



How to audit an MT5 backtest

Subject	An MT5 Strategy Tester HTML report for an Expert Advisor (EA)
Method	Five questions, a 16-item checklist you can apply by hand, a field map of the report, and a prop-firm rule replay
Reader	Anyone buying an EA, or validating their own before putting a fee or capital behind it
Does not do	Predict future results, estimate returns, or tell you whether a challenge will be passed
Version	1.0, updated 2026-10-15
Changelog	1.0 (2026-10-15): first public release.
Licence	CC BY 4.0. Share freely, attribution appreciated.

§1

Premise

1. A backtest report is a claim, not evidence

An MT5 report is a claim about the past, produced by a program whose settings you cannot see. The inputs, tick data, spread, commission and date range were chosen by whoever ran it. The report prints the result and hides the choices. It can tell you what the EA would have done on that data, under those settings, and nothing about whether either was realistic.

Authenticity is a separate problem. An HTML report is a file, and files can be edited. Nothing below proves a report is genuine, and VerifyEA does not verify authenticity either. See §5.6.

§2

The five questions

2. The five questions to ask of any report

Q1 Is there an out-of-sample segment at all?

One report, one date range, and a long list of tuned inputs cannot support a stability claim.

Q2 What are the modelling quality and the tick data?

Poor data, assumed costs or a short history invalidate the run before any metric matters.

Q3 Is the profit carried by a handful of trades?

Read the distribution and the trade count, not the total.

Q4 Is the equity curve a straight line?

Four structural fingerprints: grid recovery, time clustering, identical sequences, lot-size progression.

Q5 Does it survive reordering?

Reshuffle the same trades and look at the drawdown distribution, not one lucky sequence.

§3

Checklist

3. The checklist

Apply it by hand to any report. A box ticked means you have seen the evidence yourself, not that the seller said so. Any box you cannot tick is an open question, and should stay open until answered.

C1 to C16

Report and EA name

Reviewed on (date)

Q1. Out-of-sample (§5.1)

☐ **C1** **An OOS report exists**, covers dates the optimisation never used, and is compared to the in-sample result on the same measures, not just "also profitable".

☐ **C2** **The number of input parameters and optimisation runs is disclosed**. Many parameters tuned over many runs means a good-looking result may be chance.

Q2. Data quality (§5.2)

☐ **C3** **Modelling quality is stated** and the model is "Every tick based on real ticks", not an OHLC-only or generated-tick run.

☐ **C4** **The tick history source is named**, and the bars, ticks and period cover trending and ranging markets.

☐ **C5** **Spread is variable or realistic and was set on purpose**, and commission and swap match the broker you will use. None is left at a default.

☐ **C6** **The trade count is large enough** that the average trade, and any claim built on it, means something.

Q3. Concentration (§5.3)

☐ **C7** **Remove the five largest winners**. The result is still positive and still reasonable.

Q4. Equity curve and structure (§5.4)

- ☐ **C8** **No grid recovery:** lot size does not jump (about 2x) after a loss, repeatedly.
- ☐ **C9** **No lot-size escalation** across runs of consecutive trades.
- ☐ **C10** **No time clustering:** entries are not concentrated in one narrow window of the day.
- ☐ **C11** **No identical sequences:** the same symbol, direction, lots, and time offset do not repeat three or more times.

Q5. Reordering (§5.5)

- ☐ **C12** **Reshuffled drawdown has been checked** (Monte Carlo on closed trades). You looked at the deep percentiles, not the median, and you know the seed.
- ☐ **C13** **Worst plausible drawdown fits your risk limit**, or the firm's loss limit if you are going through a challenge.

Provenance (§5.6)

- ☐ **C14** **You hold the raw report file**, not a screenshot. It looks like a standard MT5 report, its totals agree with its own trade list, and you can reproduce it or trust the source.

