

Out-of-sample tests of retail technical signals on ODTE SPX options

Synapse Research · Oct 4, 2026 · Working paper

Abstract

None of the retail technical signals we tested predicts intraday SPX direction well enough to pay for a same-day (ODTE) option. A 15-minute at-the-money ODTE trade needs 59.0% direction accuracy to break even; our own intraday trend read reaches 52.4% at that horizon ($n = 1,212$ sessions). We tested candlestick and chart patterns (58 patterns, 59,954 occurrences), momentum oscillators (RSI, MACD, stochastics, ADX, CCI), the Strat and ICT setups, eight constructions of auction market theory, dealer second-order greeks, and a 15-signal machine-learned ensemble. Each test fixed its out-of-sample split before it ran (the later ones under pre-registration documents), was computed point in time on 5-minute /ES futures bars, and judged on sessions held out of the fit. The ensemble's out-of-sample AUC was 0.502 over 857 sessions, and its accuracy fell below a constant long position. The one apparent success, an RSI entry gate, turned out to be a side-selection confound: 134 of its 143 fires were puts. We report the designs, the null results, and five pitfalls that produced false positives along the way.

1. Introduction

Same-day S&P 500 options now trade in large size, and most of the education sold to the traders who buy them is built on chart signals: candlestick patterns, momentum oscillators, the Strat, ICT, and auction market theory. These signals are usually shown on hand-picked examples. They are rarely tested point in time, out of sample, against the actual cost of the option.

We ran those tests while building a ODTE SPX trading engine, because each signal was a candidate input. This paper reports them. It makes three contributions:

1. **A break-even bar.** We measure the direction accuracy a short-dated at-the-money trade needs to pay for itself, so every signal has a fixed number to beat.
2. **Null results, judged out of sample.** Each family was specified before it ran and judged on held-out sessions; the later studies carry pre-registration documents. Almost all of them fail, and we report the failures with their designs.

3. Pitfalls. Five mechanisms produced false positives during the work: side-selection confounds, indicator warm-up at a same-day entry, silent data degeneracy, a change in how the data was recorded, and optimistic fills.

The paper leaves out the internals of the strategies we trade. It covers the method and the signals that did not survive, plus two structural filters that did, described only as far as needed to show the contrast.

2. Data

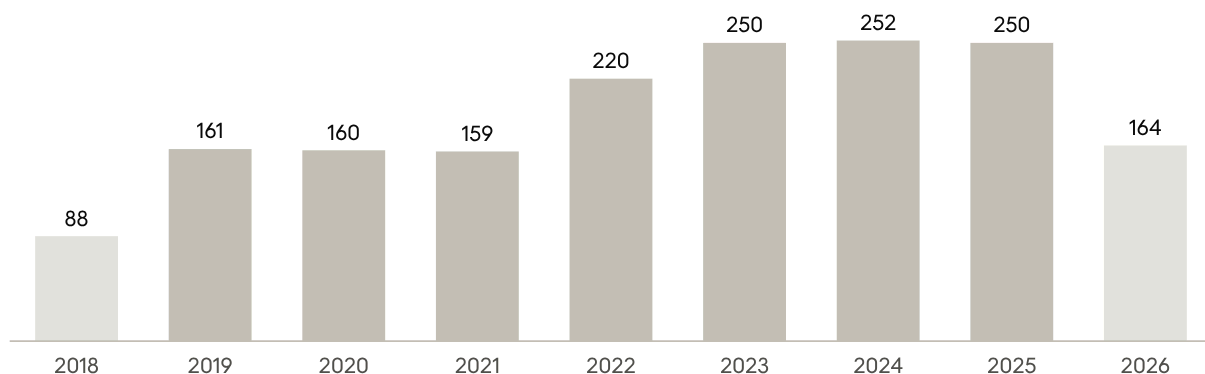
Every test draws on three data sets: intraday SPXW option chains, /ES futures bars, and prior-day open interest.

- **SPXW chains.** 1,704 sessions from June 2018 to October 1, 2026. Each snapshot holds about 100 strikes, calls and puts, bid and ask, every five minutes; since August 2026 the capture runs about once a minute. Coverage is densest from May 2022.
- **/ES futures bars.** 5-minute bars drive every directional read, translated to SPX points through the futures basis. SPX direction is never inferred from SPY.
- **Open interest.** 1,719 sessions of prior-day open interest, the input to the 09:35 pin read.

