

trend_sma.py

Strategy stress test summary. The full report has every finding, test and figure.

Assets	SPY	Data period	1993-01-29 to 2026-10-02
Development period	2000-01-01 to 2012-12-31	Costs and fills	5 bps per side; filled at the next close
Generated	2026-10-09 01:05 UTC	Engine	Quantgrounds 0.2.0
Code fingerprint	60ad d660 8f0a 878d	Replay code	1/60add6608f0a878d/C/SPY/2000-01-01/2012-12-31/5/43

Verdict

No major weaknesses, but no edge over buy & hold either; 2 tests with warnings.

Diagnostic score

43 / 100

Weak

Weighted mean of six sub-scores. Capped at 43 by the edge over buy & hold (-0.07 Sharpe). Not a forecast.

Sub-score	Weight	Score	Based on
Edge over buy & hold	25%	33	out-of-sample Sharpe 0.97 vs 1.05 held
Risk-adjusted return	15%	78	out-of-sample net Sharpe 0.97
Parameter robustness	15%	92	94% of Sharpe kept (median), 91% at worst $\pm 10\%$
Cost resilience	15%	100	break-even at 26.0x your cost
Generalization	15%	78	100% of held-out assets positive, Sharpe ratio 0.55
Drawdown at equal risk	15%	76	35.2% at equal volatility vs 55.2% held

Stress tests

Stress test	Result	Status
Out-of-sample periods	Sharpe 0.31 in-sample, 0.97 out-of-sample	Pass
Walk-forward	Sharpe 0.65 over 28 one-year tests; 21 of 28 positive	Pass
Parameter sensitivity	Worst $\pm 10\%$ change keeps 91% of Sharpe (sma_days)	Pass
Transaction costs	Break-even at 129.8 bps per side (26.0x baseline)	Pass
Execution delay	Sharpe 0.73 to 0.68 with a one-day delay	Pass
Unseen assets	Positive on 10 of 10 held-out assets; median Sharpe 0.41	Pass
Market regimes	Weakest: high volatility, Sharpe -0.24 over 546 days	Pass
Look-ahead bias	Prevented by design: on_bar() sees past data only	Pass
Resampled histories	Sharpe 90% range 0.47 to 1.00	Pass
Random timing	Beats 79% of 1,000 random timings with the same exposure (after the development period)	Note
Market exposure	Out-of-sample Sharpe 0.97 vs 1.05 held; full history: beta 0.42, alpha 3.9%/yr (t 2.5)	Warning
Research selection	Variants tried not given; selection risk unknown	Not run
Code review	No issues found; deterministic	Pass
Drawdowns and concentration	Max drawdown -24.2%; longest underwater 4.4 yrs	Warning

Key metrics

Metric	Strategy (net of costs)	Buy & hold SPY
CAGR	8.4%	10.8%
Sharpe ratio (90% range)	0.73 (0.47 to 1.00)	0.65
Maximum drawdown	-24.2%	-55.2%
Total trades / yearly turnover	108 / 6.4x	1 / 0.0x

Most important findings

F1 **Medium** Profits depend on a handful of trades

Evidence. Top trade = 18% of net trade P&L; top 5 = 68%; top 10% (11 trades) = 108%. Largest winners entered 1996-07-31, 2020-05-27, 1994-12-15.

Likely cause. Removing a few outlier trades would change the conclusion, so the average trade isn't the story.

Change to try. Check whether those trades reflect a repeatable rule or one-off events, and size positions so no single trade dominates.

F2 **Medium** No edge over buy & hold

Evidence. Out-of-sample Sharpe 0.97 vs 1.05 held (same assets, same days, same costs). It does cut drawdowns: at buy & hold's volatility its worst drawdown is -35.2% against -55.2%. The score is capped at 43.

Likely cause. the trading adds little or nothing over holding the same assets

Change to try. Make sure the strategy delivers its purpose (for example lower drawdowns at equal risk); otherwise simplify.

F3 **Medium** Very long time underwater

Evidence. Longest stretch below a previous equity peak: 4.4 yrs.

Likely cause. Returns arrive in bursts separated by long flat or losing periods.

Change to try. Decide in advance how long you would tolerate being underwater, and compare that with this figure.

F4 **Low** Some evidence of timing skill, not conclusive

Evidence. Over the years after the development period, rotating the strategy's own positions against the market 1,000 times (same exposure, holding periods and costs; only the timing changes) gives a median Sharpe of 0.81; 95% of rotations stay below 0.97. The strategy's Sharpe of 0.89 beats 79% of them.

Likely cause. random timing would beat this often enough that it could be luck

Change to try. Treat the timing as unproven until it holds on data you haven't used.

What held up

- + Out-of-sample Sharpe (0.97) held up relative to in-sample (0.31).
- + Parameters are robust: 94% of Sharpe kept (median), 91% at worst $\pm 10\%$.
- + Cost resilient: break-even at 129.8 bps per side (26.0x baseline).
- + Tolerates execution delay: a one-day delay keeps 92% of the Sharpe ratio (0.73 to 0.68).
- + Generalizes: profitable on 100% of 10 held-out assets, but its Sharpe beat buy & hold's on only 20% of them.
- + Positive alpha after market exposure: 3.9% a year ($t = 2.5$) at a beta of 0.42.
- + Walk-forward holds up: Sharpe 0.65 when parameters are re-chosen from past data only (21 of 28 test years positive).

Educational diagnostics only; not investment advice. Results are simulated on historical data and do not predict future returns. The score summarizes this report's tests; it is not a measure of trading skill or future profitability.