Matched Random Entry as a Control for Technical Trading Rules: Evidence from 533,275 Mechanised Setups

Maksim Aleksandrovich Andreev, Joseph Howden
Quark Quantitative Research

Abstract

A trading rule with a stop at a technical invalidation level and a target at a multiple of the risk taken produces a characteristic return distribution — a hit rate near two in five, wins near +1.5 R and losses near −1.1 R — whether or not the entry carries any information. Evaluating such a rule against zero therefore tests the geometry rather than the signal. We evaluate seven mechanised technical entry families against a matched random entry instead: the same universe, the same number of trades, the same calendar-month distribution, the same stop distance as a fraction of price, the same target multiple, the same position sizing and the same measured transaction costs, with only the entry day randomised. Across 533,275 setups on 7,359 United States equities over 2016–2022, five of the seven families are indistinguishable from that control, at minimum detectable effects of 0.029 to 0.097 R against an empirically plausible effect of 0.13 R. A second control that holds the name fixed and randomises the day within the same month is beaten by the random day in five of the seven families, by margins from 0.111 to 0.592 R; the two families that survive both controls are the two counter-trend ones, at +0.128 and +0.134 R, of which the measured transaction toll consumes about half. Two management rules that raise the hit rate — moving the stop to breakeven at one R, and scaling out half the position there — lower expectancy in all seven families, and a trade-level decomposition of 529,733 resolved trades shows why: the breakeven rule scratches 8.28% of trades that were winners averaging +1.35 R and rescues 9.69% that were losers averaging −1.16 R, netting 0.000 R to three decimal places. Conditioning the families on market regime produced a rule clearing a Bonferroni-corrected within-regime control at $z = 4.89$ in training and 1.48 on an unopened holdout of 897 sessions; every candidate expectancy roughly halved out of sample and so did its matched random control, so the differences contracted rather than reversed, and the only candidate stable across the two windows was the one frozen in advance because it had failed. At the realised search width of 1,538 trials the deflated Sharpe criterion refuses buy-and-hold of the index as well, which makes it a measurement of the width of the search rather than of the merit of a rule. The matched control is the discriminating instrument in this design, and it is the one the technical-rule literature most often omits.

1. Introduction

The empirical record on technical trading rules is a record of controls. Fama and Blume (1966) compared filter rules against buy and hold and found the comparison decisive against the rules once commissions entered. Brock, Lakonishok and LeBaron (1992) reported moving-average and range-break rules that outperformed a set of bootstrapped null models fitted to the index, and Sullivan, Timmermann and White (1999) showed that the same rules, evaluated inside the universe of rules the literature had searched, no longer supported the inference; Lo, Mamaysky and Wang (2000) formalised five chart patterns as kernel-smoothed conditions and found conditional return distributions that differ statistically from unconditional ones while noting the distance between that finding and a tradeable result. The surveys since (Park & Irwin, 2007) treat the choice of null as the methodological centre of the field, and the general form of the problem — that a statistic computed on a rule selected from a large search is not distributed as the same statistic computed on a rule chosen in advance — was set out by Lo and MacKinlay (1990).

The null those studies use is a return series — either a null model of the price process, or the unconditional distribution of the same instrument. A trading rule as it is actually specified is a longer object than a return: it carries an entry, an invalidation level, a target expressed as a multiple of the distance to that level, a position size derived from the same distance, a maximum hold and a management rule, and the return distribution it produces is a joint consequence of the signal and the geometry. Under a two-to-one target and a stop at a technical level, a rule with no information whatever produces winners near +1.5 R, losers near −1.1 R and a hit rate near 0.40, because the target is reached from one side of a random walk and the stop from the other. Those three statistics are the ones practitioners report, and they are the ones a null model of returns does not directly address.

The control used here holds the entire apparatus fixed and randomises only which day the trade was taken. It is a matched randomisation rather than a simulated null: the same universe, the same count of trades, the same calendar-month distribution, the same stop distance as a fraction of price, the same target multiple, the same sizing, the same management and the same per-name transaction costs. Whatever the geometry contributes is present in both arms and cancels in the difference, so the difference isolates the timing information in the entry, which is the quantity the rule claims to supply.