The archive holds 1,704 sessions of intraday SPXW chains

1,704 sessions counted from the store, by calendar year

Lighter bars are partial years: 2018 starts in June, 2026 runs to October 1



Synapse chain store, counted October 1, 2026 · as published on synapseengine.ai/research

Two gaps shape the tests. Chain snapshots carry no volume or open-interest field, so dealer-positioning tests use the separate open-interest store. And 2021 holds 159 sessions against about 250 in later full years, so a split that leans on 2021 has less data than its date range suggests. Each test in Section 5 states its own window, since the studies ran at different points as the archive grew.

3. Method

Every test follows the same four rules, and Table 1 records where each family stands against them.

1. **Point in time.** A signal is computed only from bars that existed at its decision minute. Nothing is normalized over the full session, and bars are regular-hours 5-minute /ES unless a study states otherwise.
2. **Out of sample.** Each study fixes a date split before evaluation, or walks forward for fitted models. Verdicts come from held-out sessions only.
3. **Uncertainty by session.** Signals that fire several times a day are correlated within the day, so confidence intervals resample whole sessions (day-block bootstrap, 2,000 to 5,000 draws) rather than individual events.
4. **A bar to clear, not just a coin.** Beating 50% is not enough. A directional signal has to approach the break-even accuracy in Section 4, and a filter has to improve the trading result it gates.

We ran many tests. At 95% confidence about one null interval in twenty excludes the null by chance, so isolated exclusions are not claimed as findings. Several studies since September 2026 (chart patterns, dealer greeks, breadth) wrote a pre-registration document before any outcome was joined; the June 2026 studies fixed their splits in advance but did not archive one. We mark the difference rather than blur it.

Table 1. Provenance of each test family

Family	Window and sample	Outcome measured	Pre-registration on file	Result
Chart and candlestick patterns (58)	/ES, 1,249 sessions, 2021-01-04 to 2026-07-08; out of sample from 2024	Direction 30 minutes ahead	Yes, Sept 2026	No interval sits above 50%
Momentum oscillators (RSI, MACD, stochastics, ADX, CCI, Williams %R)	2021 to 2025; splits 2021-23, 2024, 2025	Residual trade P&L, gate P&L	No; splits fixed in advance, June 2026	No tradeable signal

Family	Window and sample	Outcome measured	Pre-registration on file	Result
The Strat, ICT	Direction study June 2026; P&L on 360 trades, 2024-25	Direction; trade P&L	No; June 2026	Coin flips; no P&L as filter or entry
Machine-learned ensemble (15 signals)	Walk-forward, 857 out-of-sample sessions	Direction from 10:00 to the close	No; June 2026	AUC 0.502
Auction market theory (8 constructions)	/ES, Sept 2026	Direction hit rate, P&L	Partly; the magnitude split was pre-registered	48% to 55%, no edge
Dealer vanna, charm, vomma	1,209 sessions, Sept 2026	Added value over an open-interest gate	Yes, Sept 2026	0 of 24 cells pass
Market breadth (ADD, VOLD, TICK)	1,226 sessions, 2021-2026	/ES move over the next 5 to 60 minutes	Yes, hashed, Sept 2026	No lead; /ES leads breadth
Break-even bar	1,212 sessions	Accuracy a 15-minute ATM trade needs	Results document, Sept 2026	59.0% needed

4. The break-even bar

Before judging any signal we fixed the number it has to beat: a 15-minute at-the-money ODTE SPXW trade breaks even at 59.0% direction accuracy, and the bar barely moves with strike.

