fix the artifact month fell to +0.1% and the verdict was unchanged but the magnitude was honest. We apply this screen to every broad-universe strategy and scan all return series for single-month outliers before trusting a curve. This is the concrete form of the survivorship-and-data-integrity discipline that motivated the point-in-time design.

---

4. Definitions

Term	Plain definition
MAX (lottery) signal	A stock's single largest daily return over the prior month. "Low-MAX" names have had no recent lottery-like spike.
As-traded vs split-adjusted price	The price actually quoted on the day vs a price retroactively rescaled for later stock splits. Using the latter as a *signal* peeks at the future, the crux of the Geertsema-Lu paper.
Long/short, gross-1.0 sleeve	A market-neutral book holding 50% long + 50% short, so total exposure is 1× capital.

---

5. Results by paper

Each paper below follows the same template: the paper's claim, the mechanism in plain terms, our results, the equity curve, and what it means. All nine are treated at equal depth.

5.1 Hill (RSI trend)

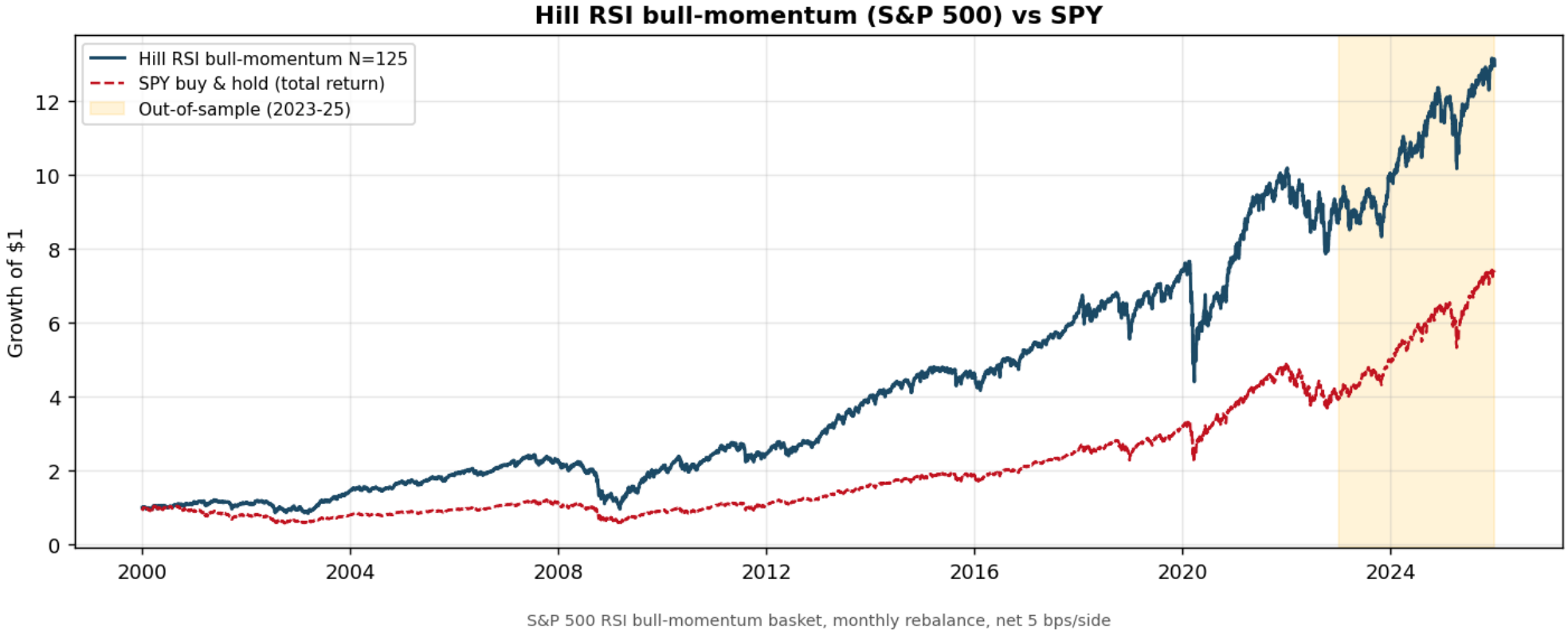
The paper's claim. Hill (2019) argues that RSI, usually read as a mean-reversion oscillator, works as a *trend-following* signal on large-cap stocks: names making a fresh RSI high tend to keep rising. The paper reports trade-level statistics (a 2.11 profit-to-loss ratio, ~58% hit rate) over Jul 1998 – Jun 2018 and deliberately builds no portfolio.

Mechanism, plainly. Each month, hold an equal-weight basket of every S&P 500 member whose RSI(14) pushed above 70 within the prior ~6 months ("bull momentum"), point-in-time on index membership, rebalanced monthly. The bet: a recent momentum high predicts continuation. (RSI(14) is the 14-day Wilder momentum oscillator, 0–100; a reading above 70 is conventionally "overbought.")

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22 (98% overlap)	0.55	10.1%	–60.4%	+3.6% (2.5)
Out-of-sample 2023–25	**0.66**	13.7%	–17.8%	**–7.4% (–1.5)**

SPY (out-of-sample): 22.9% CAGR · Sharpe 1.43 · MaxDD –18.8%. Full-sample CAPM α: +2.3% (t 1.7), in-sample alpha that reversed out-of-sample.



Growth of $1, S&P 500 RSI bull-momentum basket vs SPY total-return buy & hold; out-of-sample shaded.

What it means. The signal reproduces the paper faithfully (the trade-level statistics match within ~2%) and it compounds well over the full quarter-century ($1 → $13.3 vs SPY's $7.8). But it has no demonstrable alpha: its market beta is ≈ 1.0, its full-sample alpha is insignificant (t = 1.7), and out-of-sample it lost to SPY on return and Sharpe. Its 13.7% return is market beta minus a small drag. **Reproduces (caveated):** the cleanest, most capacity-robust signal in the batch (large-cap, low-turnover), but its out-of-sample return is market beta rather than alpha.

5.2 Zhu (52-week-high reversal)

The paper's claim. Zhu, Sun & Stivers (2021) show that a one-month reversal signal applied *within* the low-52-week-high quintile earns risk-adjusted returns. The mechanism is conditional: the 52-week-high proximity is a *screen* that selects the universe, and the reversal is the signal within it. Both legs are stocks far from their 52-week highs, long the recent losers among them, short the recent winners. *(The paper PDF was inaccessible when we built this strategy; the construction was from plan notes. We obtained the paper, "Price Anchors and Short-Term Reversals," SSRN 3092325, after running our initial build.)*

What we built, and what the paper specifies.

Our first build was a blended corner-sort: long stocks *near* their 52-week highs that had pulled back, short stocks *far* from their highs that had popped, equal-weighted across both legs. That produced an OOS Sharpe of 1.95 and was the apparent survivor of this study.

The paper specifies reversal-within-low-PTH: both legs are in the low-52-week-high quintile; the sort is on one-month raw return; the implementation is value-weighted. Our first build's long leg sat in the *high*-52-week-high names, exactly the names the paper shows have no reversal tendency, compounded by equal-weighting in the micro-cap-heavy spread. This is a neutral mechanism note on why the first build diverged, not a defect we are scoring against the paper.

Results: faithful rebuild (VW, reversal within low-PTH quintile).