Prop-firm rules (§6)

- ☐ **C15** **Rules are taken from the firm's current page** and dated. Any inconsistent value is recorded as unknown.
- ☐ **C16** **Trade list replayed against each rule** (profit target, daily loss, overall loss, minimum trading days, best-day where it applies). A balance-only pass is read as balance-only.

§4

Field map

4. Where to look: a field map of an MT5 report

This is a map of where each field lives in a Strategy Tester report. It is a schematic with no values in it. It is not a sample report, and it is not evidence of anything. Wording varies slightly between MT5 builds.

Settings

Expert		
Symbol		A
Period		
Inputs		B
Company		
Initial Deposit		C
Spread, Model		D

Tester settings; not every build prints all of them.

Results

History Quality		E	Total Net Profit		
Bars			Profit Factor		G
Ticks		F	Maximal drawdown		H
Symbols			Total Trades		I

Balance Drawdown and Equity Drawdown are listed separately; read both. Largest profit trade sits further down this block.

Every figure in this block is the seller's own summary.

Balance / equity graph (an image in the report)

Not drawn here. Check it against Q4.

Deals (the trade list)

Time	Type	Volume	Commission	Swap	Profit	Balance

One row per deal. Sort by Profit, count the rows, and add the Commission and Swap columns yourself.

Field map, not a sample report. Grey bars stand where a value would be printed. No performance values are shown.

A **Symbol, Period.** The instrument and the date range. A short range or an unusual instrument is a Q2 question (§5.2).

B **Inputs.** Count the parameters. Many inputs tuned together is the multiple-testing risk in §5.1.

-
- C Initial Deposit.** Needed to turn a drawdown into a percentage, and to replay prop-firm rules (§6).
-
- D Spread and Model.** Whether ticks are real, and whether spread was set or defaulted (§5.2). If the report does not print them, ask for the tester settings.
-
- E History Quality.** Printed as History Quality or modelling quality, depending on build. Read it together with the model (§5.2).
-
- F Bars, Ticks, Symbols.** How much data sits behind the run. Very few ticks over a long period means thin or generated data (§5.2).
-
- G Profit Factor.** Gross profit divided by gross loss. It should agree with the Gross Profit and Gross Loss lines (§5.6). On its own it says nothing about how the profit was made.
-
- H Maximal drawdown.** Compare the balance and equity versions. Depth matters more than return (§5.5).
-
- I Total Trades.** Count the deal rows yourself and see that they match (§5.3, §5.6).
-
- J The deals table.** The real evidence. Counts, concentration, lot sizes, timing, commission and swap all come from here (§5.3, §5.4).
-

§5

In detail

5. The five questions, in detail

Each subsection below explains one question from §2 and names the tell to look for. Sections are numbered so you can cite them, for example "§5.3".

§5.1

Out-of-sample

5.1 Is there an out-of-sample segment at all?

Optimising an EA on a date range fits it to that range. The honest test is to run it on data it never saw, the out-of-sample (OOS) segment, and compare the two. A single in-sample run with no OOS comparison cannot support any claim that the result is stable.

The tell: one report, one date range, no second report. Ask for the OOS report and the date at which the split was made. If the OOS period was chosen after seeing results, it is not out-of-sample.

Parameters and the multiple-testing problem

Every combination of inputs that an optimiser tries is another attempt. Try enough combinations on the same history and some will look profitable by chance alone, even for a strategy with no edge at all. The report shows only the combination that was kept. It does not show how many were thrown away.

So the base rate matters. The more parameters an EA has, and the more optimisation runs were made, the more likely it is that a profitable-looking backtest is an accident of the data and not a property of the strategy. Ask how many inputs were optimised and how many runs were made. A seller who cannot answer has told you something.

VerifyEA applies the OOS test strictly. Given one report, it returns

UNVERIFIABLE for stability, because it refuses to guess. Give it a second report with `--oos` and it compares the two.

§5.2

Data quality

5.2 What are the modelling quality and the tick data?

The report states a modelling quality percentage and a model. "Every tick based on real ticks" uses recorded ticks. An OHLC-only run, or one that generates ticks from bars, means the price path inside each bar was partly invented. For an EA that depends on stops, scalping, or tight entries, the result can change materially with the data.