Section 2 states the design and the four mechanical choices that decide the answer. Section 3 defines the two control variants. Section 4 reports the unconditional results and the control differences, with the minimum detectable effect stated per family so that the five nulls can be read as informative rather than merely quiet. Section 5 reports the two management rules whose sign the trade-level record explains exactly. Section 6 decomposes the outcome of every trade into exhaustive labels and reports what fraction of the loss is reachable by any exit rule. Section 7 follows a conditioned rule from a training pass through a sealed holdout and separates the decay of the rule from the decay of its control. Section 8 reports the deflated Sharpe diagnostic at the realised search width and explains why it does not discriminate here. Section 9 gives the limits.

2. Design

2.1 The seven families

Each family is specified mechanically enough to be executed several hundred thousand times, and each carries the stop a discretionary trader would describe rather than a percentage: the level at which the pattern is no longer the pattern.

FAMILY	ENTRY CONDITION, EVALUATED ON THE CLOSE OF THE SIGNAL BAR	INVALIDATION LEVEL
Trend pullback	Close above a rising 200-day average and above the 50-day; the five-day low within 1% of the 21-day exponential average; close within 6% of it; reversal bar	Five-day low less 0.10 average true range
Breakout from compression	Prior twenty-day range narrower than three average true ranges; close above the prior twenty-day high; volume above 1.5 times its twenty-day average; close above the 50-day	Range low less 0.10 average true range, floored at two average true ranges below the close
Support reaction	A confirmed five-bar fractal pivot low, at least six and at most 120 bars old, at or below 1.03 times the close; today's low within 2% of it; close back above it; reversal bar	Pivot low less 0.5 average true range
Structure break	Prior close below the 50-day; the ten-day low higher than it was ten bars earlier; close above the prior twenty-day high; volume above average	Ten-day low less 0.10 average true range
Bullish momentum divergence	Today's low undercuts the low of the window twenty to six bars back while the fourteen-day relative strength index sits more than three points above its own minimum over that window; index below 45; reversal bar	Three-day low less 0.25 average true range
Band reclaim	Prior close below the lower twenty-day two-sigma band; today's close back inside it and above the prior close; close above the 200-day	Three-day low less 0.25 average true range
Volatility expansion	Fourteen-day average true range above 1.5 times its value twenty bars earlier; close at or above the prior twenty-day high; close above the 50-day above the 200-day; volume above average	Close less 1.5 average true range

The indicators are the standard ones (Wilder, 1978; Bollinger, 2001). A moving-average convergence divergence condition was folded into the momentum-divergence family as a quality input rather than run as an eighth family, which is stated because it is a choice about the trial count. Position size is the risk budget divided by the distance from entry to invalidation, whole shares, long only, no leverage, a minimum ticket of fifty dollars, at most eight concurrent positions, a cap of 20% of equity per position, and sale proceeds settled the following session in a cash account.

2.2 Four mechanical choices

Four conventions decide most of the result, and each was fixed in the direction that makes a positive finding harder.

The decision is made on the close of the bar preceding the entry and the fill occurs at the next open. The verification for this requires every past signal to be bit-identical when a name's future is truncated, over the entire overlap rather than a padded interior, because a one-bar look-ahead appears only at the last overlapping index and trimming the edge would make the check unable to fail; the failing control is a detector that reads the following close, and the control does change. The severity of a timestamping error of that kind, measured on a different detector family under the same discipline, is roughly one R per trade (Andreev & Howden, 2026f).

When a bar touches both the stop and the target, the stop is assumed to have been reached first. This is the assumption most capable of manufacturing a false edge if made the other way: on a synthetic case the optimistic rule books +2 R where the conservative rule books −1 R. The ambiguity arises on 0.00% to 0.21% of trades by family, so the headline figures do not depend on it.

R is denominated in the risk planned on the signal close rather than the risk realised at the fill. Dividing by the realised distance from fill to stop appears more precise and is a trap: when the open gaps toward the stop the denominator approaches zero and the arithmetic manufactures outcomes in the thousands of R. Planned risk is also the quantity the position is actually sized on. A gap toward the entry therefore costs more than one R when the trade subsequently fails, and the distribution in Section 4 shows that happening.