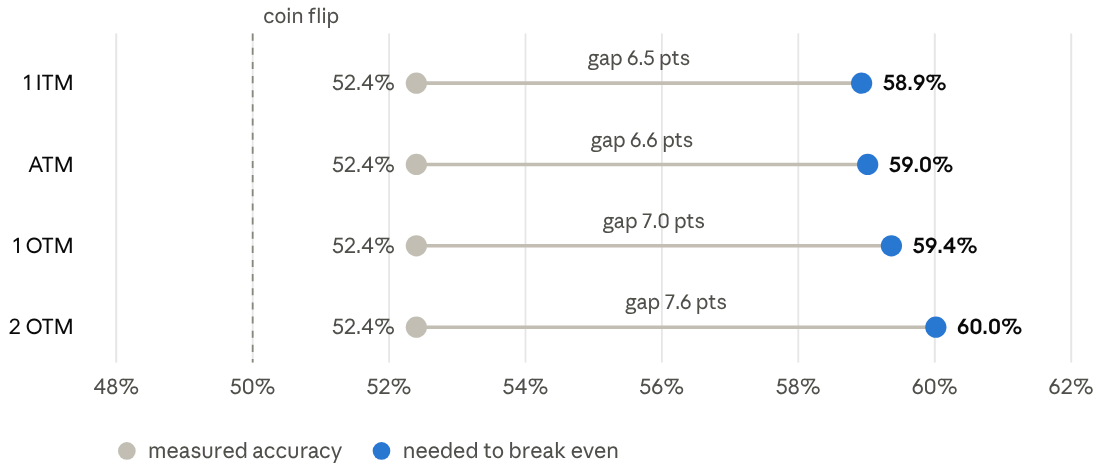
We bought one ODTE SPXW option at the recorded ask at the first signal of an intraday trend read each session, sold it at the recorded bid 15 minutes later, and charged \$1.20 in fees. A call was "right" when /ES moved its way over those 15 minutes. Solving for the accuracy that sets expected P&L to zero, from the realized averages:

$$p^* = |E[\$ \text{ when wrong}]| / (E[\$ \text{ when right}] + |E[\$ \text{ when wrong}]|)$$

At the money a right call averaged +\$296.71 and a wrong one -\$427.27, so $p^* = 59.0\%$. Losses run larger than gains because the trade pays the spread twice and 15 minutes of decay whichever way the market goes.

A 15-minute ODTE trade needs 59% direction accuracy at every strike

Measured direction accuracy against the accuracy that breaks even, one SPXW option, 1,212 sessions



Win-rate and moneyness study, run 2026-09-16 · SPXW ODTE, ask in and bid out, \$1.20 fees, 2021-01-04 to 2026-05-13

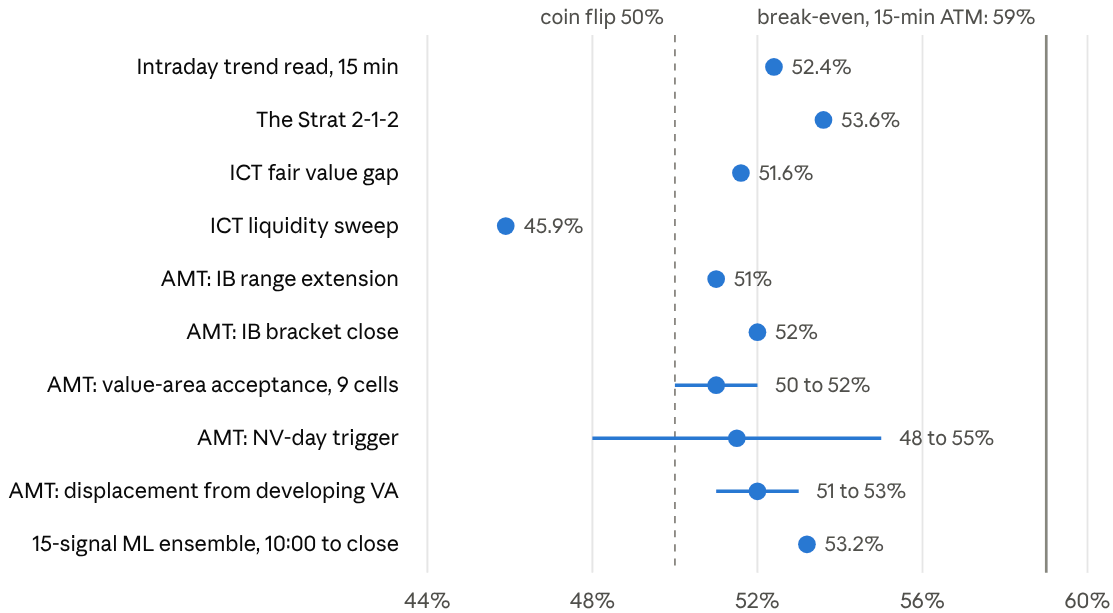
The read itself was right 52.4% of the time, 6.6 points short. Every directional result in Section 5 should be read against this bar, not against 50%. The bar applies to a 15-minute hold; longer holds change the payoff and are outside this paper.