The tell: modelling quality well short of 100%, a model other than real ticks, a short history, or a period too narrow to include both trending and ranging markets. A run on poor data is invalid before any metric matters.

Where the ticks came from

Ask for the tick history source: the broker's own history, or a third-party download. The Bars and Ticks lines say how much data sits behind the run (field F on the map). Different sources give different spreads and gaps, so a run on one broker's ticks does not carry over to another's.

Spread, commission, swap

With real ticks the spread varies as it was recorded. With generated ticks the spread is usually one fixed value, either the current spread at the time of the run or a number someone typed in. Check whether it was set on purpose or simply left at the default. A fixed, narrow spread flatters any strategy that trades often.

Commission appears per deal in the deals table, and swap is charged on positions held overnight. A column of zeros means a cost was assumed away, unless the broker charges none. Add up both columns yourself and compare them with the profit they come out of. Slippage is not modelled in a standard backtest at all.

This is a manual check. VerifyEA reads the modelling quality from the report but does not issue a verdict on it.

§5.3

Concentration

5.3 Is the profit carried by a handful of trades?

Total profit hides how it was earned. Sort the trade list by profit. If removing the five largest winners turns the result flat or negative, the edge is a few events, and a few events do not repeat on demand.

The tell: a small number of trades far larger than the median. Read the distribution, not the total.

How many trades is enough

There is no universal threshold, and anyone who quotes one is guessing. The principle is that a claim needs enough trades that removing a few of them does not change the answer, and the more parameters were tuned (§5.1) the

more trades you need. A result built on a few dozen trades is a handful of anecdotes however good the ratio looks.

For reference, VerifyEA's own floors are about arithmetic, not reliability. Monte Carlo refuses to run on fewer than 30 closed trades, and the sequence checks report **UNVERIFIABLE** below 5. Clearing a floor means the calculation can run, not that the result is trustworthy.

Concentration is a manual check. Export the trade list, sort, drop the top few, and re-sum.

§5.4

Equity curve

5.4 Is the equity curve a straight line?

Real strategies lose in clusters. A curve that climbs almost without pause is usually produced by a structure that hides losses by holding or enlarging losing positions until they recover. It works until the one sequence it cannot recover from, and on a live account that sequence can end the account in a single move.

VerifyEA detects four structural fingerprints by name. Each leaves a mark in the trade list, which means you can look for them yourself.

grid_recovery

After a losing trade the next position is larger, often close to double, and the pattern repeats along a ladder such as 0.1, 0.2, 0.4, 0.8 lots. One step-up can be coincidence. Repeated post-loss step-ups, or a run of near-doubling sizes, is a recovery grid: it wins small and often until a long enough adverse move arrives, then the size of the last step makes the loss large.

time_clustering

Entries bunched into a narrow window, for instance a large share of all trades opening within one hour of the day, far denser than the rest of the distribution. It can mean the result depends on one session's conditions, or on a recurring event such as a rollover. Either way the headline numbers describe one narrow slice of the market, not a general edge.

identical_sequence

The same trade repeating: same symbol, same direction, same lot size, entered at the same position in the time grid, three or more times. Real markets rarely hand a strategy the same trade at the same moment repeatedly. A repeat often means copied, replayed, or fabricated history, or a rule that fires on a clock rather than on price.

lot_progression

Position size escalating over a run of consecutive trades, each step clearly larger than the last. Ordinary risk scaling moves in small steps with equity. A steep escalation means risk rises when the strategy is losing, which is the same logic as grid recovery applied to size alone.

Reading the result: each of these checks returns **FLAGGED** , **NOT_FLAGGED** , or **UNVERIFIABLE** . With too few trades to test, the answer is **UNVERIFIABLE** , never a silent **NOT_FLAGGED** . A flag is a reason to look closer, not a verdict on the seller.