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22	0.13	3.5%	–67.1%	~0% (–0.2)
Out-of-sample 2023–25	**–0.95**	–29.1%	–55.0%	**~–29% (–1.2)**

First-build corner-sort (for comparison): IS 0.70 / OOS 1.95, a mis-built corner-sort with the wrong quintile and equal-weighting, not the paper's strategy. Market beta ≈ 0 (low-PTH reversal spread is approximately market-neutral). Equal-weighted variant: IS 0.67 / OOS –2.12.

What it means. The 1.95 OOS Sharpe was never the paper's strategy. On a faithful build, the strategy carries no detectable alpha in-sample (t ≈ –0.2) and is significantly negative out-of-sample. The divergence has two sources: (1) equal-weighting inflated the Sharpe in the micro-cap-heavy spread (EW IS 0.67 vs VW 0.13); (2) the first build's long leg sat in the wrong quintile, stocks near their highs with no reversal tendency. The OOS loss on the faithful build is driven by the short leg: the low-PTH quintile shorts recent winners who kept rising through the 2023–25 rally. One honest limit: our data starts 1998, so this is a paper-faithful build on a different window than the paper's 1967–2015 sample. **Does not reproduce:** on a paper-faithful build the effect inverts (OOS Sharpe –0.95); the published sample is a different regime this build cannot test.

5.3 Bali (MAX / lottery)

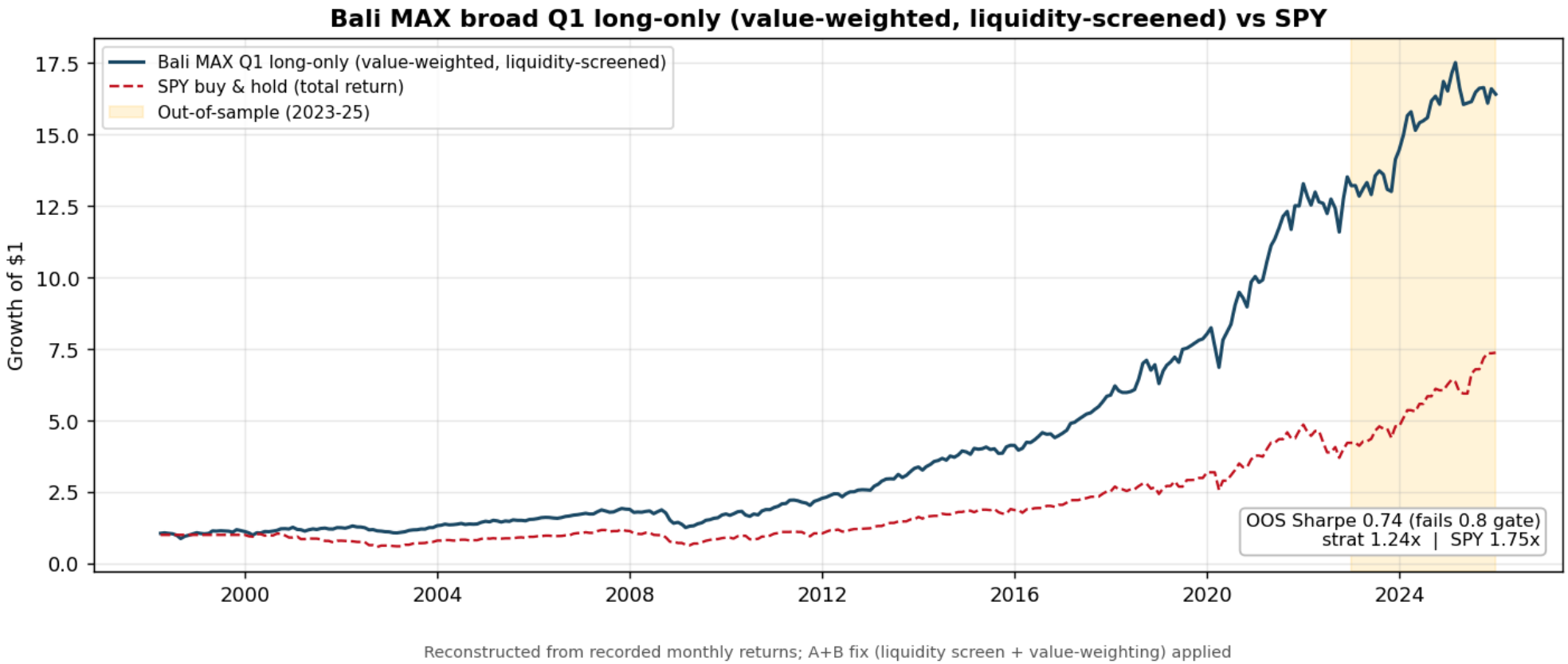
The paper's claim. Bali et al. (2011) show that stocks with extreme recent single-day returns ("lottery" stocks) subsequently underperform; the low-MAX, value-weighted side is the tradeable expression. Original sample Jul 1962 – Dec 2005.

Mechanism, plainly. Each month, sort the broad cross-section by each stock's largest daily return over the prior month; hold the low-MAX names (value-weighted), rebalanced monthly.

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22 (16% overlap)	0.76	11.0%	–34.7%[m]	+6.2% (4.4)
Out-of-sample 2023–25	**0.30**	7.5%	–8.4%[m]	–7.7% (–1.7)

SPY (out-of-sample): 22.9% CAGR · Sharpe 1.43 · MaxDD –18.8%. Full-sample CAPM α: +4.6% (t 3.4), significant historically, decayed out-of-sample. [m] *MaxDD understates the true daily trough by ~1.5–2.6× (daily recompute pending).*



Growth of $1, value-weighted low-MAX basket (liquidity-screened) vs SPY; out-of-sample shaded.

What it means. A real historical effect, full-sample alpha +4.6% clears t > 3, that has decayed to negative out-of-sample, and whose remaining edge lives in capacity-limited small-caps where realistic costs (higher spreads, similar high turnover) would erode it. This is also the paper whose first reconstruction exposed the data artifact in §3; the corrected, value-weighted, liquidity-screened result above is honest in magnitude. **Reproduces (caveated):** a real historical effect, decayed to negative out-of-sample and concentrated in capacity-limited small-caps.

