QAUDIT

Independent audit of trading systems - deterministic engine 1.0.0

VERDICT

GO

blocking 0 major 0 minor 0 info 6

qaudit audit report

temoin-sain | Audit complet

PERIOD COVERED	USABLE / SUPPLIED TRADES	INITIAL BALANCE	NUMBER OF TRIALS USED (N)
2021-01-11 -> 2024-02-24	**600 / 600**	**10 000.00 USD**	**10**

Header

HOW TO READ THIS SECTION

The audit's identity card. The SHA-256 fingerprint is a unique signature of the trade file analysed: change a single character in the file and the fingerprint changes. The seed and the versions make it possible to reproduce exactly the same report. "N" is the number of variants of the strategy tried before this one was kept, declared by the requester or set by default by the method.

Audited system	**temoin-sain**
Engine version	**1.0.0**
Method version	**1.0**
Seed	**0**
SHA-256 fingerprint of the data	**d4f73aca2abb4c9fc7d98e36556914619a30fd503e57dbe4ac6471c2fb446d5f**
Monte Carlo iterations	**10 000**
Number of trials used (N)	**10**
Period covered	**2021-01-11T16:40:28 -> 2024-02-24T04:29:18**
Usable / supplied trades	**600 / 600**
Currency	**USD**
Initial balance	**10 000.00**
Service	**Audit complet (audit_complet)**

Verdict

HOW TO READ THIS SECTION

Three possible verdicts: GO (green light), GO CONDITIONAL (amber: checks are missing), NO-GO (red: a test failed or a blocking flaw exists). The verdict follows from thresholds fixed in advance, never from an appraisal. Findings are ranked by severity: blocking, major, minor, info.

Verdict: GO.

Findings: 0 blocking, 0 major, 0 minor, 6 informational.

In plain words: what this report says

Green light. The results provided passed all five statistical tests and no serious flaw was found in the way the trading is

done. This means the observed performance is neither explained by chance, nor carried by a few lucky trades, nor manufactured by questionable data - over the period and data provided.

The verdict covers only what was given to the auditor: the closed trades in the file, over the period covered. It rests on no human judgement and no overall score: every test has a threshold fixed in advance by the method, and the verdict follows mechanically from those thresholds. The same file, audited again tomorrow with the same version, yields exactly the same report.

The five questions put to the results, and the numeric answer for this file:

PASSED `Is chance ruled out?`
When several versions of a strategy are tried before one is kept, the best one always looks good, even if none has an edge. This test computes the probability that the performance is real once that effect is corrected, for 10 declared variant(s): 97.6 %. Pass threshold: 50.0 %. Test passed.

PASSED `And with bad luck?`
The profit factor is the total won divided by the total lost (1.00 = break-even). The same trades were replayed 10 000 times in a different order, in blocks, to see what the same strategy would yield with less luck: in the 5 % least favourable scenarios the profit factor falls to 1.187. Threshold: 1.00 - below it, the strategy loses money in those scenarios. Test passed.

PASSED `Do the losses come clustered?`
The drawdown is the largest fall between a peak and the trough that follows it. The observed one (2 042.87) is larger than in 78.1 % of random reorderings of the same trades. A very low value (under 5 %) would signal a curve too smooth to be honest; a very high value means the losses came clustered, which is harder to live through than the final result suggests. Test passed.

PASSED `Would a system with no edge do as well?`
10 000 systems with no edge at all, but winning and losing amounts of the same size as this one, were simulated; 0.0 % of them did as well or better. Threshold: 5.0 % - above it, this system cannot be told apart from a system with no edge. Test passed.

PASSED `Does it hold outside the periods it was tuned on?`
A strategy tuned on one period tends to shine there and disappoint afterwards. This test compares its performance on periods not used for tuning: it keeps 65.4 % of it. Threshold: 40.0 %. Test passed.

No serious flaw was found in the way the trading is done: no martingale, no grid, no abnormal concentration of profit, no sign of tampered data.

What remains true despite the green light: a statistical test measures the past that was provided, not the future. The drawdown figures in the "Benchmarks" section show the size of temporary losses it is reasonable to expect from this system as it was observed.

This report states facts; it recommends no settings and predicts nothing.

Key indicators

HOW TO READ THIS SECTION
Net profit: gains minus losses, commissions and swaps included. Profit factor: total won ÷ total lost (1.00 = break-even). Hit rate: share of winning trades. Win/loss ratio: average gain of a winning trade ÷ average loss of a losing trade. Expectancy: average gain per trade, all trades included. Sharpe: average gain relative to the variability of results (the higher, the steadier the gains). Maximum drawdown: largest fall between a peak and the following trough. Losing streak: longest run of consecutive losses.

Net profit	**12 751.41**
Profit factor	**1.380**
Hit rate	**46.2 %**
Average win / average loss	**1.609**

Expectancy per trade	**21.25**
Sharpe per trade	**0.143**
Annualised Sharpe (measured frequency)	**1.98**
Trades per year (measured)	**192.5**
Maximum drawdown	**2 042.87**
Maximum drawdown (% of equity peak)	**12.9 %**
Compound annual growth rate	**30.2 %**
Longest losing streak	**9**

Decision barriers

HOW TO READ THIS SECTION
Each row is a test with a threshold fixed by the method. "Passed" = succeeded; "triggered" = failed (in capitals, because one is enough for a NO-GO); "indecision band" = too close to the threshold for an automatic decision, an auditor must rule; "not evaluated" = missing data. There is no overall score: a single failed test makes the verdict.

