ARKAYA RISK
Building resilience capital

The governance signal under radical uncertainty

A tail-risk, value-investor and reflexivity reading of what verifiable governance evidence can and cannot do. v4 · 1 July 2026 · Strategic paper · shareable.

The test

The argument that verifiable evidence reduces the uncertainty a counterparty prices is an economic one, and it answers well to economists. Its sharpest test comes from elsewhere. The people who have spent careers on uncertainty and loss are not the theorists but the practitioners: the tail-risk school that hedges the rare catastrophe, and the value investors who buy under irreducible doubt. Both are professionally suspicious of any claim that risk has been made smaller. If the argument survives Nassim Taleb and Mark Spitznagel on one side, Howard Marks, Warren Buffett and Charlie Munger on the other, and George Soros on the flank where a priced signal feeds back into the thing it measures, it is worth more than a proof. This note runs it through all six, the tail-risk school as the principal objection, the value investors as independent corroboration from the opposite tradition, and reflexivity as the attack the other five leave standing. It survives, but only because of what it does not claim, and the discipline that survival demands is the point.

The tail-risk objection

Take the hardest reading first. Taleb's life's work is that the losses that matter are the ones outside the model, the consequential and the unpredictable, and that risk metrics do not merely fail to see them but actively manufacture the confidence that makes them ruinous [1]. A number that looks like knowledge invites the gearing that turns a shock into a collapse. The credit ratings and value-at-risk models trusted before 2008 did not sit passively beside the crisis; they were part of its cause, because they let institutions believe a tail had been measured and could be borne. On this reading a governance signal is a candidate to become the next such number: a reassuring metric that the market gears against until the thing it did not capture arrives.

The objection has to be taken at full strength, because it is correct about a real failure mode. Spitznagel's version is quieter and just as sharp: risk mitigation is worth having only if it mitigates the risks that actually matter, and much of what markets buy as protection addresses the risks that do not [2]. A signal that reduced the wrong uncertainty, or that was believed to reduce more than it did, would be worse than none, because it would be priced as safety.

Why the argument survives it: addition by subtraction

The argument survives because of what the signal acts on. It does not touch the tail. It acts on the epistemic gap between two parties, the concealment component, what one party can hide from another, which is a different quantity from the irreducible uncertainty of extreme outcomes; the distinction is Knight's, and it is the reason the signal and the tail-risk school do not collide [3]. Reducing what a counterparty can conceal is not a forecast of the catastrophe and makes no claim to see it. It removes an opacity rather than adding a prediction.

That is consistent with one form of risk reduction the tail-risk school generally accepts. Taleb's via negativa holds that robustness comes from removing sources of harm, not from forecasting them, that you improve a system more reliably by subtraction than by addition [4]. Verifiable evidence is subtractive in exactly this sense. It removes the lemons discount, the pooled price, the premium a counterparty charges for what it cannot see. It adds no model of the future and asks to be trusted with

none. A signal built this way is robust to the Taleb critique not despite refusing to predict but because it refuses to predict, and the discipline that keeps it honest is the refusal to emit a score or forecast the outcome, attesting only the verifiable present state. A number that describes what is known cannot become a false claim about what is not, unless its users force it to, which is the tension held to the end of this note.

Buffett: risk is not knowing

The value investors arrive at the same place from the opposite direction, and they make the argument in plain speech. Speaking to students at Columbia in 1993, Warren Buffett gave the definition the academy still resists: risk comes from not knowing what you are doing [5]. Risk, on this account, is epistemic before it is statistical. It is a property of the investor's knowledge, not only of the asset's variance.

The margin of safety that Benjamin Graham built the discipline on, stated in Security Analysis in 1934 and carried to the general reader in The Intelligent Investor fifteen years later, is the direct expression of it: you demand a discount to intrinsic value precisely because your knowledge is incomplete, and the discount is the price of your own uncertainty [6]. This is the uncertainty loading of the economic argument, in an investor's language. Narrow what the buyer cannot know about the asset, and the margin of safety it must demand narrows with it. Buffett's preference for transparent, comprehensible businesses with durable economics, inside a circle of competence, is the same mechanism read forward: he pays up for what he can know and discounts what he cannot, and verifiable evidence is a way of moving a counterparty from the second category toward the first. The governance signal does not tell him the business is good. It widens the set of things he can verify about it, which is the only thing his method ever asks of new information.

Marks: the cycle and the probability of loss

Howard Marks is the friendliest of the six to the argument and the most exacting about its limits. His definition of risk is the probability of a permanent loss of capital, not volatility, and his central and inconvenient claim is that risk of this kind cannot be objectively measured and is highest exactly when it is least perceived [7]. A signal that fed the perception that risk had been removed would, on his account, be most dangerous at the moment it was most trusted.

The argument concedes the point rather than resisting it, and the concession is the safeguard. Verifiable evidence does not repeal the cycle; it acts on the counterparty-specific component of a price, not on the market's appetite for risk at large, which swings on its own and is highest, as Marks says, when complacency is deepest. His second-level thinking, the edge that comes from reading the same facts better than the consensus, is the same edge the signal industrialises: it compresses the informational gap the consensus is mispricing. But he would insist, and the argument agrees, that the signal must never be allowed to say risk has been eliminated, because the belief that it has is the most dangerous belief in markets. Evidence moves the price toward the truth. It does not abolish the cycle that moves the price around the truth.