5. Results

No directional family came close to the bar. Every direction read we measured landed between 46% and 55%, and the fitted ensemble did worse than a standing long position.

Every directional read we tested lands between 46% and 55%, short of the 59% bar

Direction hit rate out of sample; whiskers span the cells or day types a study reported



Study results 2026-06 to 2026-09 · horizons differ by study, see Table 1

Horizons differ by study (Table 1), so the 59% line is there for scale; it is exact only for a 15-minute at-the-money hold. Some reads sit at 52% to 55% partly because SPX drifted up over the sample: from 10:00 to the close, guessing "up" every day was right 54.7% of the time in the ensemble's test window.

5.1 Chart and candlestick patterns

We detected 58 patterns causally on 5-minute /ES bars over 1,249 regular sessions, 2021-01-04 to 2026-07-08: bar patterns, reversal geometries, continuation shapes and harmonics, 59,954 directional records in all. A detection scored a hit when the close six bars later agreed with the pattern's textbook direction, given a move of at least 0.10%. In-sample ran through 2023; everything below is out of sample.

No candlestick or chart pattern has an interval above a coin flip

Out-of-sample hit rate 30 minutes ahead, 2024 to mid-2026, with 95% day-block bootstrap intervals

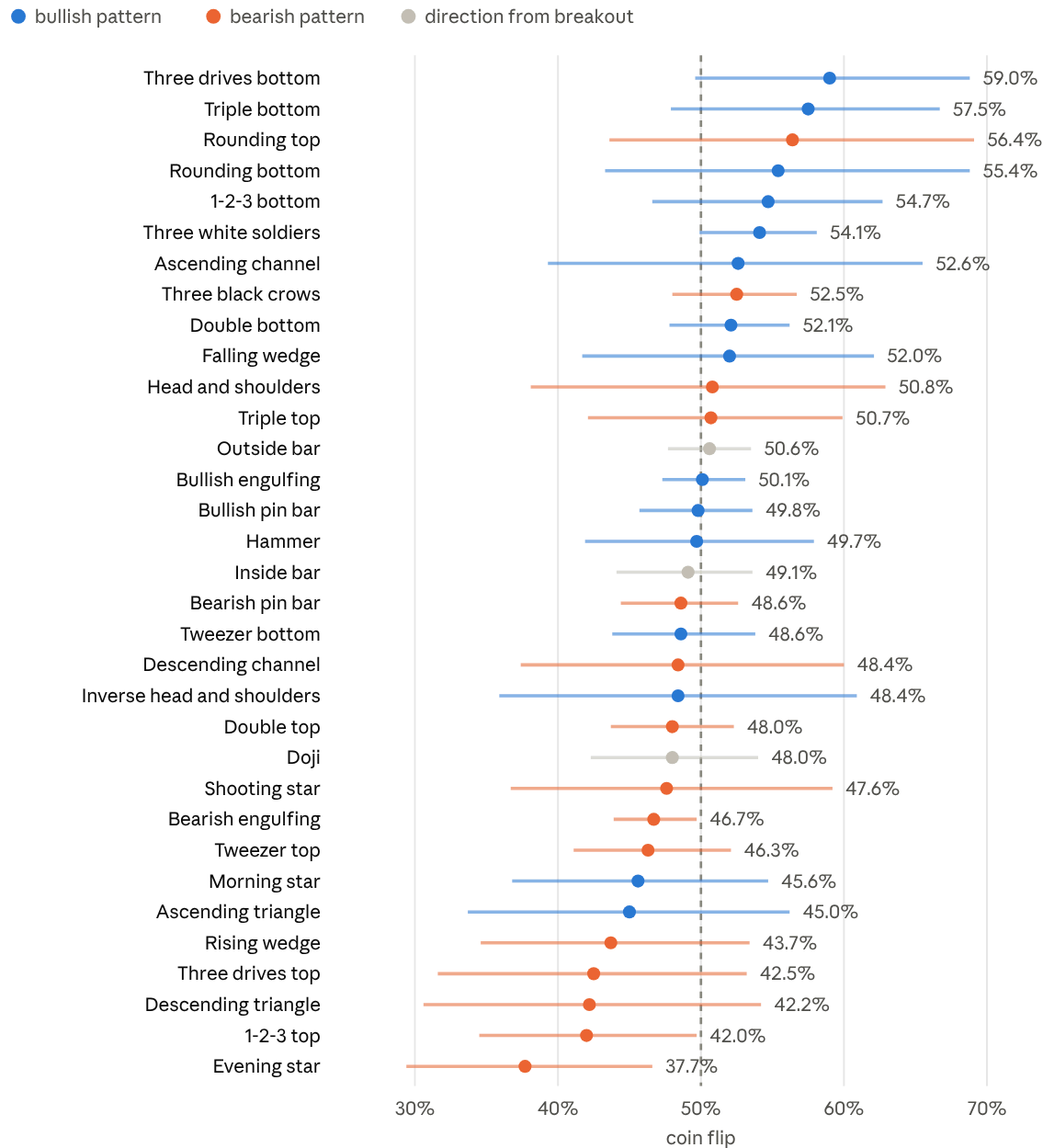


Chart-pattern study, run 2026-09-02 against a pre-registration · 33 patterns with at least 30 graded out-of-sample records

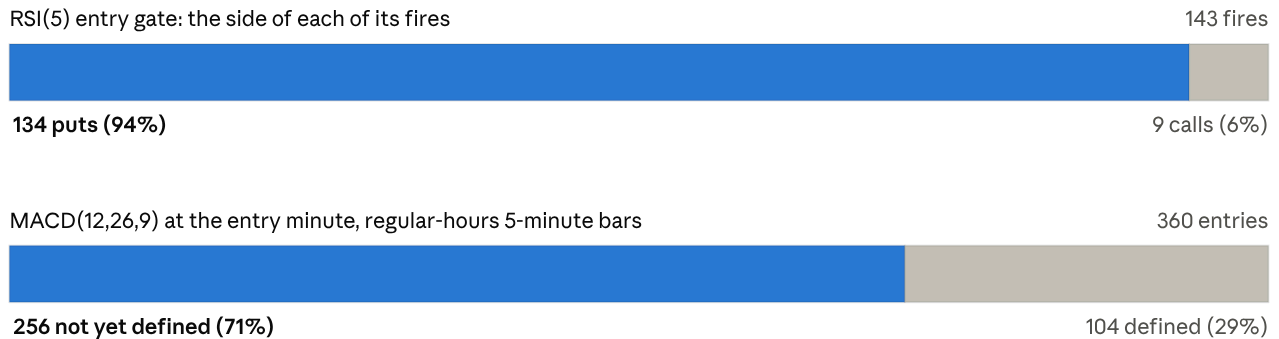