A setup whose entry bar opens at or below its own invalidation level is voided rather than traded, on 0.3% to 1.1% of signals by family. Skipping trades that opened badly is a selection, so the rate is reported.

2.3 Universe, window and costs

Every name-day in a survivorship-aware daily bars cache is admitted on the signal bar if the close is at least five dollars, the twenty-day median dollar volume is at least two million dollars, and at least 250 bars of history exist. Delisted names are present and are traded through their delisting. The screen is evaluated on the signal bar. This leaves 7,359 names and 621,539 sampled eligible name-days, which is also the sampling frame for the random control, across 1,762 training sessions covering 2016-01-01 to 2022-12-31. A holdout of 2023-01-01 to 2026-08-01 was sealed by truncation inside the bar loader, verified by a check requiring the loader's maximum returned date to be 2022-12-30 and, as its failing control, requiring the raw file on disk to carry bars into 2026.

Transaction costs are the operator's own measured fills: 28.3 basis points round trip, of which 5.4 are spread and fees and 22.9 are a placement penalty measured against the session close, charged half on entry and half on exit, plus the per-name excess of a point-in-time Corwin and Schultz (2012) high-low spread estimate over the 5.4 basis-point baseline. Two further conventions are reported throughout: a frictionless reading, and a doubled placement penalty at 51.2 basis points. The placement term is 81% of the flat toll and was estimated on five fills, which is stated because it decides the sign of five of the seven families.

3. The matched random control

For each family, a control sample is drawn with the same number of trades, the same distribution over calendar months, the same stop distance as a fraction of the entry price, the same two-to-one target, the same management rule and the same cost treatment, and is resolved by the identical exit engine against the same bars. Eight replications are drawn per family. Confidence intervals on the difference are month-clustered bootstraps, which is the resampling unit the overlap structure requires (Künsch, 1989), since a family fires on one to three hundred names on a given day and those trades share a market path.

> **Two questions, two variants.** The *any-name* control draws both the name and the day at random from the eligible pool in the same month, and answers whether the rule selects anything at all. The *same-name* control holds the name the rule selected and draws a random day within the same month, and answers whether the rule selects the right day inside a name it has already chosen. A rule can pass the first and fail the second, and five of the seven do.

The control is not a null model of the price process and does not attempt to be one. It is a randomisation of the treatment while the design is held fixed, so the estimand is the incremental contribution of the entry timing to a rule whose geometry, universe and costs are otherwise identical. Sullivan, Timmermann and White (1999) address a different failure — the rule was chosen after seeing the data — and their correction, and the reality-check and superior-predictive-ability procedures built on it

(White, 2000; Hansen, 2005), are complementary to a matched control rather than substitutable for it. Section 8 reports the trial correction for the same rules.

4. Results

4.1 Unconditional expectancy

Baseline is the fixed two-to-one target, no quality filter and no minimum structural reward-to-risk requirement, at the 28.3 basis-point toll.

FAMILY	TRADES	MEAN R	95% CI	HIT	AVG WIN R	AVG LOSS R	STOPPED	TARGET	TIMED OUT	HOLD (DAYS)
Trend pullback	250,008	−0.097	[−0.162, −0.032]	0.412	+1.36	−1.12	0.499	0.209	0.292	11.3
Breakout from compression	7,116	−0.057	[−0.141, +0.031]	0.413	+1.27	−0.99	0.456	0.161	0.383	12.7
Support reaction	197,102	−0.151	[−0.210, −0.093]	0.377	+1.57	−1.19	0.564	0.255	0.181	8.7
Structure break	5,833	+0.013	[−0.090, +0.111]	0.496	+0.83	−0.79	0.278	0.074	0.648	16.5
Momentum divergence	41,603	−0.007	[−0.107, +0.109]	0.414	+1.52	−1.09	0.530	0.252	0.218	10.0
Band reclaim	24,097	−0.069	[−0.166, +0.029]	0.398	+1.61	−1.18	0.550	0.278	0.172	8.8
Volatility expansion	3,974	−0.012	[−0.192, +0.166]	0.397	+1.55	−1.04	0.526	0.253	0.221	9.4

Two intervals exclude zero and both exclude it on the negative side. The hit rate, the average win and the average loss vary little across seven rules whose entry conditions have almost nothing in common, which is the geometry asserting itself. The stop is not a clean −1 R: between 4% and 10% of trades by family lose more than 1.5 R, because a level placed under a swing low carries no obligation on the market and the fill is the next open.