§5.5

Reordering

5.5 Does the result survive reordering?

A backtest shows one sequence of trades. The same trades in a different order produce a different equity curve and a different worst drawdown. A Monte Carlo test reshuffles the closed trades many times and reports the distribution of drawdowns, so you see the range instead of one lucky path.

Look at the deeper percentiles of that distribution, not the median. Drawdown depth matters more than total return because it is the quantity that ends accounts: a 40% drawdown needs a 67% gain to recover, and every prop-firm loss limit is defined on drawdown, not on profit.

VerifyEA's Monte Carlo is seeded: the same seed gives the same numbers, so a result can be reproduced and checked. It covers closed trades only. It is a description of the reordering risk in the trades you have, **not a forecast**.

5.6 Is the report what it claims to be?

Everything above assumes the file is a genuine, unaltered Strategy Tester output. VerifyEA cannot confirm that, and neither can you from the numbers alone. What you can do is look for the tells that a report was made or changed by hand.

The tells: a screenshot of a "result" instead of the raw report file; a report that does not look like a standard MT5 export, such as custom HTML, unusual fonts, or a layout from a third-party page; totals that disagree with their own trade list (Total Trades against the number of deal rows, Profit Factor against Gross Profit and Gross Loss, the Balance column against running profit); a pasted or cropped table with missing rows.

Any one of these is a reason to ask for the original file, and for the settings and the broker account it came from. A seller who offers only an image, or who cannot produce the original, has given you a claim, not a report. If the seller holds the only copy, treat everything else as conditional.

6. Replay the trades against the rules before paying a fee

Prop firms publish their evaluation rules: a profit target, a daily loss limit, an overall loss limit, a minimum number of trading days, and at FTMO (1-step) a best-day rule that caps how much of the total profit one day may represent. Because the rules are published, a trade list can be replayed against them before you pay a challenge fee.

6.1 What the replay can and cannot see

An MT5 report contains closed trades. It does not contain intraday floating equity, the open-position profit and loss that most firms use to measure daily and overall loss. So VerifyEA's compliance replay is balance-scoped: it uses closed-trade balance.

Compliance verdicts and what each one means

FAIL	Conclusive. A breach is visible on closed-trade balance.
PASS_BALANCE_ONLY	No breach is visible on balance. It does not mean none occurred on equity: floating losses inside a day are not in the report.
UNVERIFIABLE	The data or the rule value cannot support a verdict, so none is given.

The overall result across all checks is one of **NO_BLOCKING_FINDINGS** , **FINDINGS_REQUIRE_REVIEW** , or **INSUFFICIENT_EVIDENCE** . None of these says a challenge will be passed. A clean replay of the past on balance is a necessary screen, not a prediction.

6.2 When a published rule is unclear

Firms change rules, and their own pages and help articles sometimes disagree on a value. Where a published value is inconsistent, report the rule as unknown instead of picking the convenient number. VerifyEA's bundled presets follow this: a rule with conflicting sources carries no value, and the check that needs it comes back **UNVERIFIABLE** . Always confirm the rule on the firm's current page and record the date.

VerifyEA ships presets for FTMO (1-step, 2-step), E8 (one, custom), The5ers (2-step, high-stakes), FundedNext, Alpha Capital, and Instant Funding.

verifyea rules list shows them with their versions.

7. If you want a tool to do the mechanical parts

VerifyEA is an offline command-line tool that audits MT5 Strategy Tester HTML reports. Nothing is uploaded. `verifyea audit report.html` runs the red-flag checks and the compliance replay, `--oos` adds the stability comparison, and `report --format html` renders the result. The free tier covers the first 100 trades of a report and the FTMO presets, and runs offline.

whop.com/quantsmith/verifyea-offline-audit-for-mt5-strategy-tester-reports-9d

This document is educational. It is not financial advice and it does not predict results. Published by Quantsmith. Firm rules change; confirm each rule on the firm's own page. How to audit an MT5 backtest, v1.0, 2026-10-15, CC BY 4.0.