ISSUE 01 · AUDIT REPORT

The Honest Backtest

What we tested. What we kept. What we retired. And why.

> ***6** strategies tested across three rigor passes. **1** deployed live. **5** retired with names and reasons.*

Universe: SPY (NYSE Arca) · Period: ~6 years intraday
Method: TradingView Strategy Tester · 3 rigor passes (v1 → v2 → v3)
Costs in v2/v3: 0.05% commission · 2-tick slippage · 2% sizing
Issued: 2026-05-06 · TradeGladiator

Why this document exists

The standard playbook in retail-trading-signal marketing is to show one chart with a perfect entry, attach a 90% win rate that quietly omits sample size, and point you at a checkout. We've watched enough of those companies disappear to know it works once and corrodes everything after.

The alternative is harder and slower: run the same strategy through realistic costs, count every trade, publish the failures by name, and only deploy what survives. This is that document, for our first audit cycle.

STRATEGIES TESTED	DEPLOYED LIVE	RETIRED, BY NAME
6	1	5
across three rigor passes	Liquidity Sweep Reversal v2	with reasons published

The single number that matters

> ***6 tested. 1 deployed. 5 retired with names and reasons.*** *That ratio is the contract. We expect future audits to look the same — most strategies don't survive realistic costs, and that's the entire point of running the audit.*

Methodology

Every strategy was run through three increasingly strict cost regimes. v1 confirms the code compiles and fires. v2 introduces realistic commission, slippage, and small sizing. v3 layers in bug fixes where v2's audit revealed actual code defects.

Pass	Sizing	Commission	Slippage	Targets	Purpose
v1	10% per trade	0%	0 ticks	Default	Confirm setup logic compiles and fires; surface obvious shape errors.
v2	2% per trade	0.05%	2 ticks	Custom R:R	Apply realistic-cost discipline. Survivors are real candidates.
v3	2% per trade	0.05%	2 ticks	Custom R:R	v2 + targeted bug fixes after a re-read. Tests whether code corrections rescue strategies.

Promote gate

PF ≥ 1.2 across ≥ 30 trades after realistic costs, drawdown ≤ 5% of equity. Below those numbers we do not deploy, even if v1 looked promising.

Results — six strategies, three passes

PF = profit factor (gross profit / gross loss). N = trade count. The green-highlighted row is the only strategy that survived all three passes.

Strategy	TF	v1 PF / N	v2 PF / N	v3 PF / N	Best
LIQUIDITY_SWEEP_REVERSAL	15m	7.12 / 9	**1.21 / 43**	1.07 / 57	v2 ✓
TREND_PULLBACK_EMA	15m	0.93 / 521	0.73 / 261	0.76 / 243	retired
SMC_FVG_CONTINUATION	15m	1.07 / 1237	0.65 / 779	0.59 / 944	retired
SMC_BREAKER_RETEST	15m	33.49 / 3	0.59 / 9	0.18 / 6	retired
OPENING_RANGE_BREAKOUT	5m	0.82 / 101	0.20 / 9	0.31 / 39	retired
MEAN_REVERSION_VWAP	15m	0.02 / 20	0.68 / 3	0.15 / 567	retired

PROMOTED · LIVE IN PRODUCTION

LIQUIDITY_SWEEP_REVERSAL v2

Profit factor **1.21** across **43 trades** on SPY 15m · Drawdown **$1,720** against the 2% sizing equity curve · Win rate **35%** × R:R **2.5** · v3 attempted to improve via tighter stop and adjusted R:R; PF dropped to 1.07. The v2 settings remain the production version.

Production scope: SPY 15m only · 2% sizing · 0.05% commission baked in · Cross-symbol generalization (NQ, ES, individual momentum stocks) is Phase 5 work.

The bug fixes that didn't save anything

A v2 audit pass surfaced three real defects in the underlying code. We fixed all three. v3 measures whether those fixes turned losing strategies into winners.

Strategy	Defect	Fix
Trend_PullbackEMA	**pullbackTouchedBull** and **pullbackTouchedBear** evaluated **identically** despite their directional names — both tested *low<=ema and high>=ema*. The bull/bear asymmetry the strategy claimed to enforce did not exist in the code.	Replaced with proper directional logic: prior bar above EMA + bar low touched EMA + close back above EMA, mirrored for shorts.
SMC_BreakerRetest	OB retry inside the *chochLookback* window read the **current** bar's structure, not the structure that existed at CHoCH time. After CHoCH the structure can flip — the retry would then act on a stale or wrong-direction OB.	Snapshot the CHoCH structure to a **var smc.StructureResult chochStructure** at trigger time and reuse it for retries.
OpeningRange_Breakout	The breakout level check was true on **every bar** that held above the ORB high. With no "entered this break" guard, the strategy re-entered every following bar.	Added **var bool enteredLongThisBreak / enteredShortThisBreak**, reset on isNewDay.

