ASHITA ORBIS

Falsifiers for a Portfolio

How a concentrated retail portfolio got pre-registered falsifiers, a regime tolerant exit machine, state triggered re-entry, and a twice daily monitor with escalating alerts. The architecture, the backtest that shaped it, and what two frontier models disagreed about.

JUNE 11, 2026 · SYSTEMS · ASHITAORBIS.COM/POSTS/048-FALSIFIERS-FOR-A-PORTFOLIO

FINANCE · FALSIFIERS · FABLE · CLAUDE · INVESTING · SYSTEMS · MONITORING · EXIT-RULES · AI-ASSISTANCE

> **Editorial note (August 2026):** One sentence in the closing section — "a system built in June has survived nothing by August" — appeared verbatim in both this post and its companion, Auditing the Vibes, published the same day, and carried a deadline that has since passed. Both instances have been reworded, each in its own terms. Nothing about the argument changes.

A thesis is falsifiable when you can state, before the outcome, what observable evidence would mean it had broken. Scientific fields adopted preregistration because researchers kept discovering that hypotheses written after the data have a suspicious tendency to fit it. Investors face the identical failure mode with worse incentives and no journals, because a position you already hold will recruit any evidence into its defense. The structural story behind a stock (this company matters, this industry is the future) is usually true at every point of the cycle, including all the way down. The system this post describes applies preregistration to a concentrated retail portfolio. Every thesis decomposes into falsifiers with named public observables, watched by a monitor twice a day. Exit rules are chosen for robustness to regime uncertainty rather than optimality within a guessed regime, and the rules for getting back in make the exits cheap when they turn out to be false alarms. The story of why I needed this (five years of instinct that outperformed the index 4.84x to 1.85x while never writing a thesis down, audited by a model in an exercise that was equal parts vindication and exposure) has its own post. This one is the machine.

The context the machine was built for: a portfolio with 63 percent in a single memory semiconductor, satellites across the AI supply chain, and a modest line of credit, held by someone whose audited edge is episodic conviction and whose audited weakness is everything that surrounds a conviction once it is sized. The design constraint that matters most is that the operator cannot be trusted at extremes, which is not self-deprecation but the central empirical finding of the audit, and the machine treats it as a specification: every rule is written so that no judgment is ever required while afraid or euphoric. Past me decides. Present me executes, or does nothing.

Anatomy of a falsifier

A falsifier in this system has four parts: the claim being tested, a public observable, a deteriorating condition (WATCH), and a broken condition (FIRED). The discipline is in the observable. "Memory demand stays strong" is not a falsifier; it is a mood. "DRAM contract prices decline for two consecutive quarters, per the industry trackers that publish them" is a falsifier, because two people who disagree about the thesis would still agree on whether the condition occurred. The main position carries twelve, decomposing the thesis (memory prices rising on real shortage, long term agreements damping the cycle, the AI buildout actually getting built) into independently checkable failure modes. The first five are the obvious ones: contract pricing reversing, server memory availability normalizing, customers walking their agreements, hyperscaler capital expenditure cuts, the company's own margin guidance breaking. The other seven are subtler, and most were not my idea: peer margins peaking first, inventory days rising while prices stay high, high bandwidth memory demand weakening enough to push capacity back into the commodity market, supplier capital spending accelerating after demand softens, Chinese entrants qualifying at premium customers, the company's share of the next platform slipping, and credit stress in the buildout's financing stack. The last four came from a second model's independent review of the first model's list, and the additions turned out to be the early warning channel. The financing stack falsifier and the platform share falsifier were the first to move to WATCH, days after being written, on news the original list would have had no slot for.

Each satellite position carries its own smaller set, designed around the specific way that thesis would break: a customer churn repeat for the optics moonshot, a purchase order failing to convert by a fiscal deadline for the grid storage play, two consecutive quarters without a new named foundry customer for the turnaround. Satellite assessments run weekly rather than daily, because the positions are small enough that attention is the scarce resource being budgeted. Position rules are account aware, which in Canada matters more than it sounds, because losses realized inside a tax free account are dead forever. So the rules locate speculative coin flips in the taxable account, where a failed thesis at least harvests a deduction, and keep the registered accounts for positive expected value and decade horizon holds. They also note, uncomfortably and permanently, that one existing lottery ticket lives in the wrong account because the rule was written after the buy.

The monitor

The assessment loop is mundane and the mundanity is the point. A script wakes before the morning session, pulls prices for every position, and applies drawdown tripwires from tracked peaks. It then hands the falsifier table to a model with web search and one instruction: check each observable against current evidence, last thirty days weighted heaviest, return CLEAR or WATCH or FIRED per falsifier with the source named, and never invent a source. It runs again after the close, catching the structured evening news dump, with lightweight price checks every half hour in between. Status changes post to a channel; decision class events (a falsifier firing, the largest position crossing its drawdown tripwire, a price target being reached) escalate to direct messages that repeat every fifteen minutes until acknowledged, because

the operator mutes channels, and a tripwire that fires unheard is a tripwire that does not exist. A weekly digest summarizes the whole board even when nothing changed, which sounds redundant until you need to trust the system's silence, and a catalyst calendar holds the dated events (earnings windows, fiscal deadlines for promised purchase orders, the expiry of standing orders) so that nothing load bearing depends on anyone remembering anything.

