ALDERLUX

THE INSIDER · RESEARCH · #0034

The Alderlux Null Test

Last reviewed 13 November 2026

Alderlux runs a test on claims about levels, indicators and patterns. It takes the claim and sets it against a second market, built from the real record with its memory removed. If the claim performs the same there, it was never reading structure.

The test we run

We run it where a claim asserts that something works. What it finds is what we publish, including when what it finds is nothing.

Where a published finding has been through the test, the finding says so. Where it has not, we do not present it as tested.

The rest of this page explains what the test does, what it governs, and what it costs us.

Why a test is needed at all

A market moves. Prices rise and fall, retrace part of a move, and return to levels they have visited before. That happens whether or not anything meaningful is driving it.

So when an analyst says a level “holds” a certain proportion of the time, there are two possible explanations. The first is that the level is doing something — that the market has a memory, and that memory shows up at that price. The second is that any two prices a similar distance apart would have produced the same number, and the level is incidental.

A published figure that does not rule the second out is not evidence of the first. It may be perfectly correct and still tell you nothing, because nothing has been excluded. Separating the two is what turns a number into a finding.

The twin

The test is to build a second market and run the same measurement against it.

The twin is constructed from the real record. It keeps the statistical properties the test needs to hold constant — volatility, drift, the distribution of daily movement — and removes the structure the test is asking about. It is a market that moves like Bitcoin and remembers nothing.

There is no single way to build one, and that matters. Different procedures keep different features and therefore test different questions. Which properties a twin preserves, which it destroys, and what the comparison can and cannot establish are stated for each measurement rather than assumed from this page.

The measurement is then run against both. If the result is the same on the twin, the pattern was never reading structure. It was reading the geometry of two prices being a certain distance apart, which is a property of any market, including one made of noise. The comparison also runs the other way, which is less obvious and has mattered more: a market with no memory produces more apparent structure than most people expect, so a record that fails to do what noise does is itself a finding. More than once that has been the stronger result.

To be published as a finding, a pattern must beat its twins - not one of them, but the spread of results across many. A single structureless series can be beaten by luck, so the comparison is against the whole distribution those series produce. An interval that merely excludes zero does not clear that bar, and is not treated as though it does.

That is a precise and limited claim, and the limits are part of the standard. Beating a twin is evidence against the null model that twin represents. It is not proof that a pattern is profitable, and not proof that it will hold on data it has not seen. How strong the evidence is depends on how the twin was built, what was measured, and how many things were measured before this one.

The method is not ours

Testing a measurement against data that keeps a series' gross statistical properties while destroying its structure is established practice in the statistical literature, where it is known as surrogate data testing. Alderlux did not invent it and does not claim to have.

What this page commits to is when the test is run. Running it before publication, and running it afterwards to defend a figure already in print, are different activities with different incentives. Alderlux runs it first, and publishes the result either way.

What this governs, and what it does not

Definitions are not governed

What a structure is called, and what rules define it, cannot fail a null test. Taxonomy is publishable as taxonomy.

Claims about what works are governed

Any statement that a level, indicator or pattern predicts, holds or performs must have passed the twin, or be labelled as not yet tested.

Claims about how often something happens are governed

Frequency is an empirical claim, and "rare", "common" and "usually" are frequency claims wearing ordinary clothes. Either the figure is measured and printed, or the source is cited, or the word is kept alongside a plain statement that the rate is not something we have measured, or the word is deleted. Sometimes the fourth route is the only honest one: where no figure exists and none can be derived, saying so is better than deleting the observation or inventing a number for it. What is never acceptable is the word standing on its own, carrying the authority of a measurement that was never made.

Untested figures may be published, labelled

Research does not wait for certainty. A figure that has not been through the twin is published as provisional — and the label travels into the summary, the bullet list and the index entry, because the summary is what gets skimmed and quoted.

What the test costs us

It disqualifies work we have already done, including work this house previously relied on. It has sent findings back for rebuilding more than once, and it has retired readings we would have been content to publish a year ago.

It also makes us slower than firms that publish the first interesting number they find.

We accept both. A research house is worth exactly what its worst published claim is worth, and the cheapest way to protect that is to run the test before publication rather than to defend the claim afterwards.

The hardest part is the twin, not the test

Building a market that is genuinely structureless is more difficult than measuring against it. A twin that is too smooth, or that misreads the tails, will flatter or destroy a finding without any change to the underlying data. The procedure that removes memory also disturbs the way volatility arrives in clusters, so a verdict counts only once it holds across the settings that govern this. A result that changes when they change has not been tested; it has been sampled.

For that reason: no result built on a new or modified twin is published until someone independent has reproduced the headline figures from scratch — given the record and the question, but not our workings.

What you can hold us to

Every research note states the window it was measured over, start and end.

Any claim that something works or happens at a given rate has been through the twin, or is labelled as not yet tested.

Where a reading is retired, that is published. The standard that caused the retraction is the same standard that produced the original claim.

Borrowed frameworks are credited to whoever built them.

Where more than one variant was tested — levels, definitions, windows — the note says how many were tried, and reports the result for the set rather than only for the best of it.

This page is permanent.

Related reading: A Number Is Not a Finding and The 200-Poke Theory: Statement, Test and Withdrawal, both of which apply this standard.

Alderlux provides education and information only. We do not provide regulated financial advice, investment recommendations, legal or tax advice. Alderlux is not authorised by the Financial Conduct Authority.