4.2 The control differences

FAMILY	SETUP MEAN R	ANY-NAME CONTROL	DIFFERENCE	95% CI	*P*
Trend pullback	−0.097	−0.097	−0.001	[−0.060, +0.049]	0.98
Breakout from compression	−0.057	−0.049	−0.008	[−0.055, +0.037]	0.74
Support reaction	−0.156	−0.165	+0.009	[−0.020, +0.038]	0.55
Structure break	+0.013	+0.053	−0.040	[−0.087, +0.006]	0.10
Momentum divergence	+0.003	−0.125	**+0.128**	[+0.062, +0.209]	<0.001
Band reclaim	−0.069	−0.203	**+0.134**	[+0.063, +0.212]	0.001
Volatility expansion	−0.012	−0.009	−0.003	[−0.105, +0.088]	0.92

The same-name variant is the more severe of the two.

FAMILY	SETUP MEAN R	RANDOM DAY, SAME NAME	DIFFERENCE	95% CI
Trend pullback	−0.097	+0.036	−0.133	[−0.155, −0.111]
Breakout from compression	−0.057	+0.443	−0.500	[−0.547, −0.452]
Support reaction	−0.156	−0.045	−0.111	[−0.134, −0.086]
Structure break	+0.013	+0.327	−0.314	[−0.351, −0.279]
Momentum divergence	+0.003	−0.310	+0.313	[+0.258, +0.385]
Band reclaim	−0.069	−0.431	+0.362	[+0.287, +0.437]
Volatility expansion	−0.012	+0.580	−0.592	[−0.700, −0.498]

Read jointly, the two tables place the five null families precisely. They select names whose behaviour over the month is the universe average, and within those names they select days that are worse than average. The two momentum-continuation families are the extreme cases: buying the names that a compression breakout or a volatility expansion selects, on a random day of the same month, earns +0.443 and +0.580 R, and buying them on the day the condition fires earns −0.057 and −0.012 R. The continuation day is the expensive day, which is consistent with the short-horizon reversal that follows the kind of volume-and-range event these rules require, and inconsistent with the interpretation under which the event is the start of the move. It is also not in tension with intermediate-horizon momentum (Jegadeesh & Titman, 1993), which is measured on formation periods of months rather than on the single session in which a range or volume condition resolves.

The two families that survive both controls are the two counter-trend ones, and they survive by the opposite mechanism. Their matched controls are the weakest month-matched pools in the study, at −0.125 and −0.203 R, because the names are selected while falling; the rule then picks a day inside that population which is materially less bad. That is timing information, it is worth about 0.13 R per trade, and the measured toll on those two families is 0.063 and 0.100 R, so roughly half of what the information is worth is spent acquiring the right to trade on it.

4.3 Power

An unpowered null is absence of evidence presented as evidence of absence, so the minimum detectable effect is stated per family rather than in aggregate. The half-width of the month-clustered interval on the difference is the design's resolution.

FAMILY	MINIMUM DETECTABLE EFFECT (R)	READING
Support reaction	±0.029	Detects a quarter of the surviving effect
Breakout from compression	±0.046	Detects a third
Structure break	±0.046	Detects a third
Trend pullback	±0.054	Detects 40%
Momentum divergence	±0.073	(survived)
Band reclaim	±0.074	(survived)
Volatility expansion	±0.097	Detects 74%; the weakest of the seven

The plausible effect size is taken from this study's own two survivors, +0.13 R, rather than from the literature. Every one of the five null families resolves an effect below that, four of them well below, so an edge of the surviving magnitude would have been visible in any of them. The volatility-expansion null is the weakest and should be read as covering effects above roughly +0.10 R, though its same-name difference of −0.592 R with an interval nowhere near zero is not power-limited under any reading.