The verdict follows mechanically from the barriers below; no composite score.

BARRIER	VALUE	NO-GO THRESHOLD	INDECISION BAND	STATUS
Deflated Sharpe ratio (DSR)	0.976	< 0.500	[0.450 ; 0.550]	**passed**
Profit factor, 5th percentile of the block bootstrap	1.187	< 1.000	[0.950 ; 1.050]	**passed**
Percentile of the observed drawdown among order permutations	78.140	< 5.000	[3.000 ; 8.000]	**passed**
p-value against the no-edge null	0.000	>= 0.050	[0.030 ; 0.080]	**passed**
Out-of-sample walk-forward efficiency	0.654	< 0.400	[0.350 ; 0.500]	**passed**

Critical findings

HOW TO READ THIS SECTION
The flaws serious enough to weigh on the verdict, with the figure that establishes them. One blocking finding is enough for a NO-GO; several major findings lead at best to a GO CONDITIONAL. These findings appear in every version of the report, whatever the service ordered.

No finding at this level.

All findings

HOW TO READ THIS SECTION
Everything the auditor noted, including minor remarks and information items (for instance a test that could not run). An "info" line is not a flaw: it says what was done, or what could not be done and why.

INFO `dsr_lo_variance`
Variance of Sharpe ratios across variants unknown: Lo approximation (V = 0.0017) for N = 10 trials. Conservative correction: the true DSR is probably lower.

INFO `execution_realism_skipped`
Detector "execution_realism" not run: no price series provided.

INFO `lookahead_skipped`
Detector "lookahead" not run: no price series provided.

INFO `missing_optional`
Fields absent from the source: commission, swap. Analyses that depend on them are disabled and reported as such.

INFO `null_empirical`
The "no-edge" null keeps the empirical distributions of gains and losses and only neutralises their frequency (p0 = 0.3832). It answers: "is the observed hit rate explainable by luck, given this gain/loss profile?" - it is not an efficient-market null.

INFO `stop_geometry_skipped`
Detector "stop_geometry" not run: no price series provided.

Probabilistic and deflated Sharpe

HOW TO READ THIS SECTION
The Sharpe ratio measures the steadiness of gains. The PSR is the probability that the true Sharpe is positive given the number of trades and the shape of the results. The DSR answers the harder question: does this performance exceed what chance would have given to the best of N attempts? If the requester tried more variants than declared, the true DSR is lower than the one shown.

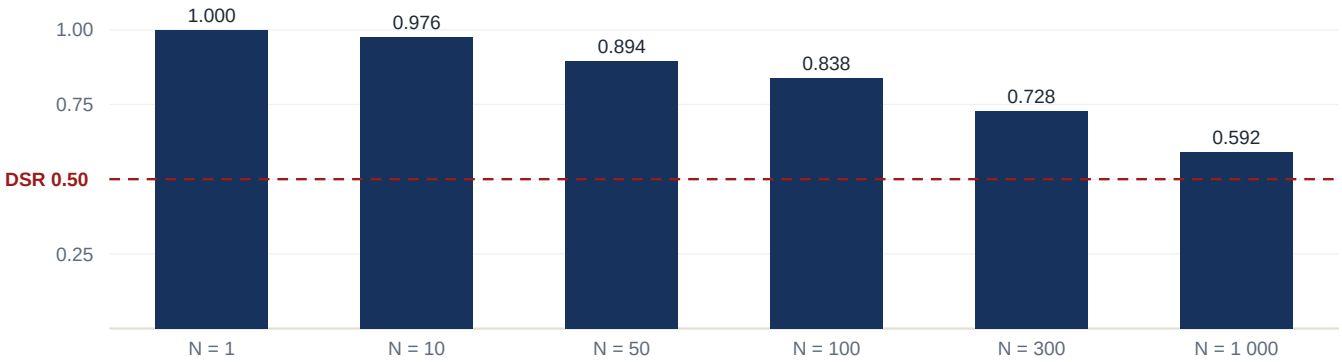
PSR: probability that the true Sharpe is positive, corrected for skewness and fat tails. DSR: the same probability evaluated against the best Sharpe expected by chance after N trials.

Probabilistic Sharpe ratio (PSR)	**0.9998**
Deflated Sharpe ratio (DSR)	**0.9760**
Expected maximum Sharpe under the null (SR0)	**0.0646**
Number of trials used (N)	**10**
Variance of Sharpe ratios across variants (V)	**0.001684**
Source of V	**Lo approximation (conservative: the true DSR is probably lower)**

Sensitivity to the number of trials

HOW TO READ THIS SECTION
The same DSR recomputed for different numbers of variants tried. Read the row that matches your actual practice: if you tested 50 settings before keeping this one, the N = 50 row is the one that counts. The chart shows from which number of attempts the performance stops being distinguishable from chance (red line).