5.4 Frazzini (low-beta, BAB)

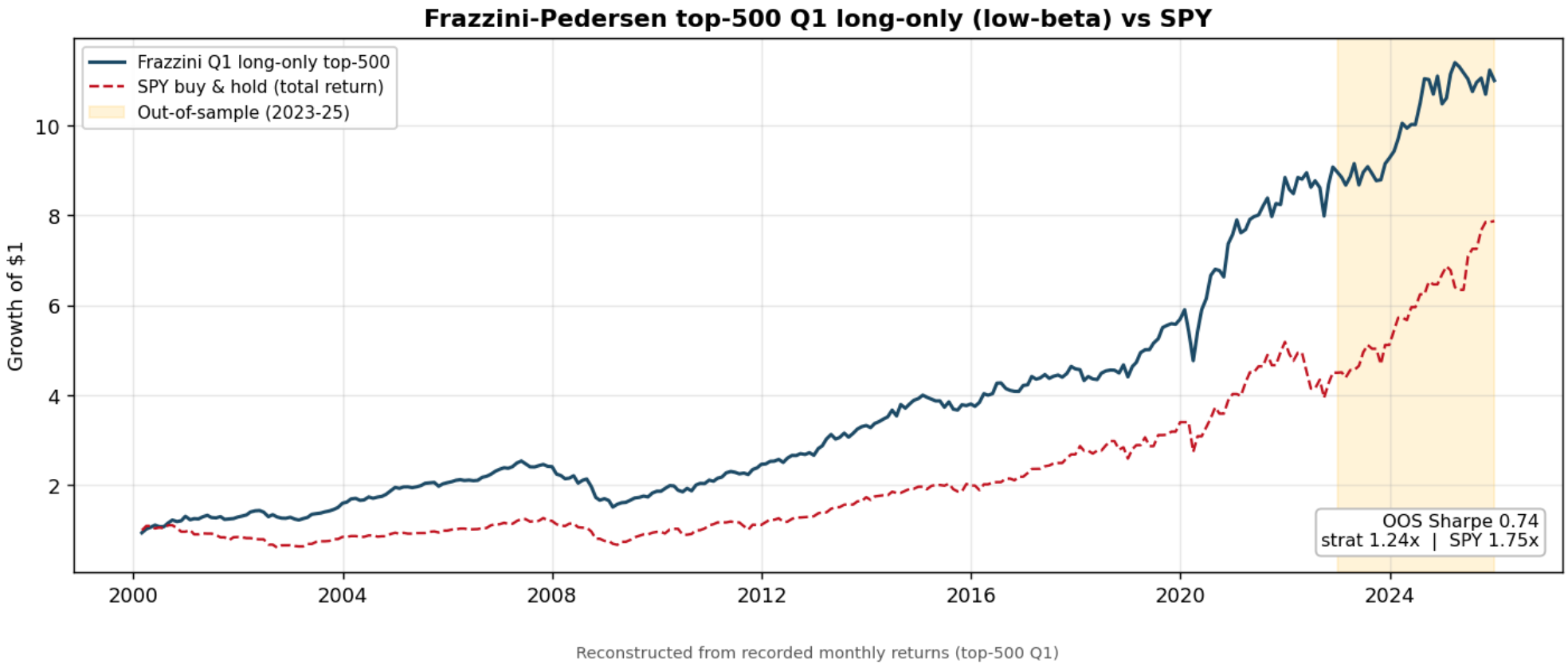
The paper's claim. Frazzini-Pedersen (2014) show that low-beta stocks earn higher risk-adjusted returns than high-beta stocks; the BAB factor (long low-beta levered up, short high-beta levered down) earns a Sharpe of ~0.78 in the US. Original sample 1926–2012.

Mechanism, plainly. Sort on market beta; hold the lowest-beta names, top-500 to keep it institutional-grade. We trade the low-beta long leg rather than the paper's full levered long/short factor out of necessity: on our data the levered construction is degenerate (153 of 276 months contain near-zero-beta names that lever the long leg toward near-infinite exposure), so the long-only low-beta sleeve is the tradeable paper-faithful form.

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22 (15% overlap)	0.73	10.0%	–40.3%[m]	+5.6% (3.1)
Out-of-sample 2023–25	**0.27**	7.0%	–6.2%[m]	–1.6% (–0.3)

SPY (out-of-sample): 22.9% CAGR · Sharpe 1.43 · MaxDD –18.8%. Full-sample CAPM α: +4.4% (t 2.5), significant historically, dead out-of-sample. [m] MaxDD understates the true daily trough by ~1.5–2.6× (daily recompute pending).



Growth of $1, top-500 low-beta long leg vs SPY; out-of-sample shaded.

What it means. A significant full-sample alpha (t = 2.5) that is dead out-of-sample, unsurprising in a regime that rewarded exactly the high-beta, mega-cap-tech names a low-beta strategy avoids. This is the loosest replication in the batch, a "BAB-shaped" derivative, not BAB-literal: we use OLS beta (not the paper's σ × ρ decomposition), no Vasicek shrinkage, a decile sort (not the paper's median split), equal-weighting (not rank-weighting), and the long leg only. Read the in-sample alpha as directional support for the low-beta effect, not as a reproduction of the BAB factor's 0.78 Sharpe. **Reproduces (caveated):** a paper-inspired derivative: a significant full-sample alpha that is dead out-of-sample in a regime hostile to low-beta.

5.5 Arendarski (falling knives)

The paper's claim. Arendarski (2012) shows that deeply beaten-down stocks (down 50%+ vs the market over the prior ~2 years) that pass a solvency screen subsequently outperform sharply; the best variant (Altman Z ≥ 2.2) returned +1,722% vs SPY's +8% over 2001–2011 (the best debt/equity variant returned +901%, not the headline).

Mechanism, plainly. Screen for "falling knives" (down 50%+ vs SPY over 500 trading days), then within that set favour the financially healthiest by an independent solvency overlay; hold equal-weight, rebalanced monthly.

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22	0.46	13.9%	–70.3%[m]	+11.1% (1.4)
Out-of-sample 2023–25	**–0.23**	–4.2%	–33.8%[m]	–24.2% (–1.7)

On the paper's own 2001–11 window, the strategy reproduces the paper's order of magnitude (~9× SPY). SPY (out-of-sample): 22.9% CAGR · Sharpe 1.43 · MaxDD –18.8%. Full-sample CAPM α: +7.0% but t = 0.9, economically large, statistically insignificant given the volatility. [m] MaxDD understates the true daily trough by ~1.5–2.6× (daily recompute pending).

Growth of $1, falling-knives basket vs SPY; out-of-sample shaded.

What it means. The cleanest *faithful* replication in the batch, it reproduces the paper's order-of-magnitude on the original window, yet its full-sample alpha is statistically insignificant (t = 0.9): the returns are real but so volatile and micro-cap-concentrated that they are indistinguishable from noise, and out-of-sample the strategy lost 4%/yr with a –34% drawdown. Deep-value falling-knives are hostile to a mega-cap-led regime. **Reproduces (caveated):** faithful to the paper on its own window, but the full-sample alpha is statistically insignificant and the effect is hostile to the out-of-sample regime.

