

Backtest Honesty Checklist

A backtest is not a result. It is a claim about the past, produced by a program whose settings were chosen by someone with an interest in the claim looking good. That does not make every backtest false — it makes it *unverified* until you have asked how it was made.

This sheet is the sequence to run before you believe one. It rates no product and contains no performance figures. Every verdict on it is one you reach from evidence you gathered yourself.

Section A — Read the report, not the curve

The equity curve is the most persuasive chart in trading and the least informative. Demand these five numbers from the Strategy Tester report, and treat a missing one as an answer in itself.

Number	Where it is	What it exposes
Equity drawdown	Not balance drawdown — look for the equity figure.	Balance drawdown hides floating losses. Equity drawdown is what closes accounts.
Total trades	Alongside the period tested.	A long period with few trades is a small sample wearing a long label.
Profit factor	Gross profit ÷ gross loss.	Meaningless without the trade count next to it.
Average win / average loss	Both, plus the win rate.	A high win rate with a large average loss means one bad day erases many good ones.
Cost assumptions	Spread, commission and swap used in the test.	A run at zero spread describes a market that does not exist.

Write the numbers down before you look at the curve. Once you have seen a rising line, every figure afterwards gets read in its favour. That is not cynicism, it is how reading works.

Section B — The data underneath the test

- ☐ **Every tick, on real ticks.** “Open prices only” is not a backtest of an intraday system; it decides each bar at its open and cannot see what happened inside it.
- ☐ **Modelling quality is stated and high.** The Strategy Tester prints it — if the report does not say, the test was not run to be judged.
- ☐ **Spread is modelled, not fixed at zero.** A fixed spread is still a simplification, but zero is a fiction.
- ☐ **Commission and swap are included** if your broker charges them. A strategy that holds overnight and ignores swap is being measured on a different instrument to the one you would trade.
- ☐ **The history is long enough to contain a bad regime** — a trending year and a ranging year, not a single quiet stretch.
- ☐ **You can say which broker's data it used,** because ticks are not identical between brokers.

Section C — The optimisation audit

This is where most attractive backtests come from, and it is rarely dishonest on purpose — a parameter search is a normal step that quietly becomes the result.

- ☐ You know **which inputs were optimised**, and roughly over what ranges.
- ☐ You know **how many combinations** were tried. A search over many combinations will always find some that look excellent in hindsight.
- ☐ The chosen settings were then run **on a period they were not chosen on**.
- ☐ **Walk-forward or rolling tests** exist, rather than one continuous tuned run.
- ☐ The settings are **not on the edge of their range** — a value that only works at its maximum is a fitted value.
- ☐ Small changes to an input **do not destroy the result**.

The ratio worth estimating. Compare the number of trades to the number of adjustable inputs. When a system has many inputs and few trades, there is enough freedom to fit the noise — and the equity curve will look superb precisely because it was fitted.

Section D — Seven tells of a tuned result

What you see	Why it is suspicious
A curve that climbs almost in a straight line	Real trading has losing stretches. A line without them usually means losses are absorbed somewhere off the chart — averaging, grid, or a stop that is never reached in the test.
No losing months at all	Plausible over a short window, not over years of varying conditions.
The same trade appearing at the same hour, over and over	Often a fill assumption the tester grants and the broker does not.
A test period that ends just before a known crisis	Date ranges are chosen, and the end date is the easiest thing to choose.
A very large average win beside a very high win rate	Both together usually mean the losing column is incomplete.
The result collapses when spread is raised slightly	The edge is smaller than the cost of taking it.
Settings that only work at one exact value	That is the signature of fitting, not of a market behaviour.

Section E — Re-run it yourself

This is the only step that turns a claim into evidence, and it is tedious. Do it anyway — the effort is small next to a funded account.

- ☐ Same EA, same version, **your own MetaTrader, your broker's data**.
- ☐ Same symbol, same period, same starting balance as the report you are checking.
- ☐ Spread set to what *you* actually pay, not what the vendor used.
- ☐ Commission and swap set to your account's terms.
- ☐ Run it once with the published settings, and **change nothing else**.
- ☐ Compare the **shape and the drawdown**, not the final balance. A different ending with a similar character is usable information; a completely different character means the report described a different test.
- ☐ If it matches, run it again over a different stretch of the same length, chosen before you look.

What a match does and does not prove

A repeat proves the test can be reproduced from the stated assumptions. It does **not** prove the strategy will make money — it proves the claim was made honestly about a past that has already happened. Those are different sentences, and sales pages rely on you hearing the second when only the first is true.

Section F — From backtest to demo to live

Each step forward removes an assumption. Expect the numbers to get worse at every one, and treat a strategy that *improves* in live conditions as an anomaly to explain rather than a bonus.

Stage	What it assumes away	What stays unknown
Backtest	Nothing is assumed — but fills, spread and slippage are modelled, and the settings were chosen with knowledge of this period.	Everything about the future.
Forward / demo	Fills and spread are real; the settings are fixed in advance.	Whether the same fills hold under live routing, and how you behave when it draws down.
Live, small	Nothing.	Whether the edge survives real slippage, news spreads and your own decisions.

Section G — Verdict

- ☐ Five numbers written down before the curve was read (A).
- ☐ Data quality and cost assumptions known (B).
- ☐ Optimisation described, including what was tried (C).
- ☐ No unexplained item from the tells table (D).
- ☐ Re-run reproduced on my own data (E).
- ☐ A written decision: discard, adjust, or proceed to a fixed demo period (F).

Verdict: continue ☐ adjust and re-test ☐ discard ☐

Notes — what would change my mind, and the date I decided:

Backtest Honesty Checklist — a free resource from bestmt4ea.com. Free to use, print and share with credit. It is general information, not financial advice, and it rates no product.

Trading forex and CFDs on margin carries a high risk of loss and is not suitable for everyone. Leverage works against you as readily as for you. Historical results, backtests and demo results do not predict future performance. Test any expert advisor on a demo account, with money you can afford to lose, before trading it live.

bestmt4ea.com — full guides, EA reviews and free tools for MetaTrader 4 and MetaTrader 5 traders.