No pattern's interval sits above 50%. The 25 patterns with fewer than 30 graded out-of-sample records, mostly harmonics and flags, are left unscored. Three bearish patterns sit significantly below 50%, evening star lowest at 37.7% (n = 114). That mostly reflects the market's upward drift over the window rather than a tradable fade, and among 33 intervals one or two exclusions are expected by chance.

5.2 Momentum oscillators

We computed RSI, MACD, stochastics, ADX/DMI, chopiness, CCI, Williams %R and RSI divergence point in time on regular-hours 5-minute /ES bars, and joined each to the realized P&L of the trades it would have gated, with splits 2021-23, 2024 and 2025. None added anything beyond momentum and implied volatility. MACD was partly redundant with implied volatility (correlation 0.35), and its confirm-side rule lost money in both test years.

One result looked strong: an RSI(5) "don't chase" gate lifted the backtest's Calmar ratio from 4.4 to 16. Split by side, 134 of its 143 fires selected puts, in a window where puts were the paying side. It was a side selector in disguise, and its timing value within each side was weak.

The RSI gate was a put selector, and MACD rarely exists at a 0DTE entry



Oscillator residual study, June 2026 · /ES 5-minute regular-hours bars, 2021 to 2025

Warm-up is the second problem. Regular-hours 5-minute bars leave about six bars at a 10:00 entry, while MACD(12,26,9) needs about 34 for its signal line. It was defined on only 104 of 360 entries. Textbook swing oscillators mostly do not exist yet at a same-day decision.