4.4 Costs and the sign

Re-scoring every trade individually at four cost conventions reproduces each family's realised mean to zero absolute error at the as-run convention, which establishes that the re-scoring is arithmetic rather than a second model.

CONVENTION	TREND PULLBACK	BREAKOUT	SUPPORT	STRUCTURE	DIVERGENCE	BAND RECLAIM	EXPANSION	POOLED
Frictionless	+0.086	+0.073	+0.074	+0.090	+0.135	+0.139	+0.099	+0.088
Spread and fees only	−0.019	−0.010	−0.058	+0.041	+0.045	+0.014	+0.035	−0.026
As run, 28.3 bp	−0.097	−0.057	−0.151	+0.013	−0.007	−0.069	−0.012	−0.106
Doubled placement, 51.2 bp	−0.175	−0.105	−0.244	−0.016	−0.059	−0.151	−0.059	−0.187

The pooled toll is 0.194 R per trade against a gross expectancy of +0.088 R, and 1.86% of all trades are profitable before costs and negative after. The relation is arithmetic and generalises beyond this study: the cost in R is the round-trip cost as a fraction of price divided by the stop distance as a fraction of price. At a 3% stop, 28.3 basis points is 0.094 R; at a 1% stop it is 0.28 R. The toll measured in the unit the trader keeps score in therefore grows as the stop tightens, and a tight stop is what a small account reaches for, because a tight stop is what permits a position of meaningful size. In this sample the family with the widest stops (9.5% of price) carries the smallest toll (0.036 R) and is the only one with positive after-cost expectancy.

5. Two management rules and the arithmetic of their sign

Six management variants were run on every family: fixed targets at 1.5, two and three times risk; a trailing stop at 2.5 average true ranges; moving the stop to breakeven on a one-R print; and scaling out half the position at one R. Expectancy moved by hundredths of an R and never changed sign in any family. Three regularities hold across all seven.

Wider targets are better in six of seven families, which locates the problem in the losses rather than in truncated winners. The trailing stop raises the stopped share from about 0.53 to about 0.77 and lowers expectancy in six of seven, converting open profit into realised stop-outs. The two variants that most resemble discipline are the two that cost the most: the breakeven stop lowers expectancy in all seven families while dropping the hit rate from about 0.41 to about 0.33, and scaling out half raises the hit rate to about 0.50, the highest of any variant in every family, while lowering the mean in all seven.

Both variants are the mechanised form of a well-documented disposition to realise gains early and defer losses (Shefrin & Statman, 1985; Odean, 1998), and both are ordinarily defended by the hit rate they produce rather than by the expectancy. The trade-level record resolves the breakeven result exactly, because it stores the first bar at which each trade printed one R and the worst low strictly after it. Among all trades, 8.28% printed one R, traded back to the net entry, and finished positive regardless, at an average outcome of +1.35 R; a breakeven stop scratches every one of them. A further 9.69% printed one R, traded back to entry, and finished negative, at an average outcome of −1.16 R; a breakeven stop rescues every one of them.

```
0.0969 × 1.155 − 0.0828 × 1.351 = 0.000 R  (to three decimal places)
```

The rule is a swap at nearly identical average size, and its measured sign is negative only because a scratched ticket still pays its round trip. Its effect on the hit rate is mechanical: converting losses into flat outcomes is what a hit rate counts. The generalisable statement is that any stop-tightening variant should be pre-registered with both its expected scratch rate on winners and its expected save rate on losers, since the sign is decided by the difference of two quantities that are close in this sample and need not be in another.

6. What the trades were

Every one of the 529,733 resolved trades carries exactly one label, assigned by a decision tree in which each trade takes the first branch it matches, so the shares sum to one. The rebuild that produced the labels reproduces every published family mean to six decimal places and passes seven internal consistency checks relating the excursion scan to the exit logic, with zero violations.