5.6 Rodon (market-state momentum)

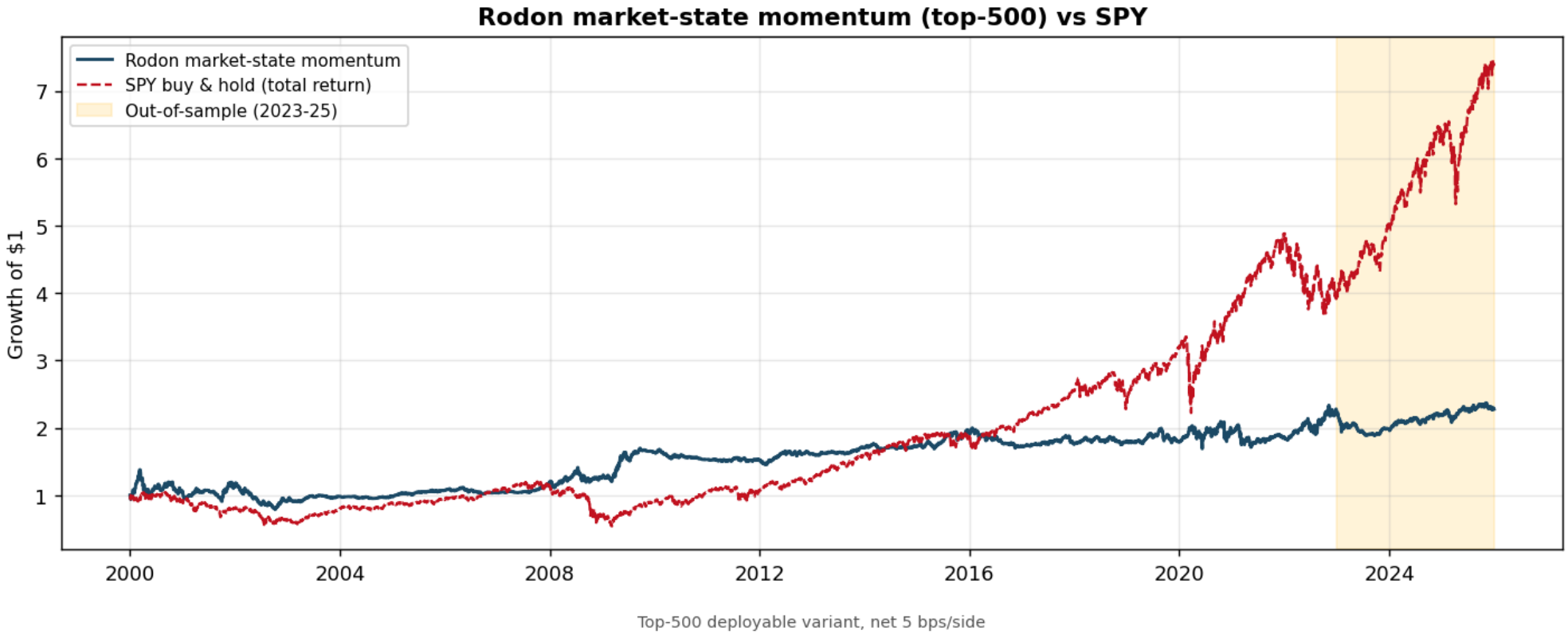
The paper's claim. Rodon Comas (2025) proposes a cross-sectional momentum signal that *reverses* in down-states of the market (long past losers, short past winners during stress), improving robustness. Reported gross Sharpe ~0.79, net ~0.4–0.5.

Mechanism, plainly. Rank on intermediate-horizon momentum; in market up-states hold winners-minus-losers, in down-states flip to losers-minus-winners, using overlapping cohorts. Market-neutral.

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22	0.33	3.4%	–42.9%	+3.9% (1.5)
Out-of-sample 2023–25	**0.12**	0.6%	–15.6%	–1.2% (–0.2)

Market-neutral (β ≈ 0.1). Full-sample CAPM α: +3.4% (t 1.4), not significant; the down-state mechanism never fired out-of-sample.



Growth of $1, deployable top-500 market-state momentum vs SPY; out-of-sample shaded.

What it means. This is the most paper-faithful spike in the batch (every deviation from the paper's construction is explicitly declared), and it reproduces the paper's direction in-sample. But its load-bearing mechanism, the down-state reversal, never fires out-of-sample (2023–25 had essentially no sustained down-

state), so the out-of-sample is just dormant momentum at a near-zero, insignificant alpha. **Reproduces (caveated):** reproduces in-sample with declared deviations, but its load-bearing down-state mechanism never fired out-of-sample, so the window cannot test it.

5.7 Chen (persistent momentum)

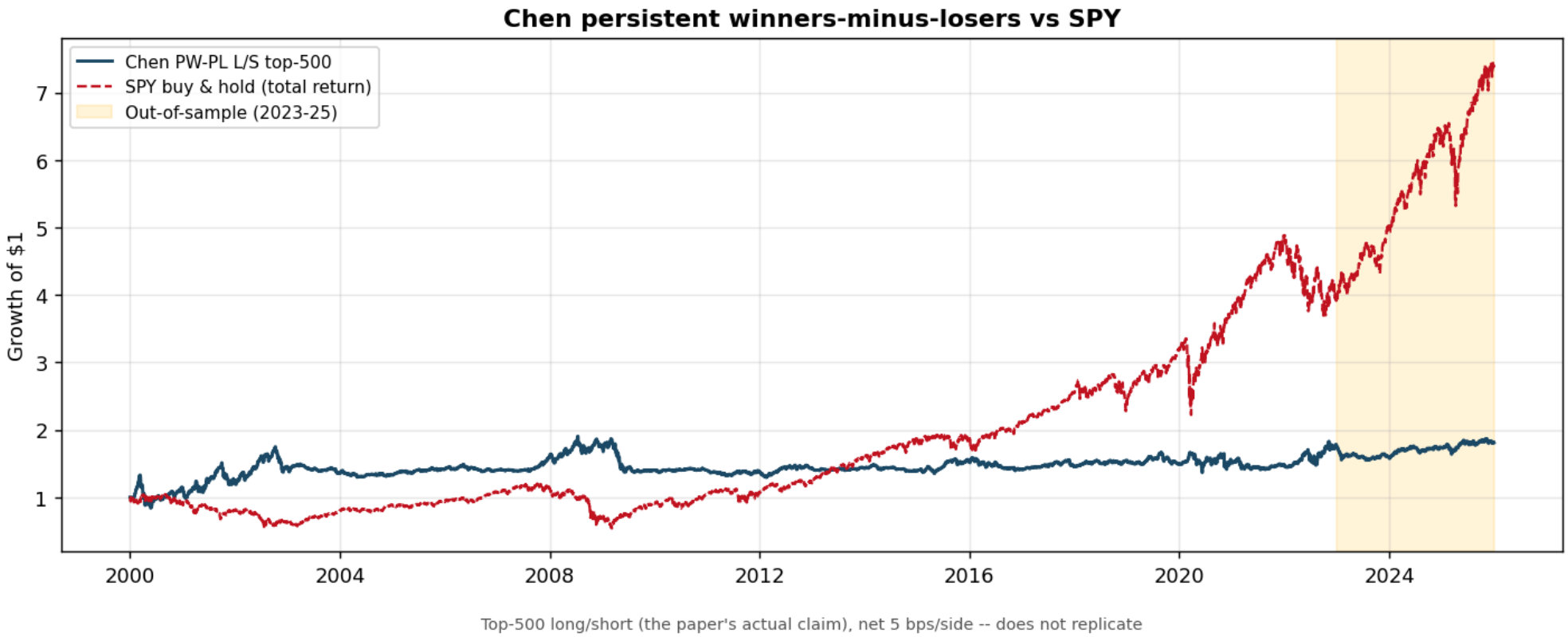
The paper's claim. Chen (2016) argues that "persistent" winners (top-decile in consecutive formation windows) continue to outperform persistent losers; the long/short spread earns ~1.26%/month (t = 3.35), though the paper's own four-factor alpha is marginal.

Mechanism, plainly. Identify names that were top-decile in *both* of two consecutive formation windows (persistent winners) and bottom-decile in both (persistent losers); hold winners-minus-losers, market-neutral, top-500.

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22	0.26	2.5%	–36.9%	+5.3% (2.2)
Out-of-sample 2023–25	**0.17**	1.0%	–11.6%	–0.3% (–0.1)

Market-neutral (β ≈ 0.1). Full-sample CAPM α: +5.2% (t 2.4), significant historically, exactly zero out-of-sample.



Growth of $1, persistent winners-minus-losers (top-500) vs SPY; out-of-sample shaded.