One rule in the monitor exists because of a specific historical failure, and it is the rule I would export to other people first. In April 2025 I sold a thesis I still believed into the bottom of a 36 percent drawdown because a prestigious magazine convinced me the macro situation was unprecedented, and the postmortem finding was not that the article was wrong but that I had no procedure for adjudicating between new information and eloquent fear. The rule: a frightening narrative, whatever its prestige, must name which falsifier it would move, and then you wait for that observable at data speed rather than narrative speed. A narrative that cannot name its falsifier is noise by a definition I registered while calm. This is not a rule against macro awareness; it is a router that forces macro awareness through the same gate as everything else. In the weeks it has run, every alarming headline has either named its falsifier, in which case the falsifier was already being watched, or failed to, in which case the headline was, by construction, not about my thesis.

Exits chosen for regime tolerance

The audit's most actionable finding was that my exit plan contained only an upside trigger, a price at which I would reconsider, with no branch at all for the cycle turning first. The model's response was to backtest exit rule families against the position's actual history: sixteen drawdowns exceeding 40 percent since 1995 by the audit's cycle count, median modern depth in the mid 50s, roughly one every two years. Buying and holding through each historical bust path ends near breakeven by cycle reset but visits a trough around 44 cents on the dollar and spends up to three years underwater, with one path requiring two decades to recover. Tight trailing stops protect the trough but exit permanently into every false alarm. The decisive experiment ran the same rules through a synthetic regime in which my own thesis about damped cyclicality is true, with busts at half historical depth. Every rule family inverts there: holding becomes nearly optimal, stops become pure cost, and the question "which exit rule is best" turns out to be the question "which regime are we in" wearing a costume. The honest answer is that I do not know which regime we are in, and the only family that performs acceptably in both is the staged trim: a quarter of the position at one drawdown threshold, another quarter deeper, the remainder held through, landing within a few points of the winner in either world. Robustness to being wrong about the regime replaced optimality as the design principle, and I now think most retail exit debates are people arguing about optimality inside regimes they cannot verify.

The trim is state gated, not purely price gated: it arms when a cluster of falsifiers sits at WATCH, deterioration without confirmation. At that point the first tranche becomes a standing stop limit order at the broker, placed while calm, with a wide limit band so that anything short of a catastrophic gap executes without requiring courage from anyone at the bottom. And every exit has a symmetric re-entry, because the audit was clear that the expensive mistake is not selling a local bottom; it is selling a local bottom and

having no rule that brings you back. When the falsifier board returns to clear for two consecutive assessments, the tranche cash redeploys at whatever the price is, higher or lower, no judgment invited. The arithmetic on a false alarm round trip came out to one or two percent of the position, which is the known, accepted, preregistered insurance premium, and knowing it in advance is what makes the whole structure psychologically executable: the system's worst case is a number, not a humiliation.

Around all of this sits a downturn playbook whose default is freezing, because the audit found the operator's freeze instinct is genuinely good (it once held a 47 percent drawdown to a 170 percent exit). The exceptions wrapped around the freeze are pure arithmetic: tax loss harvesting in the taxable account, a ladder written in advance for swapping the weakest claims into the strongest one if quality ever goes properly on sale, and dry powder deployment in tranches at preset depths. The design principle repeats once more, because it is the entire system stated as one sentence: no decision at an extreme, in either direction, ever.

Two models, one point apart

The obvious objection to a model auditing the person who operates it is sycophancy, and the mitigation is structural rather than hopeful: the entire analysis, stripped to percentages, went independently to a second frontier model from a different company, no shared context, no access to the first model's work. It returned a median bust estimate within one point, scenario impacts within two points across the board, and a nearly verbatim exit verdict: staged selling is the only family robust to regime uncertainty. That is not proof of correctness. But two systems with different training and different incentives converging on uncomfortable numbers is not what flattery looks like. The disagreements were more valuable than the agreement: the second model contributed the financing stress and platform share falsifiers the first had missed, and both models, on the record, rate the central position several times larger than their frameworks would size it. That disagreement is preserved in the system rather than resolved, because the operator's conviction is itself a tracked input now, the residual risk the machine is built around, and I find I prefer being the documented anomaly in a working system to being the undocumented author of an unexamined one.

What this machine does not do is pick stocks, and after two weeks of running it I think that absence is the design's strongest claim rather than its limitation. The audited record says conviction calls were never the missing capability; everything else was, the falsifiers and exits and provenance and procedures for fear, all of it discipline rather than insight, and discipline is precisely the thing that cannot be summoned at the moment it is needed, only built in advance by someone calm. The model's actual contribution was to make calm available on demand, including the specific calm required to look at five years of winning and ask what, exactly, had been doing the winning. Two weeks of operation is not a track record, and this one has not yet met a market that disagreed with it; I will report honestly when the machine meets its first real bust. But the prior system met its first real test in April 2025 and failed it in the most human way possible, and the difference between the two failures is that the next one will be legible: every rule that breaks will have been written down, which means every rule that breaks can be fixed. Vibes do not have changelogs. That, in the end, is the whole argument.