5.3 The Strat and ICT

Both were tested twice and failed both times. On direction, the Strat's 2-1-2 reversal was right 53.6% of the time, ICT fair value gaps 51.6%, and ICT liquidity sweeps 45.9% (June 2026).

We then re-tested both on actual trade P&L, using real /ES bars over 2024 and 2025 and 360 trades, once as a veto on an existing entry and once as a standalone entry clock. Neither paid either way. They were not redundant with what we already had: their

correlation with momentum and implied volatility was only 0.07 to 0.19. They carry information about the tape, just not about its next move.

The one signal that survived this track was structural, not directional: an efficiency-ratio filter (net move divided by path length) that stands a trade down on sessions going nowhere, validated on 1,131 days. Neither study archived a pre-registration or its per-cell sample sizes, so we treat these figures as preliminary and plan a re-run under a hashed pre-registration.

5.4 Auction market theory

Auction market theory got the most room: eight independent constructions of a directional signal on /ES, because its claims are specific enough to test. All eight landed between 48% and 55%, and none held up out of sample.

Table 2. Eight directional constructions of auction market theory

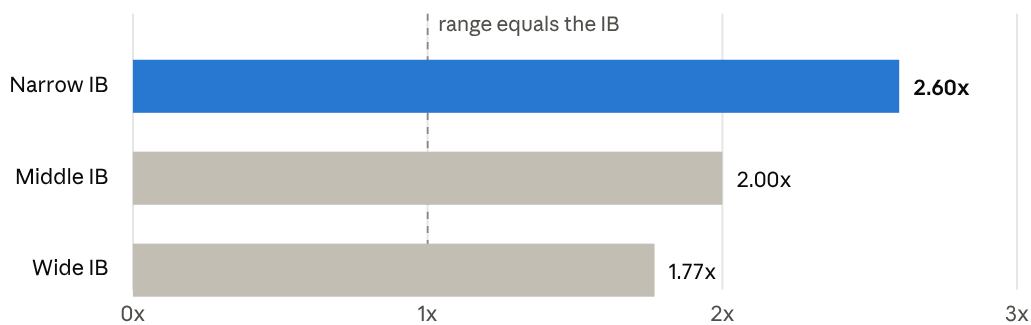
#	Construction	Result
1	Initial-balance (IB) range extension, any trade beyond the IB	51%; negative per trade out of sample
2	IB range extension, a bracket close beyond the IB	52%
3	Follow the first extension, hold to the close	Positive in sample, negative out of sample
4	Follow, exit on a break of the opposite IB edge	Positive in sample, about zero out of sample
5	Follow, reverse on a break of the opposite IB edge	Best shape; the only arm whose halves agree, but its interval spans zero
6	Value-area acceptance, all 9 cells of bars held by bar length	50% to 52%
7	Neutral-day trigger, second bracket breaks the first	55% overall, 48% on the medium-IB days it is meant for
8	Displacement out of the developing value area, barrier test	51% to 53%, against 50% to 51% for random entries

We also closed tails and poor highs, low- and high-volume nodes as price magnets (a 2% effect), and multi-day composite value areas. The composites fired at a median 09:40 at every lookback, which means they were detecting the overnight gap rather than the auction.

What the framework does get right is range. Narrow-IB sessions extended to 2.6 times their initial balance and wide-IB sessions to 1.77 times, with the same ratios in and out of sample under a pre-registered split. Part of this is mechanical, since the IB sits inside the day's range, but the ordering matches the framework's trend-day claim.

