

lookahead_bug.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	298d 9ece 279b f40e	Replay code	1/298d9ece279bf40e/C/SPY/2000-01-01/2012-12-31/5/20

Verdict

Not tradeable as written: the backtest uses information from the future.

Diagnostic score

20 / 100

Serious issues

Weighted mean of six sub-scores. Capped at 20 because look-ahead was confirmed. Not a forecast.

Capped at 20 because look-ahead was confirmed (uncapped 98).

Sub-score	Weight	Score	Based on
Edge over buy & hold	25%	100	out-of-sample Sharpe 2.61 vs 1.05 held
Risk-adjusted return	15%	100	out-of-sample net Sharpe 2.61
Parameter robustness	15%	94	96% of Sharpe kept (median), 92% at worst $\pm 10\%$
Cost resilience	15%	100	break-even at 22.8x your cost
Generalization	15%	95	100% of held-out assets positive, Sharpe ratio 0.89
Drawdown at equal risk	15%	100	14.8% at equal volatility vs 55.2% held

Stress tests

Stress test	Result	Status
Out-of-sample periods	Sharpe 1.88 in-sample, 2.61 out-of-sample	Not valid
Walk-forward	Sharpe 2.39 over 29 one-year tests; 29 of 29 positive	Not valid
Parameter sensitivity	Worst $\pm 10\%$ change keeps 92% of Sharpe (smooth_days)	Not valid
Transaction costs	Break-even at 114.0 bps per side (22.8x baseline)	Not valid
Execution delay	Sharpe 2.27 to 2.26 with a one-day delay	Not valid
Unseen assets	Positive on 10 of 10 held-out assets; median Sharpe 2.02	Not valid
Market regimes	Weakest: bear / crash, Sharpe 0.72 over 577 days	Not valid
Look-ahead bias	Confirmed: past values of 'slope' change when future data is removed	Fail
Resampled histories	Sharpe 90% range 2.05 to 2.49	Not valid
Random timing	Beats 100% of 1,000 random timings with the same exposure (after the development period)	Not valid
Market exposure	Out-of-sample Sharpe 2.61 vs 1.05 held; full history: beta 0.42, alpha 22.1%/yr (t 15.4)	Not valid
Research selection	Variants tried not given; selection risk unknown	Not valid
Code review	No other issues (1 look-ahead pattern, see look-ahead bias); deterministic	Pass
Drawdowns and concentration	Max drawdown -9.5%; longest underwater 0.7 yrs	Not valid

Key metrics

Metric	Strategy (net of costs)	Buy & hold SPY
CAGR	30.3%	10.8%
Sharpe ratio (90% range)	2.27 (2.05 to 2.49)	0.65
Maximum drawdown	-9.5%	-55.2%
Total trades / yearly turnover	406 / 24.1x	1 / 0.0x

Most important findings

F1 **Critical** Look-ahead bias confirmed: the backtest uses future data · line 15

Evidence. Truncation test: when data after 2003-03-05 was removed, indicator 'slope' changed on 9 of 2544 earlier rows (reaching 8 bars back). 4 of 4 checks failed. Backtest as written: Sharpe 2.27, CAGR 30.3%. Recomputed point-in-time (what you could actually have traded): Sharpe 0.35, CAGR 3.5%, max drawdown -41.9%, in the market 60% of days.

Likely cause. line 15: `smooth = prices.rolling(params["smooth_days"], center=True, min_periods=1).mean()` (centered rolling window uses future bars)

Change to try. Remove `center=True`; trailing windows (the default) only use past data. Until fixed, every other number in this report is contaminated, and the score is capped.

What held up

None reported: with confirmed look-ahead, every performance-based strength would be an artifact.

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.