This system stays significant up to at least 1 000 tested variants (grid bound).



N TRIALS	DSR
1	0.9998

N TRIALS	DSR
10	0.9760
50	0.8942
100	0.8377
300	0.7279
1 000	0.5923

Resampling

> **HOW TO READ THIS SECTION**
> Three simulations on the same trades, for three questions. Bootstrap: with less luck, what profit and what drawdown? (the p05 / p50 / p95 ranges bracket 90 % of scenarios). Permutation: is the observed drawdown due to the order of the trades? No-edge null: how many systems with no edge at all, but the same sizes of gains and losses, would do as well?

Three simulations, three questions: the bootstrap measures how unlucky the same edge can get; the permutation tests whether performance owes to trade order; the null counts edge-less systems that would do as well.

The "no-edge" null keeps the empirical distributions of gains and losses and only neutralises their frequency (p0 = 0.3832). It answers: "is the observed hit rate explainable by luck, given this gain/loss profile?" - it is not an efficient-market null.

Bootstrap method	**stationary blocks (Politis-Romano), automatic length**
Mean block length	**1.00**
Profit factor p05 / p50 / p95	**1.187 / 1.378 / 1.596**
Net profit p05 / p50 / p95	**6 679.99 / 12 692.53 / 18 690.96**
Drawdown p50 / p95 / p99	**1 706.73 / 2 897.87 / 3 678.94**
Observed drawdown	**2 042.87**
Percentile of the observed drawdown among permutations	**78.1**
Permutation drawdown p50 / p95	**1 670.50 / 2 608.53**
Equilibrium hit rate p0	**38.3 %**
p-value (net profit >= observed)	**0.0005**
Net profit of null systems p50 / p95	**-103.24 / 5 835.66**

Regimes and periods

> **HOW TO READ THIS SECTION**
> Trades are sorted into three tiers according to how agitated the market was when they took place (low, medium, high volatility), then by quarter. A healthy strategy earns across several regimes and several quarters; if all the profit comes from a single tier or a single period, the overall result is misleading. Rows in red are losing.

Classifier used: volatility terciles (rolling standard deviation of P&L, no prices); unclassified trades: 2.

REGIME	TRADES	NET PROFIT	PF	HIT RATE	SHARE OF NET
low volatility	200	6 002.33	1.523	47.0 %	47.1 %
medium volatility	199	2 065.67	1.188	44.2 %	16.2 %
high volatility	199	4 849.49	1.446	47.7 %	38.0 %

QUARTER	TRADES	NET PROFIT	PF
2021-Q1	31	1 840.05	2.473
2021-Q2	52	1 132.66	1.399
2021-Q3	47	1 388.94	1.605
2021-Q4	49	826.58	1.275
2022-Q1	51	-1 007.03	0.712
2022-Q2	48	2 181.53	1.880
2022-Q3	45	1 269.35	1.569
2022-Q4	62	1 066.97	1.331
2023-Q1	49	684.72	1.234
2023-Q2	38	-647.49	0.754
2023-Q3	60	1 302.22	1.350
2023-Q4	42	1 886.74	1.931
2024-Q1	26	826.17	1.579

Drawdown markers

HOW TO READ THIS SECTION
Benchmarks from the simulations: the drawdown exceeded in only 5 % (p95) and 1 % (p99) of scenarios. They serve to know what to expect, and to recognise the day the system no longer behaves as in its history. They are not recommended settings.

Statistical drawdown markers from the block bootstrap, to compare with the live drawdown. They are not strategy settings.

Observed maximum drawdown	**2 042.87**
Bootstrap drawdown, 95th percentile	**2 897.87**
Bootstrap drawdown, 99th percentile	**3 678.94**

Conditions for production

HOW TO READ THIS SECTION
What would have to be provided or fixed for the audit to be complete. Each condition points to a precise test that could not be run or that stayed in its uncertainty zone.

No condition.

Scope and limits

HOW TO READ THIS SECTION
What the audit could not see, and what it does not claim to do. These limits are part of the result: a report that did not state them would be misleading.

- Only closed trades are analysed: floating P&L and open positions are unknown.
- No price series: stop sensitivity, look-ahead and execution realism untested.
- The walk-forward relies on results supplied by the client, never replayed.
- No prop-firm ruleset: no replay.
- The "no-edge" null keeps the empirical distributions: it is not an efficient-market null.
- The auditor observes; it recommends no parameter setting.
- The variance of Sharpe ratios across variants is unknown: Lo correction, conservative.

Footer

> **HOW TO READ THIS SECTION**
> The identifiers that make this report verifiable: engine version, method version (the thresholds), seed, file fingerprint, service. Anyone holding the same file and the same version obtains the same report, byte for byte.

qaudit 1.0.0 - method 1.0 - schema 1.0 - seed 0 - fingerprint d4f73aca2abb4c9fc7d98e36556914619a30fd503e57dbe4ac6471c2fb446d5f - service audit_complet.

Reproducible: same inputs + same version + same seed = same report, byte for byte.