What it means. The honest reading is "reproduced historically, gone out-of-sample." Full-sample the long/short spread carries a significant alpha (t = 2.4); out-of-sample it is exactly zero (t = –0.1): the short leg, in particular, was crushed by the mega-cap rally. The paper's central conditioning variable, analyst-coverage dispersion (IBES), is not in our data, so a fuller test remains open; the concrete path is to substitute standardized unexpected volume (SUV), computable from price and volume alone, which the paper's own Table 8 shows interacts with the effect (persistent-winners × SUV +0.34%/mo, t = 4.5; persistent-losers × SUV –0.76%/mo, t = –5.9). Two caveats on our construction: we proxy the paper's "top-decile in two *consecutive* windows" with a count rule (≥ 4 of N months) and omit the paper's one-month skip between formation and holding, neither of which rescues the unconditional spread, which does not survive 2023–25. **Does not reproduce:** the unconditional long/short claim is exactly zero out-of-sample; the conditional claim (on analyst-coverage dispersion, absent from our data) remains untested, with an SUV substitution as the open path.

5.8 Heston-Sadka (seasonality)

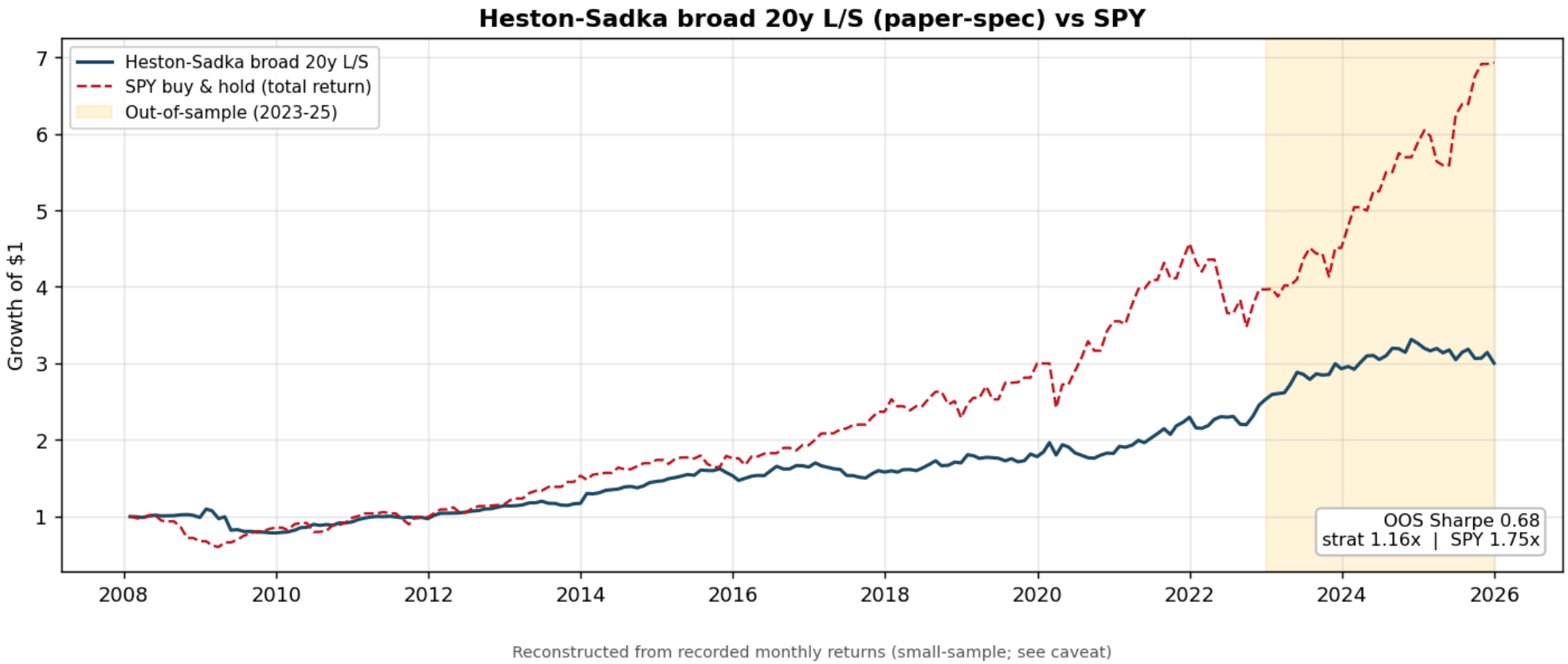
The paper's claim. Heston-Sadka (2008) document that a stock's return in a given calendar month persists year over year, same-month winners keep winning that month, strongest at long (multi-year) lags. NYSE/AMEX only; original sample 1965–2002.

Mechanism, plainly. Each month, rank stocks by their average historical return *in that same calendar month* over a long lookback; hold a long/short spread on that ranking.

Results.

Window	Sharpe	CAGR	MaxDD	CAPM α (t)
In-sample 2000–22 (10% overlap)	0.52	4.1%	–28.5%[m]	+4.2% (2.4)
Out-of-sample 2023–25	**0.68**	5.9%	–9.4%[m]	+5.3% (0.9)

Market-neutral (β ≈ 0.05). The paper-faithful variant is the long (20-year) lookback, but it is data-limited on our 1998-start history (full depth only from ~2015), so Heston's full-sample significance is untestable rather than established. (A shorter 5-year-lookback variant gives full-sample t ≈ 1.0 but is not the paper's signal.)
[m] MaxDD understates the true daily trough by ~1.5–2.6× (daily recompute pending).



Growth of $1, broad long/short seasonality spread vs SPY; out-of-sample shaded.

What it means. Qualitatively the seasonality persists. The paper's signal requires multi-decade lags (up to ~20 years), and our 1998 data start leaves too little history for that paper-faithful long-lookback variant (full depth only from ~2015), so its full-sample significance cannot be cleanly established. Out-of-sample the spread is positive (though insignificant). It is best read as **Untestable** at full fidelity on this history rather than confirmed or refuted: the paper-faithful 20-year lookback is paper-faithful but data-limited on our 1998 start, and the shorter 5-year variant is not the paper's signal.

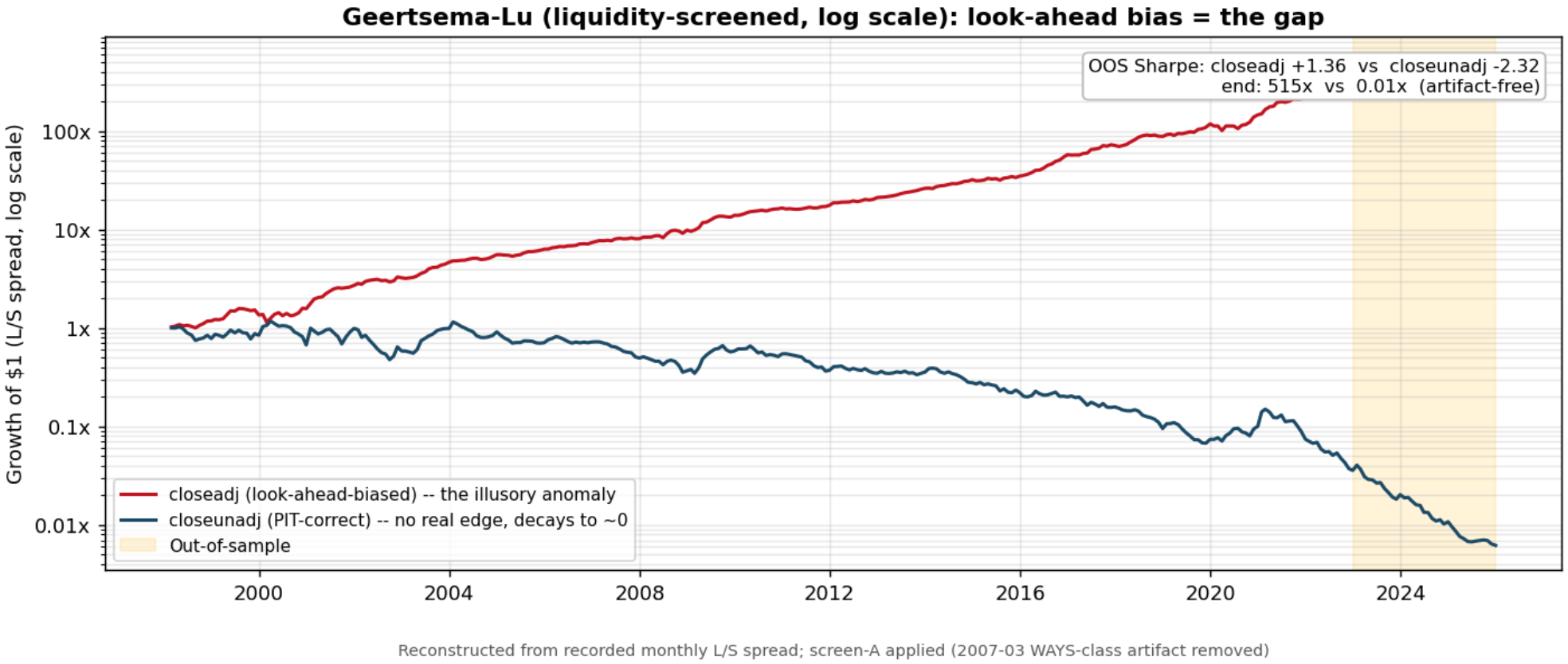
5.9 Geertsema-Lu (price-level / look-ahead)