OUTCOME	DEFINITION	SHARE
Never worked	Stopped, having never printed +0.5 R at any point	31.97%
Target reached	Reached the two-to-one target	23.22%
Partial win	Finished positive without reaching the target	16.72%
Giveback from a half-R print	Printed at least +0.5 R, finished at or below zero	9.13%
Giveback from a one-R print	Printed at least +1 R, finished at or below zero	9.05%
Gapped stop	The exit bar opened below the stop	6.09%
Drift timeout	Ran out the twenty-day clock at a small loss	2.90%
Toll only	Profitable gross, negative after the round trip	0.92%

Across the pooled book, 55.4% of trades never printed one R, and among losing trades the figure is 83.1%; the average best excursion of a losing trade is +0.39 R. The median trade reaches its own best point on the third bar of an average 10.2-day hold, so this is not a holding-period effect. No exit rule can improve a trade that was never in profit, which places roughly a third of all trades outside the reach of the management layer entirely and returns the question to entry selection.

Entry selection does not answer it. Ranked by the area under the receiver operating characteristic curve at separating a trade that never worked from every other trade, no entry-time feature is informative: the setup's own composite quality score is the best separator at 0.398, an effect in the expected direction and a weak one; the point-in-time spread estimate reaches 0.532; and every trend, momentum, volatility, regime and liquidity feature lies between 0.457 and 0.515. The largest single class of loss in the sample is not screenable from anything available on the setup row at the time the decision is taken.

Two further per-trade quantities constrain the design rather than the rules. Exits through a gap below the stop occur on 6.09% of trades and overshoot the planned loss by an average of 0.51 R, and 43.1% of all trades lose more than one R at an average of −1.35 R, against a sizing rule that treats one R as the worst case. And 53.9% of stopped trades are trading higher five sessions after the exit; twenty sessions after the exit the average stopped trade sits +0.24 R above where it was sold, against +0.19 R after a target exit and +0.13 R after a timeout. That comparison is suggestive rather than conclusive, since conditioning on having traded down to a low will produce a positive forward number in any upward-drifting series; when the index's own drift over the same window is netted out in each trade's R units, the excess is −0.020 R at five sessions and −0.148 R at twenty. There is no post-exit alpha to recover, and holding longer is not the repair.

7. Conditioning, freezing and the sealed holdout

The most common defence of a null result on technical rules is that the test period contained regimes in which the rules are not supposed to be used. The seven families were therefore re-run inside eight definitions of market regime, including a follow-through-day state machine of the kind the practitioner literature specifies (O'Neil, 1988), with the random control re-drawn *inside each cell* rather than inherited from the unconditional run. Four momentum-shaped families inside a confirmed uptrend cover zero against a within-regime random entry under every definition, and the change in expectancy from conditioning lies between −0.023 and +0.005 R. The richest cell in the study is the one the practitioner rule says to stand aside in: with the volatility index at or above 30, six of seven families post their best absolute expectancy, trend pullback at +0.263 R against −0.127 unconditional. The within-regime control removes the whole of it, at +0.118 R with an interval of [−0.103, +0.286]. Selecting the regime did something; selecting the setup inside it did not.

Of 119 cell-level contrasts, twelve survive a Benjamini and Hochberg (1995) correction at 0.05 and seven survive Bonferroni; every positive survivor at either threshold belongs to one of the two counter-trend families, whose advantage is present in uptrend, downtrend, calm and stressed cells alike. The regime filter is not what produces it.

Four candidates were then frozen for a single reading of the sealed holdout, with every parameter written down in advance: two band-reclaim books on different liquidity universes, a two-leg book combining one of them with the momentum-divergence family, and a fourth candidate frozen *because it had failed*, as a falsification instrument. The multiplicity bar was fixed before the freeze at the 800 trials the conditioning search had spent, giving a two-sided z of 4.00. The holdout covers 2023-01-03 to 2026-07-31, 897 sessions, and was opened once, behind nine controls of which seven had to refuse and two are the look-ahead invariant of the production code and a deliberately leaky twin the same check must reject.