AMT reads range, not direction: narrow-IB days extend 2.6 times their IB

Day range divided by initial-balance range, sessions split into terciles by IB width, 1,249 sessions



AMT framework study, September 2026 · narrow and wide terciles identical in and out of sample (2.60 vs 2.60; 1.76 vs 1.77)

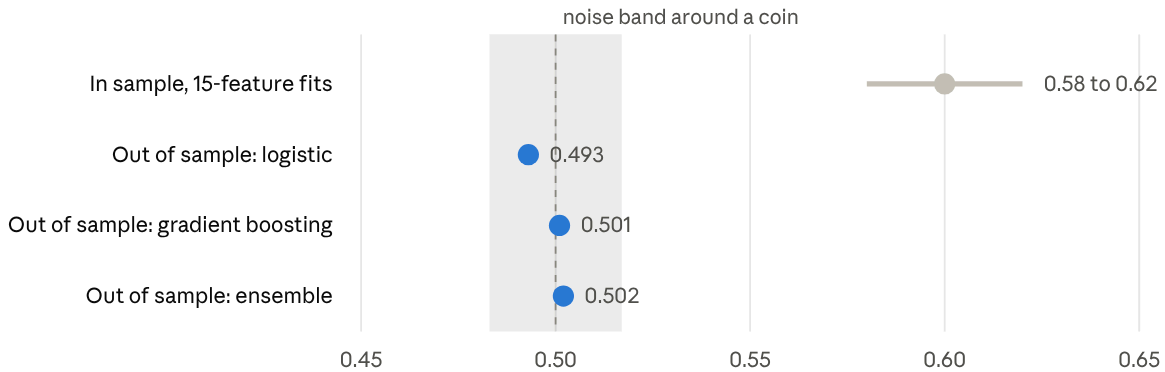
The catch for a same-day trader is timing: the IB completes at 10:30, after most entry decisions are made, so the one reliable AMT read arrives late.

5.5 A machine-learned ensemble

The strongest form of "combine everything" is a fitted classifier, since it searches for the same weightings and interactions a human reader of those signals is looking for. We extracted 15 signals at 10:00 on every session (momentum, acceleration, morning range and volatility, at-the-money implied volatility, skew, net gamma exposure, distance to the open-interest pin, distance to the prior day's high and low, range position, the gap, cumulative volume delta, day of week) and labeled each day by the sign of the move from 10:00 to the close.

The 15-signal model scored 0.58 to 0.62 in sample and 0.50 out of sample

AUC for direction from 10:00 to the close, walk-forward, 857 out-of-sample sessions



Ensemble direction study, June 2026 · noise band ± 0.017 around 0.5

Out of sample, logistic regression scored 0.493, gradient boosting 0.501 and their ensemble 0.502, all inside the noise band. Accuracy was 53.2%, below the 54.7% a constant "up" call earned over the same sessions: the best blend of every signal did worse than being long. The same inputs did predict magnitude (implied volatility separated trend days with an AUC near 0.74), which is the pattern across this paper: these signals describe how far the day will travel, not which way.

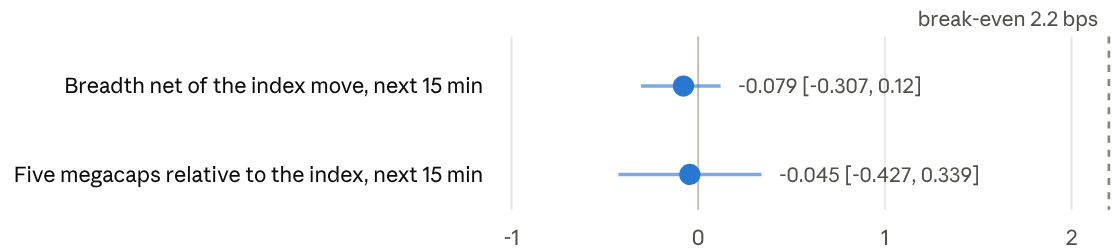
5.6 Dealer greeks and market breadth

Dealer second-order greeks. We tested vanna, charm and vomma exposure at 09:35 as additions to an existing open-interest pin gate, on a 1,209-session trade ledger that reproduced the existing gate to the cent. None of the 24 pre-registered cells passed. Two design errors surfaced and are on the record: the pre-registration read a negative interval as evidence for the inverse gate, which is wrong when the base book is profitable (skipping any profitable days scores negative), and one vomma ratio took only 7 distinct values because vomma is positive almost everywhere, which a distinct-value check caught before it was read.

Market breadth. Advance/decline, up-volume share and a 1-minute TICK proxy, built from the 503 index members, move with /ES over the same window (15-minute correlation 0.66 to 0.85) but say nothing about the next move. The pre-registered cell failed all four of its clauses, and none of 60 secondary cells survived a Benjamini-Hochberg correction. Five megacap names relative to the index failed the same way.