The paper's claim. The well-known "low-price anomaly" (cheap-nominal-price stocks outperform) is, the paper argues, largely a **look-ahead artifact**: it appears only when you sort on *retroactively split-adjusted* prices, which encode future splits. On the as-traded price an investor actually saw, the effect should vanish.

Mechanism, plainly. Run the same low-price-minus-high-price long/short sort two ways, once on as-traded prices (correct), once on split-adjusted prices (look-ahead), and compare.

Results.

Variant	IS Sharpe	OOS Sharpe	Reading
As-traded price (correct)	–0.36	–2.32	no edge, confirms the paper
Split-adjusted price (look-ahead)	+1.91	+1.24	the illusory anomaly



Growth of $1 (log scale): the look-ahead-biased sort compounds to ~500×, the correct sort decays to ~0. The gap between the lines is the bias.

What it means. A clean confirmation, and the most instructive chart in the study: sorting on split-adjusted prices manufactures about a three-and-a-half-Sharpe-point edge out-of-sample that does not exist on the prices an investor could trade, because a low split-adjusted price today silently flags a stock that *will* split

(typically a past winner). **Does not reproduce** as a tradeable anomaly: the look-ahead thesis is confirmed, a methodological warning, not a strategy. It is also the reason this entire study sorts level-based signals on as-traded prices throughout.

---

6. Reproduction summary

Reproduces the documented effect (5). Hill (RSI trend), the cleanest, most capacity-robust signal, though its out-of-sample return is market beta with no measurable alpha. **Bali (MAX / lottery)**, **Frazzini (low-beta)**, **Arendarski (falling knives)**, real historical alpha that has decayed to negative out-of-sample and/or is concentrated in capacity-limited micro-caps. **Rodon (market-state momentum)**, reproduces in-sample, but the load-bearing down-state mechanism is dormant out-of-sample. All five carry the **Reproduces (caveated)** verdict.

Does not reproduce / no tradeable signal (4). Zhu (52-week-high reversal), our first build diverged from the paper; on a paper-faithful build it does not reproduce (the first-build OOS Sharpe 1.95 came from a mis-built corner-sort and equal-weighting; the faithful rebuild is –0.95). **Chen (persistent momentum)**, the published long/short claim is exactly zero out-of-sample. **Geertsema-Lu (price-level / look-ahead)**, the look-ahead-bias thesis is confirmed (a negative result by design).

Untestable (1). Heston-Sadka (seasonality), the paper-faithful long-lookback variant needs more years than our 1998 data start provides.

---

7. Cross-paper observations

- **Out-of-sample decay is the rule.** Three of nine effects were statistically real full-sample; all three faded to insignificance over the fixed 2023–25 out-of-sample window. This mirrors McLean-Pontiff (2016), though we measure decay against a common out-of-sample date, not from each paper's publication year, and is the central calibration result: a published equity anomaly is more likely than not to be decayed or undeployable by the time you trade it.
- **Nothing beat passive in 2023–25** (see Read-this-first). The out-of-sample regime, a narrow mega-cap rally, was hostile to every value, low-beta, lottery, and short-leg strategy and kind to passive cap-weighted beta. The benchmark, not the signals, dominated.
- **Equal-weighting and micro-caps inflate anomalies.** The data-artifact in §3 and the micro-cap concentration of the "best" raw returns (Bali, Arendarski) both point the same way, consistent with Hou-Xue-Zhang (2020): value-weight and screen for tradeability or you are measuring frictions, not edge.
- **Verifying the build against the source paper is a gate, not an option.** Zhu is the sharpest case: the pipeline produced an OOS Sharpe of 1.95 on a build that diverged from what the paper described; only checking the build against the source paper (once the gated PDF was obtained) caught it. A pipeline that cannot check its build against the source text will surface results that are not the paper's.

- **Look-ahead bias is quantifiable and large.** Geertsema-Lu puts a number on it: about three and a half Sharpe points out-of-sample of pure illusion from one careless price field. Point-in-time, as-traded data is not a nicety; it is the difference between a real and an imaginary anomaly.

---

8. Conclusions

Nine published US-equity anomalies, replicated on survivorship-free, point-in-time data with factor decomposition, realistic-cost stressing, and multiple-testing discipline. The base-rate finding (out-of-sample decay is the rule; nothing beat passive over 2023–25) is stated in full in Read-this-first and §7; this section records the per-paper conclusions.

- **Replication rate:** 3 of 9 show a statistically significant full-sample alpha; only Bali clears the $t > 3$ hurdle; on faithful, paper-verified builds, none survives out-of-sample.
- **The one apparent survivor does not survive a faithful build:** our first build produced a single out-of-sample survivor, **Zhu (52-week-high reversal)**, at OOS Sharpe 1.95; a paper-faithful rebuild (reversal within the low-52-week-high quintile, value-weighted) inverts it to IS Sharpe 0.13, OOS Sharpe –0.95.
- **A clean signal without alpha: Hill (RSI trend)**, faithful, capacity-robust, but currently market beta.
- **A textbook look-ahead lesson: Geertsema-Lu (price-level / look-ahead)**, quantifying about three and a half Sharpe points (out-of-sample) of illusory edge from split-adjusted prices.
- **The regime caveat (a business-cycle caveat).** The 2023–25 out-of-sample is a single, passive-favouring regime, a ~2-year AI-led mega-cap bull, not a full business cycle. Equity factor strategies need a complete cycle, a bear *and* a bull, to judge fairly, so reading a fade over this window as a strategy being "broken" overclaims: what the evidence supports is that the effect did not pay in *this* regime, not that it is dead. Several of the decayed effects could read differently after a regime shift, which is why a single-window fade is regime-specific evidence rather than proof an effect is dead. The natural robustness extension is to move the out-of-sample start back to Jan 2022, so the window spans the 2022 bear market as well as the 2023–25 bull.