CANDIDATE	TRAINING DIFFERENCE	TRAINING *Z*	HOLDOUT DIFFERENCE	HOLDOUT *Z*	HOLDOUT *P*
Band reclaim, low-cost universe	+0.196	3.58	+0.104	1.26	0.214
Band reclaim, liquid universe	+0.215	4.89	+0.095	1.48	0.161
Two-leg book	+0.143	4.19	+0.106	2.86	0.008
Falsification candidate	+0.020	0.65	+0.030	0.87	0.401

The candidate that had cleared the corrected bar at $z = 4.89$ retains 44% of its difference and loses 70% of its z. The two-leg book retains 74% of its difference and is the only one still clearing a nominal 5% threshold, at a z that would have to more than double to meet the correction the same workflow had imposed on itself in advance.

The mechanism is visible in the control column, and it is the reason for reporting a matched control at both ends of a holdout rather than only at the training end. Every candidate's own expectancy fell by roughly half between windows — +0.296 to +0.130 R, +0.247 to +0.082, +0.183 to +0.110 — and so did the matched random entry each is measured against, from +0.100 to +0.026, +0.032 to −0.013, and +0.040 to +0.004. The differences therefore contracted rather than vanished, and a report of the candidate series alone would have attributed to overfitting a decay that the untreated arm shared. The exception is the falsification candidate, whose expectancy moved from +0.180 to +0.171 R against a control moving from +0.159 to +0.141: it is stable across windows because it is measuring the tape. It was frozen with a written prediction that it would make money for the wrong reason, and it did.

In dollar terms on the holdout window, all four books made money, none matched buy-and-hold of the index, and none came within $7.89 a day of it at the larger of the two capital levels. Sharpe ratios of 0.23 to 0.69 sit inside one Lo (2002) standard error of each other — that standard error is 0.53 at 897 observations — and every one of them lies more than one standard error below the index's 1.41. The books carry betas of 0.37 to 0.49 with alphas indistinguishable from zero in both directions, the largest Newey and West (1987) t being 1.04.

8. The deflated Sharpe diagnostic at the realised search width

The deflated Sharpe ratio (Bailey & López de Prado, 2014) adjusts an observed Sharpe for the number of configurations tried, the dispersion of Sharpe ratios across those trials, and the non-normality and length of the return series. The two searches that touched these rules spent 800 and 919 trials with 181 shared between them, giving a union of 1,538, and the observed cross-trial dispersion was 0.0764 per observation. At 897 sessions the criterion demands an annualised Sharpe of 4.09.

SERIES	HOLDOUT SHARPE	THRESHOLD	DEFLATED SHARPE	CLEARS
Band reclaim, low-cost universe	0.69	4.09	0.000	No
Band reclaim, liquid universe	0.23	4.09	0.000	No
Two-leg book	0.66	4.09	0.000	No
Falsification candidate	0.51	4.09	0.000	No
Buy and hold, index fund	1.41	4.09	0.000	No
Buy and hold, technology index fund	1.46	4.09	0.000	No

By the criterion fixed in advance all four candidates are named data-mining artifacts, and so is passive ownership of the index at more than twice the best candidate's Sharpe. No equity strategy of any description produces an annualised 4.09 over three and a half years of daily data, so at 1,538 trials against 897 sessions the criterion is measuring the width of the search rather than the merit of any rule inside it; the candidates' failure is definitional. Four parameterisations were computed — the union and the smaller trial count, each at the measured and at a softer 0.35 annualised dispersion — and the ordering never changes, with the passive benchmarks above every candidate in all of them.

The per-trade version of the same statistic, applied to the R series at the same trial count, produces the sharper illustration. It certifies exactly one candidate at 0.9995, and that candidate is the one frozen in advance as having no edge, whose per-trade return is real, positive and persistent because it belongs to the market rather than to the rule. A deflation statistic evaluated against a null of zero return will certify any positive drift that survives the trial correction. Reported without a matched control it is not the test it appears to be, and the same point applies to the overfitting diagnostics of Bailey, Borwein, López de Prado and Zhu (2014), to the multiple-testing thresholds proposed for the anomaly literature (Harvey, Liu & Zhu, 2016), and to the protocols of López de Prado (2018) and Arnott, Harvey and Markowitz (2019): they bound the damage from searching, and they do not establish that what was found is anything other than exposure.