Breadth's effect on the next 15 minutes is about zero; paying would take 2.2 bps

Basis points of /ES per standard deviation of the signal, with 95% day-clustered bootstrap intervals



Breadth round, pre-registered and hashed, 2026-09-30 · breadth: 851 in-sample days 2023 to 2026; megacaps: 211 single-name expiry days

The lead runs the other way. /ES moves first and breadth follows within seconds: a delay of 6.2 seconds for TICK and 7.8 seconds for advance/decline. A breadth signal that seems to lead the cash index is reading stale constituent prints that the future has already priced.

6. Pitfalls that produced false positives

Five times during this work a result looked real and was not. They are general to intraday options backtests, so we list how each one showed up and what caught it.

Table 3. False positives and how they were caught

Pitfall	How it showed up	What caught it
Side-selection confound	An RSI(5) gate lifted a backtest's Calmar ratio from 4.4 to 16	Splitting fires by side: 134 of 143 were puts, in a window where puts paid
Indicator warm-up	MACD(12,26,9) feeding a same-day entry rule	Counting defined values at the decision minute: 104 of 360
Silent data degeneracy	VWAP read 0.00 on every live tick for an unknown stretch, because synthesized bars carried zero volume and every consumer guarded with <code>if vwap > 0</code>	An audit of each signal's live values; the engine now grades its own signal stack every tick

Pitfall	How it showed up	What caught it
Data-era artifact	A next-25-minute direction model scored AUC 0.66 out of sample, above its 0.575 in sample	Splitting by year: the edge sat in 2026 alone (0.81 against about 0.51 in 2024-25), after the archive began recording the underlying price directly instead of inferring it from put-call parity
Optimistic fills	Backtests filled at the mid	Re-pricing every fill at the ask removed 15.6% of a butterfly backtest's total

Two tells recur. An out-of-sample score above the in-sample score is almost never a real model. And a filter that changes which side a strategy trades has to be judged within each side before its timing is credited.

7. Limitations

These results answer a narrow question: does a mechanical rule built on the signal clear the cost of a same-day SPX option, out of sample? They do not show that the signals carry no information at other horizons, in other instruments, or as context for a discretionary trader.

- **One market, one era.** Most tests cover 2021 to 2026, a single rate and volatility cycle with a strong upward drift after 2022. That drift lifts bullish hit rates and depresses bearish ones.
- **Uneven pre-registration.** The June 2026 studies (Strat and ICT, oscillators, the ensemble) fixed their splits in advance but did not archive a pre-registration or their run files. We treat them as preliminary and plan to re-run them under hashed pre-registrations.
- **Modeled execution.** Fills come from recorded bid and ask snapshots taken every five minutes, not from live orders, and the chain snapshots carry no volume or open interest.
- **Many tests.** Across hundreds of cells, some intervals will exclude the null by chance. We claim no isolated exclusion as a finding.
- **Our own read is in the comparison.** The 52.4% trend read in Section 4 is ours, and we have a commercial interest in the topic (Section 9).

8. Conclusion

For a same-day option buyer, direction is the wrong thing to predict from public chart signals. Every family we tested sat within a few points of a coin flip out of sample, against a break-even near 59% for a 15-minute at-the-money trade, and the fitted combination of all of them did worse than a standing long position.

What held up describes magnitude and structure instead: how far a session will travel (initial-balance width, implied volatility) and when to stand aside (an efficiency-ratio filter on sessions going nowhere). That is where we would point the next study, and where the pitfalls in Section 6 are most likely to fool the next researcher.

9. Disclosures and sources

Synapse Research is the research arm of Synapse, which publishes the decisions of a ODTE SPX trading engine to subscribers, so we have a commercial interest in this topic. Backtested results are hypothetical and do not guarantee future results. Nothing in this paper is investment advice.

The study log, including the pre-registration documents where they exist, is at synapseengine.ai/research. Study documents cited here: the chart-pattern results (run 2026-09-02), the win-rate and moneyness study (2026-09-16), the auction market theory framework and direction studies (2026-09-16), dealer second-order greeks (2026-09-19), and the breadth round (2026-09-30), plus internal research notes from June 2026 for the oscillator, Strat/ICT and ensemble studies.