

The most important number in finance was *self-reported* for forty-six years.

This document states the problem MIZAN exists to solve — precisely, with its literature, its price tag, and the reason every previous attempt failed. Nothing here is our claim alone: **every element below was documented by the field itself, across four decades, and never once closed.**

46	\$80B	<0.025	898
YEARS OF PUBLISHED WARNINGS, UNCLOSED	PER YEAR, DERIVED — ASSUMPTION MOVABLE	R ² OF BACKTEST SHARPE VS REALITY · 888 REAL ALGORITHMS	AI STRATEGIES FOR \$40 — ZERO SURVIVED HONESTY
0			
MECHANISMS THAT COULD CHECK THE TRIAL COUNT. UNTIL NOW			

1 · The deadlock — [Grossman & Stiglitz, 1980](#)

To verify a trading record, the examiner must inspect the research behind it. To survive, the quant can never reveal that research — **a disclosed edge stops being an edge**. These two requirements are **jointly unsatisfiable by any disclosure-based mechanism**. That is why the problem stayed open for forty-six years: every prior verifier had to pick a side. Auditors demanded disclosure and got refused by anyone with real IP. Platforms demanded custody of the strategy and became gatekeepers nobody portable trusted. Allocators, unable to check, substituted proxies — pedigree, references, the warm introduction — none of which measure the only thing that matters: whether the record is real.

2 · The number nobody could check — [Bailey & López de Prado, 2014](#); [Harvey & Liu, 2015](#)

Every statistical correction for luck — the Deflated Sharpe Ratio, White's Reality Check, Hansen's SPA, Harvey-Liu's multiple-testing haircuts — depends on a single input: **N, the number of strategies tried before the one being shown**. Understate N and every correction silently evaporates. And N has always been **typed in by the person being judged**. The [Notices of the American Mathematical Society \(2014\)](#) called failure to report the trial count a form of scientific fraud. It remained unreported — because no mechanism existed to make it reportable.

“Because most financial analysts and academics rarely report the number of configurations tried for a given backtest, investors cannot evaluate the degree of overfitting in most investment proposals.”

BAILEY · BORWEIN · LÓPEZ DE PRADO · ZHU — NOTICES OF THE AMERICAN MATHEMATICAL SOCIETY, 2014 · VERBATIM

The screening statistic the industry uses instead — backtest Sharpe — predicts out-of-sample performance with R² under 0.025 across 888 real trading algorithms.

The primary number on which trillions are allocated carries almost no information about what happens next, and the field has known this in print for a decade.

“Most claimed research findings in financial economics are likely false.”

HARVEY · LIU · ZHU — THE REVIEW OF FINANCIAL STUDIES, 2016 · VERBATIM

3 · The flood — *the AI acceleration, 2023–2026*

Generating a compelling backtest once took months of skilled work; the cost was a natural rate-limiter on fabrication. Large language models reduced it to minutes and dollars. One 2026 study generated **898 AI strategies for \$40 — zero survived an honest backtest**. Published literature now clocks models at 100% accuracy in-sample and a coin flip out of sample. Manual diligence — reading, references, judgment — cannot filter output produced faster than it can be read. **The trust problem became an arms race, and trust lost.**

4 · The price of the hole

COST	DERIVATION
~\$80B per year, recurring	If systematic capital (\$8T) underdelivers its backtested expectation by just 1% — conservative, given R ² < 0.025 — the annual shortfall is \$80B, paid by allocators who had no way to check. The assumption is stated so a sceptic can move it; halving it still leaves \$40B.
The diligence tax	\$50,000 manual audits per manager that structurally cannot check the trial count at all — the one number the corrections need.
The invisible loss	Real edge that goes unfunded because its owner cannot prove it without surrendering it: talent priced at zero for lack of a verification mechanism. Unmeasurable, and every allocator knows it exists.
The trust ceiling	The industry's answer — “invest in people you know” — caps the market at the size of everyone's rolodex. \$23T allocated on handshake-era mechanisms.

5 · Why every previous fix failed

ATTEMPT	WHAT IT VERIFIES	WHAT IT STRUCTURALLY CANNOT DO
Manual audit (\$50K)	That reported returns match statements	Cannot see the trial count, the search, or hindsight; annual, not continuous
GIPS (1999)	That the reporting <i>process</i> follows standards	Verifies process, not math; silent on overfitting, trial counts and backtests entirely
Platform track records	Live performance inside one venue's walls	Records die at the platform boundary; the platform is a player, not a referee; backtests unverified
Track-record databases	Self-reported numbers, collected	Garbage in, garbage archived — the input is the thing that needed verifying
Trust + pedigree	Where you worked, who vouches	Measures social position, not statistical truth; excludes everyone outside the rolodex

The pattern: every mechanism either demands disclosure (and gets refused), takes custody (and becomes a conflicted gatekeeper), or accepts self-reports (and verifies nothing). The deadlock of §1 is why — **the constraint was mathematical, not commercial**. No amount of diligence, regulation or reputation dissolves a jointly-unsatisfiable requirement. Only changing the mathematics does.