9. Limits

The evidence is United States equities on daily bars, screened at five dollars and two million dollars of median daily turnover, over 2016–2022 in training and 2023 to mid-2026 in the holdout. Nothing here bears on intraday intervals, on other markets, or on the discretionary practice from which these rules are drawn; a discretionary trader applying the same patterns selects instruments, sizes and moments by criteria not present on the setup row, and the seven mechanisations are a lower bound on that practice rather than a representation of it.

The holdout was opened once per candidate set and both seals are now spent, so every rule named in Section 5 and Section 6 as implied by the trade record — a sizing rule honest about the gap tail, a pre-registered scratch-and-save accounting for stop-tightening variants, a stop wide enough to sit outside the instrument's own noise — is a candidate for forward validation under pre-registration and not a result. None is adopted here.

The two surviving families are reported as beating a matched random entry, which is not the same as being profitable: their absolute expectancies after the measured toll are +0.003 and −0.069 R. Their advantage is roughly twice the toll they pay, so the useful direction of further work on them is the toll — wider stops, longer holds, fewer tickets — rather than further conditions on the entry, which Section 7 measured and found empty. The falsification candidate's expectancy is likewise near zero rather than negative, so nothing here supports an inference that these entries are systematically adverse; the finding is that the five null families supply no timing information beyond the design that surrounds them, and the same-name comparison indicates that two of those five supply timing information of the wrong sign. The costs are those of one account, with a placement term estimated on five fills that constitutes 81% of the flat toll and decides the sign of five families, so the cost convention is reported at four settings throughout rather than assumed at one.

Both windows are single realisations of one market history, so a rule validated on them is validated against a past rather than against a process (Andreev & Howden, 2026e). The matched control narrows that gap by making the comparison internal to each window, which is why the difference is reported as the estimand and the level is not.

Nothing in this paper was traded, no order was placed, no allocation, gate or risk limit was changed on the basis of it, and no position is taken on the merit of any security.

References

Arnott, R., Harvey, C. R. & Markowitz, H. (2019). A backtesting protocol in the era of machine learning. *Journal of Financial Data Science*, 1(1), 64–74.

Bailey, D. H., Borwein, J. M., López de Prado, M. & Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: the effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*, 61(5), 458–471.

Bailey, D. H. & López de Prado, M. (2014). The deflated Sharpe ratio: correcting for selection bias, backtest overfitting and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1), 289–300.

Bollinger, J. (2001). *Bollinger on Bollinger Bands*. New York, NY: McGraw-Hill.

Brock, W., Lakonishok, J. & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5), 1731–1764.

Corwin, S. A. & Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance*, 67(2), 719–760.

Fama, E. F. & Blume, M. E. (1966). Filter rules and stock-market trading. *Journal of Business*, 39(1), 226–241.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4), 365–380.

Harvey, C. R., Liu, Y. & Zhu, H. (2016). …and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.

Jegadeesh, N. & Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.

Lo, A. W. & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3(3), 431–467.

Lo, A. W., Mamaysky, H. & Wang, J. (2000). Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, 55(4), 1705–1765.

López de Prado, M. (2018). *Advances in Financial Machine Learning*. Hoboken, NJ: Wiley.

Newey, W. K. & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 53(5), 1775–1798.

O'Neil, W. J. (1988). *How to Make Money in Stocks*. New York, NY: McGraw-Hill.

Park, C.-H. & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21(4), 786–826.

Shefrin, H. & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: theory and evidence. *Journal of Finance*, 40(3), 777–790.

Sullivan, R., Timmermann, A. & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54(5), 1647–1691.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126.

Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. Greensboro, NC: Trend Research.

Andreev, M. A. & Howden, J. (2026e). Ontological limits of historical validation for wavefunction-calibrated financial systems. *Quark Quantitative Research Working Paper*.

Andreev, M. A. & Howden, J. (2026f). A look-ahead convention in intraday price-structure backtests and the size of the edge it manufactures. *Quark Quantitative Research Working Paper*.