None of the fixes flipped a losing strategy into a winning one. *This is the single most important lesson of the audit. Bug fixes can stop a strategy from over-firing — saving costs — but they can't manufacture an edge that isn't there.*

Retired — five strategies, by name

We do not quietly remove strategies. Each retirement carries a reason, and where Phase 5 may revisit it, the conditions for re-test are stated.

Strategy	Why retired
TREND_PULLBACK_EMA	Pullback bug fix didn't change behavior — the higher-timeframe filter and EMA stack already filtered directionally. 243 trades / PF 0.76 confirms no edge on SPY 15m.
SMC_BREAKER_RETEST	Bug fix exposed lack of edge. CHoCH retest is too rare on SPY 15m to compound — 6–9 setups in 6 years.
SMC_FVG_CONTINUATION	Every R:R variant loses on SPY 15m once costs are realistic. v1's marginal PF 1.07 was a cost-blind illusion.
OPENING_RANGE_BREAKOUT	Re-entry fix recovered the sample but ETF mean-reversion characteristics still bury it. Phase 5 will re-test on NVDA / TSLA before final retire.
MEAN_REVERSION_VWAP	Redesign produced 567 trades at 16% win rate — clean confirmation that the strategy is fundamentally not viable on SPY 15m, regardless of parameters.

Five lessons we're keeping

Generalizable takeaways, written for a future you who is about to test a new strategy.

1 **Realistic-cost backtests are a binary edge filter.**

v1 surfaced what looked like three winners. v2 — the same code with realistic commission, slippage, and 2% sizing — left exactly one. If a strategy's profitability lives or dies at the cost layer, it has no edge to begin with. Always run v2 before you publish numbers.

2 **Bug fixes don't manufacture edge that isn't there.**

Three of the v3 strategies had real defects. We fixed all three. None of the fixes turned a losing strategy into a winning one. A bug fix can stop a strategy from over-firing or trading the wrong direction; it cannot create alpha that the underlying setup doesn't already imply.

3 **Don't over-optimize a working strategy.**

v3 LIQUIDITY_SWEEP tightened the stop and adjusted R:R hoping to improve v2's PF 1.21. Result: 1.07. We reverted. If a strategy is already producing edge with healthy sample size, the next move is out-of-sample testing, not parameter tuning.

4 **Sample size of ≥30 trades is the gate.**

v1 saw "PF 33.49" (three trades) and "PF 7.12" (nine trades) and called both winners. Both collapsed once N grew. We now hold strategies to ≥30 realistic-cost trades before assigning any meaning to the profit factor.

5 **Strategy class determines R:R.**

Mean-reversion strategies that target VWAP-mid (~1× risk) succeed where the same strategy class targeting 2× risk fails. Trend-continuation strategies often want 2× R:R or more. Don't pick R:R as a universal knob — derive it from the empirical winner-distribution of the strategy class you're testing.

What ships, and how

Liquidity Sweep Reversal v2 is now the single strategy in the production webhook funnel for SPY 15m signals. The TradingView library retains v1, v2, and v3 as separate scripts so the audit trail is reproducible. In the TradeGladiator app, signals from this strategy carry the **live** deployment status. Strategies still in testing carry **paper_trade** — they appear in admin observability but do not contaminate live performance metrics.

Phase 5 — what's next

• Walk-forward split on LIQUIDITY_SWEEP — 80% in-sample / 20% out-of-sample to confirm temporal stability rather than period-fit.
• Cross-symbol promote test for OPENING_RANGE_BREAKOUT and TREND_PULLBACK_EMA on momentum stocks (NVDA, TSLA, META) at 5m and 15m.
• Multi-symbol promote criteria: PF ≥ 1.2 AND ≥30 trades AND DD ≤ 5% of equity AND OOS PF ≥ 80% of in-sample.
• Live paper-trade gate before any backend webhook deployment.

We will publish Phase 5 results the same way we published Phase 4-Pine: numbers, decisions, retirements by name.

How to read what we publish

Every signal in your TradeGladiator feed surfaces three fields: **strategy name** (named per this audit), **deployment status** (live or paper_trade — only live counts toward our public win-rate aggregates), and **confidence + R:R** (the numbers used to compute the entry / stop / target ladder for that specific signal).

We don't claim a strategy works until it has cleared the audit. We don't quietly retire it without telling you why. And we don't show you a paper-trade signal as if it were live.

Limitations

Single symbol (SPY). ~6 years of data. Fills modeled at TradingView default + 2 ticks slippage; real broker fills can be worse. Backtests are not future returns. We deploy on backtest evidence + paper-trade gating + live-fire observation — not backtest evidence alone.

Want the next audit when it ships?
Drop your email at **tradegladiator.com/honest-backtest**.