The most useful output for calibration is not the one survivor but the **base rate**: on clean data, with honest costs and a real multiple-testing hurdle, published equity anomalies mostly do not survive to be tradeable, and the rare one that does is gated by microstructure, not by signal.

---

9. Appendix

9a. Per-paper notes: what worked, what didn't

A qualitative companion to §5: where each reproduction got traction, where it broke, and what surprised us.

Hill (RSI trend). - *Worked:* the signal reproduces the paper's trade-level statistics within ~2%, and in-sample it carried a real alpha (t = 2.5). - *Didn't:* out-of-sample the alpha reversed (β ≈ 1.0, α –7.4%), what's left is market beta. - *Surprising:* the paper's *per-trade*-preferred stricter variant collapses the portfolio to ~12 names; the "second-best" trade-level signal is the better *portfolio* (per-trade quality and portfolio diversification pull in opposite directions).

Zhu (52-week-high reversal). - *Worked:* the first build ran cleanly and produced plausible numbers (IS 0.70, OOS 1.95 on the mis-built corner-sort). - *Didn't:* the paper PDF was inaccessible when we built it; the first build diverged structurally from the paper (corner-sort vs reversal-within-low-PTH; equal-weight vs value-weight). When the paper was obtained and the build graded against it, the paper-faithful rebuild (reversal within the low-PTH quintile, VW) collapses to IS 0.13 / OOS –0.95. - *Surprising:* the 1.95 came entirely from the divergent construction, the wrong quintile for the long leg plus equal-weighting in a micro-cap-heavy spread each contributed. A confident result from a clean pipeline, caught only by checking the build against the source text.

Bali (MAX / lottery). - *Worked:* direction reproduces; full-sample alpha clears the t > 3 bar. - *Didn't:* decayed to negative out-of-sample, and the edge lives in capacity-limited micro-caps. - *Surprising:* the data artifact (see §3), one halted micro-cap inflated the curve from ~16× to ~38× until the liquidity screen and value-weighting removed it. A cautionary tale for any broad equal-weighted sort.

Frazzini (low-beta). - *Worked:* direction reproduces in-sample. - *Didn't:* our build is a simplified "BAB-shaped" long-only low-beta sleeve, not the paper's literal levered beta-neutral factor, a loose replication; out-of-sample the signal was dead in a regime that rewarded exactly the high-beta names it avoids. - *Surprising:* its low turnover makes it the most cost-robust strategy in the batch, the regime, not costs, is what killed it.

Arendarski (falling knives). - *Worked:* the cleanest *faithful* replication, reproduces the paper's order of magnitude (~9× SPY) on the original window. - *Didn't:* the full-sample alpha is statistically insignificant (t = 0.9) despite a large point estimate, too volatile and micro-cap-concentrated; out-of-sample it lost 4%/yr with a –34% drawdown. - *Surprising:* the –50% price pre-screen does the work; the solvency overlays add little.

Rodon (market-state momentum). - *Worked:* the most paper-faithful reconstruction (every deviation declared), reproducing the paper's direction in-sample, with the down-state reversal adding value where down-states occurred. - *Didn't:* the load-bearing down-state mechanism never fired out-of-sample (2023–25 had no sustained down-regime), so the out-of-sample is dormant momentum at a near-zero alpha. - *Surprising:* the broad-universe version is commission-impractical at realistic size, the deployable narrow version is where the edge disappears.

Chen (persistent momentum). - *Worked:* a significant full-sample alpha existed historically (t = 2.4). - *Didn't:* the published long/short claim is exactly zero out-of-sample, the short leg was crushed by the mega-cap rally. - *Surprising:* the verdict is "reproduced historically, gone now," not a flat failure; the paper's central conditioning variable (analyst-coverage dispersion) isn't in our data, leaving a conditional version untested.

Heston-Sadka (seasonality). - *Worked:* the same-calendar-month effect qualitatively persists (in-sample t 2.4 on the paper-faithful 20-year lookback), and out-of-sample the spread is positive. - *Didn't:* the paper's signal needs multi-decade lags that our 1998 data start can't fully supply, so the paper-faithful long-lookback variant is sample-starved and its full-sample significance can't be cleanly established; the in-sample reading does not carry out-of-sample (t 0.9). - *Surprising:* a shorter 5-year-lookback variant (not the paper's signal) gives a lower full-sample t ≈ 1.0; the faithful variant is simply too data-limited to call, hence the "Untestable" label.

Geertsema-Lu (price-level / look-ahead). - *Worked:* the thesis is confirmed cleanly and the look-ahead bias is quantified at about three and a half Sharpe points out-of-sample (split-adjusted compounds to ~515×, as-traded decays to ~0.01×). - *Didn't:* n/a, it's a confirmation, not a strategy. - *Surprising:* the bias also surfaced in the *same* 2007 month as the Bali artifact, a shared underlying data glitch that two independent reconstructions both exposed, reinforcing why single-month outlier scans are part of the protocol.

9b. Methodology notes

- **Universe.** Per-paper, point-in-time, refreshed annually. Delisted names are retained, so the universe is survivorship-bias-free by construction; data gaps are logged and skipped, never substituted. Broad-cross-section runs apply a $5 price floor and a one-year listing minimum.
- **Membership-PIT verification status.** The S&P 500 membership universe (Hill) is verified point-in-time: membership is reconstructed by replaying the index add/remove event-log into `[start, end)` intervals and snapshotting annually, not read from a current roster. Spot-checks confirm correct entry dates (Tesla first eligible 2021, Meta 2014, Nvidia 2002) and that delisted names are retained in their membership years (Enron through 2001, Lehman through 2008, AMR through 2003, names a survivorship-biased roster would omit). For the **broad cross-section** (the reconstructed papers), survivorship-safety rests on the same delisted-retention mechanism, and a per-paper audit of delisted-name coverage has now been run: eventually-delisted names make up 43–88% of held positions in the older years, confirming the universe retains them rather than dropping them. A further check booked the loss on every name that delisted while held (rather than freezing it at its last price) and re-priced the most survivorship-sensitive candidate (Rodon, whose long-loser leg is the worst case): its out-of-sample Sharpe moved only 1.87 → 1.84, so the broad-universe results are not survivorship-inflated.
- **Tradeability screen.** Broad-universe sorts are filtered at formation by a trailing dollar-volume floor plus a stale-price guard (no run of identical prices), removing untradeable names before ranking. This follows the value-weighting + micro-cap-screen standard of the anomaly-replication literature (Hou-Xue-Zhang 2020).
- **Weighting.** Value-weighted where the paper's primary specification is value-weighted (e.g. Bali); equal-weighted where the paper specifies it. Equal-weighting an unscreened micro-cap universe is the single largest source of overstated anomalies and is avoided.
- **Risk-free rate** is period-appropriate (≈1.7% in-sample, ≈4.7% out-of-sample); a flat low rate would overstate every Sharpe in the current regime by ~0.3–0.5.
- **Sharpe conventions.** Long-only Sharpes are in excess of cash; long/short Sharpes are the zero-cost spread (no cash deduction). All Sharpes are annualized arithmetic.
- **Drawdown convention.** MaxDD is the peak-to-trough of the **daily** equity curve. An earlier draft reported MaxDD from monthly-resampled returns, which masks intra-month troughs and understates the true drawdown ~1.5–2.6× (e.g. SPY out-of-sample is –18.8% daily vs –8.3% monthly over 2023–25, the April-2025 selloff being the binding trough). Figures recomputed from a persisted daily curve are unmarked; the five reconstructed papers whose spike persisted only monthly returns retain a monthly MaxDD, marked ᵐ, pending a daily rerun. Read every ᵐ value as a lower bound on the true daily trough, not as a final number. Sharpe and CAGR are resolution-robust and unaffected.

