ARCTAN RESEARCH

How We Try Not to Fool Ourselves, or You

The discipline behind our research: a plain-language guide

Tony Volpon · Arctan Research · arctanresearch.net

> *"The first principle is that you must not fool yourself — and you are the easiest person to fool."*
>
> — Richard Feynman

The problem: with enough tries and time, luck looks like skill

Imagine ten thousand people each flipping a coin ten times. By pure chance, around ten of them will flip ten heads in a row. If those ten then write newsletters about their coin-flipping technique, they will sound like geniuses.

Quantitative investment research faces exactly this trap, at industrial scale. Financial markets give us one history — one sequence of prices — and modern computers let a researcher test millions of ideas against it. Test enough ideas and some will fit the past beautifully by accident. Academics call the result the "factor zoo": hundreds of published patterns that "predicted" returns in old data, many of which quietly stopped working the moment real money followed them. The old joke is that if you torture the data long enough, it will confess to something eventually. The confession is worthless.

Everything we do at Arctan is organized around one goal: **making it as hard as possible for us to fool ourselves**. This note explains, in plain language, the specific rules we follow. None of them are our invention — they come from the best practices of experimental science and modern finance research. What is unusual is applying all of them, all the time, **and in some cases publicly**.

Rule 1 — Use only what was knowable at the time

Suppose a company reported its annual results in February, but three months later corrected an accounting error and restated them. A careless historical test uses the corrected numbers as if investors had them in February. A strategy built that way is peeking at the answer key: it "knew" in February what the world only learned in May.

The remedy is called **point-in-time data**. Every piece of information in our research carries a timestamp of when it became knowable — the date a filing was actually published, not the period it describes — and every simulated decision may use only what carried an earlier stamp. It sounds obvious. It is also, in practice, **the single most common way historical results get silently inflated**, which is why we treat data provenance as seriously as we treat the strategies themselves. When we recently rebuilt a study on a new database, **the first week was spent not on results but on verifying timestamps, cross-checking sources, and confirming that companies which went bankrupt decades ago are still in the data with their true, painful final returns**. Databases have a way of quietly forgetting the losers; markets did not, and neither may we.

Rule 2 — Write the test down before running it

The subtlest way to fool yourself is to decide what counts as success **after** seeing the results. Move a threshold here, exclude an "unusual" period there, and almost any idea can be made to look good — each small choice defensible, the sum of them fatal.

Our answer is **pre-registration**, borrowed from clinical medicine, where it became mandatory after too many drug trials quietly changed their definition of success midstream. Before we test a hypothesis, **we write a memo that fixes everything: the exact data, the exact procedure, the exact thresholds a result must clear — we call them gates** — and, just as importantly, what we expect to happen and what we will conclude in each scenario, including failure. Then the memo is frozen. Any later change must be a dated, written amendment, made before the affected step runs — never a silent adjustment after peeking.

For our most important studies we now go a step further and **register the analysis plan on the Open Science Framework**, a public scientific registry, before execution. The registration is timestamped and cannot be edited by anyone, including us. When we eventually claim "we predicted this in advance," any reader can click a link and verify it against an independent clock. We are, to our knowledge, among a very small number of investment firms that do this.

Rule 3 — One look at the answer sheet

Any flexible model can be tuned until it fits the past. The only honest question is how it performs on data it has never seen. **So we split history: an older training period where a model is built and calibrated, and a newer held-out period that is locked away, untouched, until the design is frozen**.

Then comes the part we are strictest about: the held-out data may be examined once. Our systems literally enforce this — the evaluation script requires an explicit declaration to run and refuses to run twice. *If a test fails on the held-out data, the answer is no*. We do not adjust the model and try again, because a hundred retries against the same data is just the coin-flipping story with extra steps. One question, one answer, and the answer stands.

Rule 4 — Ask what pure noise would have found

Modern statistical machinery is powerful enough to find patterns in anything — including in randomness. So before trusting what our machinery finds in real market data, **we run the identical machinery on synthetic data: hundreds of artificial histories, built to have the same statistical texture as the real thing but containing, by construction, no genuine signal at all.**

This **placebo test** tells us what our tools would "discover" in a world with nothing to discover. Only results that clearly exceed that benchmark deserve attention. In one recent study this test did double duty: it confirmed that a pattern in the real data was genuine — stronger than anything in three hundred noise worlds — while also revealing that the finer details of the fitted model were indistinguishable from what noise produces. Both halves mattered. Knowing precisely which parts of a result to trust, and which to discard, is the difference between research and storytelling.

Rule 5 — Distrust the data itself

Bad data produces confident nonsense. So before any hypothesis is tested, the data must pass its own examination — **what we call an integrity gate**. We cross-check every series against

independent sources. We scan for impossible values, stale feeds, and gaps. And we fingerprint each series against known history: a momentum series that does not crash in the spring of 2009, or a volatility series that sleeps through the March 2020 volatility explosion, is broken, whatever its documentation says. This is also how sign errors — a series accidentally flipped upside down — get caught before they can do harm.

These checks sound pedestrian. They regularly find real problems: missing bankruptcies, silently revised histories, mislabeled conventions. Each one found early is a false conclusion that never got the chance to exist.

Rule 6 — When the answer is no, the answer is no

Most of our carefully designed tests fail. This is not a confession; it is the system working. Markets are ferociously competitive, and most plausible-sounding ideas are already reflected in prices. A research process that mostly says yes is not discovering an exceptional world — it is fooling itself.

So we treat a well-established "no" as a first-class result. Failed tests are written up with the same care as successes, filed with a full account of why they failed, and closed — **which protects real capital from ideas that only ever worked on paper**, and stops us from quietly re-testing the same seductive idea every two years. When a test fails for an identifiable technical reason, we fix the flaw and say so; when it fails cleanly, we update our beliefs and say that too. The declared expectations from Rule 2 make this bite: we score our own predictions against outcomes, in writing, including the ones we got wrong.

What this buys you — and what it cannot

None of this guarantees returns. Nothing does, and you should be skeptical of anyone who implies otherwise. What this discipline buys is something narrower and, we believe, more valuable: **when a result survives this gauntlet — timestamped data, a frozen and publicly registered plan, a single look at unseen history, a placebo benchmark, audited inputs, and the standing willingness to accept no for an answer — it has earned the right to be taken seriously**. And when we tell you something did not work, you can trust that too.

The easiest person to fool is yourself. We have tried to build a research process where that person is watched, constantly, by the most suspicious auditor we could construct: our own written, frozen, public commitments. We would rather report fewer findings that are real than many that are stories. In the long run, we believe that is also the better investment strategy.

—

This note describes research methodology. It is not investment advice, and it describes no specific strategy, position, or performance.
© Arctan Research · arctanresearch.net