“The number of small gains far exceeds the number of small losses... The discontinuity is absent in the 3 months culminating in an audit.”

BOLLEN & POOL — THE JOURNAL OF FINANCE, 2009 · VERBATIM · REPORTED RETURNS BEND, AND STRAIGHTEN ONLY WHEN AN AUDIT APPROACHES

6 · What changed — and what "solved" means

Zero-knowledge proofs became practical — at this cost and speed, only in the last two years. A computation can now be performed inside a cryptographic circuit that proves the result correct **without revealing the inputs**. Applied to the deadlock: the record can be judged while the strategy stays sealed. The two jointly-unsatisfiable requirements stopped being joint.

MIZAN's construction closes each element of the problem in turn:

THE PROBLEM ELEMENT	THE MECHANISM THAT CLOSES IT
The deadlock (§1)	The full evaluation — returns, costs, out-of-sample behaviour, the gate verdict — is computed inside a zero-knowledge proof on committed data. Anyone verifies the result offline in ~23 ms; nobody sees the strategy. Not the allocator. Not MIZAN.
The self-reported N (§2)	The committed trial ledger: the candidate set is Merkle-committed <i>before</i> evaluation; N becomes the leaf count; the winner is proven maximal in-circuit; the Deflated Sharpe is recomputed against that N. The anti-luck correction can no longer be understated by the party it constrains. To our knowledge, the first mechanism of its kind (SSRN 7187038).
The AI flood (§3)	A locked gate, identical for everyone, enforced in-circuit — fabricated backtests don't argue with diligence, they fail cryptography. And for AI strategies themselves: the sealed path proves a model's record with weights never revealed.
Hindsight (the last refuge)	The live chain: strategy hashed and anchored (Bitcoin + RFC-3161) before the data it will be judged on exists, then re-proven on the growing window, every link chained to its parent. "Was this built in hindsight?" becomes a millisecond check. First live chain public — anchored in Bitcoin block 962,013 .

7 · The evidence that it is actually solved

NOT A CLAIM — A PUBLIC RECORD, RE-VERIFIABLE BY ANYONE, FOREVER

The gate refused Apple buy-and-hold at **+272,548%** (Sharpe 0.89 vs the 1.20 bar) — **REFUSED**. It refused the S&P 500 itself, proving in-circuit that the benchmark is **99% market beta** — **REFUSED**. It refused its maker's own +895% AI model, and shipped its maker's own flagship as **NOT SIGNIFICANT** at the 95% bar. **77 credentials on an append-only public wall, 54 of them refusals** — because a gate that would pass everything certifies nothing. The standard behind it (VTR-1: 41 normative requirements, conformance vectors, public governance and errata) is published on SSRN under CC BY 4.0 — free for anyone to implement, including competitors. That is what makes it a standard rather than a product.

8 · The problem, restated as its solution

For forty-six years: *"trust me — I can't show you."*
Now: *"check me — I don't have to show you."*

Every era of finance installed a verification standard once the market outgrew handshakes — ratings in 1909, audit in 1923, GIPS in 1999. Systematic finance, the last \$23T holdout, gets its standard now. The document half is published. What remains is the requirement — and the wedge producing it is already running.

MIZAN · the problem document · shareable · verify everything: [studio.mizan.market/app/#registry](#) · the standard: [ssrn.com/abstract=7215664](#) · the mechanism: [ssrn.com/abstract=7187038](#) · reproduce a credential: [github.com/muaviamohammed/mizan-verify](#)
Literature (all openable): [Grossman & Stiglitz 1980](#) · [White 2000](#) · [Lo 2002](#) · [Getmansky, Lo & Makarov 2004](#) · [Hansen SPA 2005](#) · [Bollen & Pool 2009](#) · [Agarwal, Daniel & Naik 2011](#) · [Bailey, Borwein, López de Prado & Zhu 2014](#) · [DSR 2014](#) · [Novy-Marx 2015](#) · [Harvey, Liu & Zhu 2016](#) · [Wiecki et al. 2016](#) · [Jensen, Kelly & Pedersen 2023](#) · the full curated list with excerpts and verbatim quotations: [mizan.market/evidence](#).