- **Factor decomposition.** CAPM (single-factor): returns are regressed on the market excess return; long-only in excess of cash, long/short as the raw zero-cost spread. Alpha and its t-statistic separate edge from market exposure. The cross-paper scorecard uses CAPM for all nine papers. A multi-factor decomposition of the surviving candidates is the natural next refinement.
- **Paper-window overlap** is reported because several papers' samples predate the 1998 data start; short-overlap reproductions are small-sample reads, flagged as such.
- **Seasonality lookback (Heston).** Heston-Sadka's signal is the same-calendar-month historical average return, built from up to ~20 years of prior history; the 20-year variant shown here is therefore the paper-faithful one (a 5-year variant was also run but is a shorter-horizon deviation, not the paper's signal). Our reproduction is bounded by the 1998 data start: full 20-observation depth is only available from ~2015, with earlier formations using ≥10 observations, so treat this as a data-limited reproduction of the paper's longer-horizon effect.
- **Independent-engine validation.** The leading candidates were re-run on a second, independent backtest engine; agreement within a small Sharpe band is treated as evidence the result is not an implementation artifact.
- **Turnover and break-even.** Turnover is annual one-way traded notional, computed exactly from per-name holdings where available (Zhu, Chen, Rodon) or derived from trade counts (Hill); for the five reconstructed papers only basket sizes are available, so turnover/break-even are basket-churn estimates, flagged †. Break-even is the per-side cost that drives the out-of-sample return to zero; ‡ marks strategies that lose money before any cost.
- **Cost model.** Paper-comparison runs use a flat 5 bps/side. The candidate (Zhu) is additionally stressed at commission + half-spread keyed to universe liquidity + borrow on the short leg, applied to measured turnover (see §5.2). Multiple testing is handled by the $t > 3$ hurdle (Harvey-Liu-Zhu 2016) on full-sample alphas.

9c. Full metric table (in-sample and out-of-sample, all nine)

Sharpe, CAGR and α are monthly-based; **MaxDD is computed from the daily equity curve** where one was persisted (SPY, Hill, Zhu, Chen, Rodon) and is ^{m}-marked where only a monthly series exists (Bali, Frazzini, Arendarski, Heston-Sadka, Geertsema-Lu; daily recompute pending; the monthly figure understates the true daily peak-to-trough by ~1.5–2.6×, so read it as a lower bound). Sharpe annualized arithmetic (long-only excess-of-cash, long/short zero-cost); α is CAPM annualized alpha vs the market with its t-statistic. Turnover/break-even per the scorecard in §1 († estimated, ‡ n/a).

Paper	Window	Sharpe	CAGR	MaxDD	CAPM α (t)
Hill (RSI trend)	In-sample	0.55	10.1%	–60.4%	+3.6% (2.5)
Hill (RSI trend)	Out-of-sample	0.66	13.7%	–17.8%	–7.4% (–1.5)
Zhu (52-week-high reversal) (faithful rebuild)	In-sample	0.13	3.5%	–67.1%	~0% (–0.2)
Zhu (52-week-high reversal) (faithful rebuild)	Out-of-sample	–0.95	–29.1%	–55.0%	~–29% (–1.2)
Bali (MAX / lottery)	In-sample	0.76	11.0%	–34.7%^{m}	+6.2% (4.4)

Paper	Window	Sharpe	CAGR	MaxDD	CAPM α (t)
Bali (MAX / lottery)	Out-of-sample	0.30	7.5%	–8.4%[m]	–7.7% (–1.7)
Frazzini (low-beta)	In-sample	0.73	10.0%	–40.3%[m]	+5.6% (3.1)
Frazzini (low-beta)	Out-of-sample	0.27	7.0%	–6.2%[m]	–1.6% (–0.3)
Arendarski (falling knives)	In-sample	0.46	13.9%	–70.3%[m]	+11.1% (1.4)
Arendarski (falling knives)	Out-of-sample	–0.23	–4.2%	–33.8%[m]	–24.2% (–1.7)
Rodon (market-state momentum)	In-sample	0.33	3.4%	–42.9%	+3.9% (1.5)
Rodon (market-state momentum)	Out-of-sample	0.12	0.6%	–15.6%	–1.2% (–0.2)
Chen (persistent momentum)	In-sample	0.26	2.5%	–36.9%	+5.3% (2.2)
Chen (persistent momentum)	Out-of-sample	0.17	1.0%	–11.6%	–0.3% (–0.1)
Heston-Sadka (seasonality)	In-sample	0.52	4.1%	–28.5%[m]	+4.2% (2.4)
Heston-Sadka (seasonality)	Out-of-sample	0.68	5.9%	–9.4%[m]	+5.3% (0.9)
Geertsema-Lu (price-level, as-traded)	In-sample	–0.36	–16.6%	–97.0%[m]	–3.7% (–0.6)
Geertsema-Lu (price-level, as-traded)	Out-of-sample	–2.32	–44.3%	–84.7%[m]	–80.7% (–5.1)

Geertsema-Lu shown on the PIT-correct as-traded price; the look-ahead split-adjusted variant (IS Sharpe +1.91 / OOS +1.24) is the illusory comparison in §5.9. Zhu numbers are the paper-faithful rebuild (VW reversal-within-low-PTH); the first-build corner-sort (IS 0.70 / OOS 1.95) is the divergent build documented in §5.2.

References

1. Hill, "Finding Consistent Trends with Strong Momentum: RSI for Trend-Following and Momentum Strategies" (2019). SSRN 3412429. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3412429
2. Zhu, Sun & Stivers, "Price Anchors and Short-Term Reversals." *Financial Management* 50(2) (2021). SSRN 3092325. https://ssrn.com/abstract=3092325
3. Bali, Cakici & Whitelaw, "Maxing Out: Stocks as Lotteries and the Cross-Section of Expected Returns." *Journal of Financial Economics* (2011). NBER w14804. https://www.nber.org/papers/w14804
4. Frazzini & Pedersen, "Betting Against Beta." *Journal of Financial Economics* (2014). http://pages.stern.nyu.edu/~lpederse/papers/BettingAgainstBeta.pdf

5. Arendarski, "Tactical Allocation in Falling Stocks: Combining Momentum and Solvency Ratio Signals." WNE Working Paper No. 1/2012 (67), University of Warsaw (2012). http://www.wne.uw.edu.pl/inf/wyd/WP/WNE_WP67.pdf
6. Rodon Comas, "Winners & Losers in Motion: A Market-State Momentum Signal" (2025). SSRN 5130289. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5130289
7. Chen, "Persistency of the Momentum Effect" (2016). SSRN 2652592. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2652592
8. Heston & Sadka, "Seasonality in the Cross-Section of Expected Stock Returns." *Journal of Financial Economics* (2008). SSRN 687022. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=687022
9. Geertsema & Lu, "Revisiting the Price Effect in US Stocks." SSRN 4013958. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4013958
10. Hou, Xue & Zhang, "Replicating Anomalies." *Review of Financial Studies* (2020).
11. Harvey, Liu & Zhu, "… and the Cross-Section of Expected Returns." *Review of Financial Studies* (2016).
12. McLean & Pontiff, "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance* (2016).