Munger: invert the question

Charlie Munger's contribution is a method, and it produces the strongest version of the case. Following Jacobi, he insists on inversion: to understand a thing, study its opposite [8]. Invert the governance argument and it reads more forcefully than in its forward form. Do not ask what verifiable evidence earns the well-governed. Ask what happens to the party that cannot prove its governance. It pays the lemons premium. It is credit-rationed rather than priced. It carries the heavier holdback in a sale, the worse rating, the tighter supplier terms, and it is the first counterparty cut when capital contracts. The reward for proof and the penalty for opacity are the same mechanism seen from two sides, and the penalty is the sharper motivator.

Munger's other great theme completes the point. Incentive-caused bias, his name for the way people come to believe whatever their incentives reward, is the reason a party's account of its own governance carries no information: the incentive to report well is universal, so the report separates no one. It is also the structural indictment of the issuer-paid certificate, whose author is paid by the party it assesses. Only a signal that is costly to fabricate and independent of the assessed party escapes

the bias, which is the same conclusion the economics reaches through Spence, arrived at through psychology instead. Munger's wider teaching points the same way. His insistence on a latticework of many mental models rather than one fits an argument in which verifiable governance evidence sits alongside underwriting, financial analysis and operational diligence as a lens among several, complementary to them and a substitute for none.

Soros: the signal changes what it measures

The sixth test is the one the first five leave standing, and it comes from the practitioner of feedback rather than of loss. George Soros's reflexivity holds that in markets the measure and the measured are not independent: participants act on their picture of reality, and the acting changes the reality the picture describes [9]. Charles Goodhart gave the special case its law: a statistical regularity collapses once pressure is placed on it for control purposes, or in the later popular form, a measure that becomes a target ceases to be a good measure [10]. Applied here the objection is exact. Once counterparties price the governance signal, firms will manage to the signal rather than to the governance: author Protocols that are generous, run Signal classes that are trivial, and manufacture Velocity where governance should be. The signal, priced, corrupts itself.

The architecture's answer is that it was designed against this failure rather than surprised by it, and the defences are structural. There is no score, so there is no single number to manage to; the read decomposes to the underlying proofs, and a counterparty that gears against a compression of them has built the compression itself. The Protocol is part of the read, so the generous Protocol is visible as generous: gaming the bounds is legible as gamed bounds, and prices as such. And the record is symmetric, discoverable and producible, so the firm that manufactures evidence has manufactured exhibits against itself. What remains is the half of reflexivity Soros himself insisted on: the loop runs in both directions. A firm that must evidence its governance continuously tends, over time, to run the governance the evidence demands, in the way a business that must report to the market comes to manage what it reports. The signal does change what it measures. Built to attest rather than to score, that is the point rather than the defect: the feedback it creates runs toward the conduct, because the conduct is the only thing that moves the record.

The tension the schools are right about

The honesty the whole argument depends on shows itself here. The risks that destroy a counterparty are the ones that fake the evidence or lie wholly outside it, and a signal the market comes to trust raises the prize for forging it. Verifiability raises the cost of the lie far above that of a bare assertion, because a counterfeit that must survive independent cryptographic checking is a harder thing to produce than a false claim, and harder still the more independent the parties that would have to collude to forge it, though the cost never falls to zero, and it can do nothing about the shock no governance record would ever contain. This is where Taleb, Spitznagel and Marks are simply right, and the argument must stay humble in front of them: the signal lowers the price of the knowable and leaves the unknowable exactly where it was. Treated as more than that, geared against as though it had priced the tail, it would become the fragility they warn of. The hazard has a precise form. A market handed a standardised, verifiable read will be tempted to compress it into a crude accept-or-reject switch and gear against that. Nothing can stop a counterparty building a poor model on good evidence; but because the signal carries no score and the read decomposes to the underlying proofs, the switch is the counterparty's own construction, always traceable back to the verifiable state it flattened, and the fragility, when it comes, is in the model and not in the signal. The safeguard is not a feature to be added later. It is the refusal, built in, to forecast or to score.

What survives

Put the six together and they say one thing. The enemy is unrecognised uncertainty and the false confidence that hides it. The governance signal survives their scrutiny precisely because it works on the opposite quantity: the recognised, concealed, reducible uncertainty between two parties, which it lowers by removal, not by prediction. It refuses to measure the tail, refuses to score, and concedes that it changes neither the true risk nor the cycle. What is left is accuracy about what can be known, and nothing more, which is all any of these six ever trusted. The via negativa reading and the inversion reading meet in the same place: the value of the signal is as much the penalty it imposes on

those who cannot prove their governance as the reward it offers those who can, and neither is a claim on the future. The reflexive reading, run honestly, is the mechanism by which that penalty and that reward do their work over time. And the six would jointly insist on one further discipline, which the argument accepts: the claim that markets price this signal is itself, for governance evidence, an untested comparative static, an experiment with the result pending, and holding it as a hypothesis rather than a fact is exactly the epistemic humility the whole note argues for. Evidence is not a forecast. It is the removal of an excuse for mispricing, and that is a smaller and more durable thing than the practitioners fear a risk number will always try to be.

References

1. **Nassim Nicholas Taleb, The Black Swan, 2007.** The consequential event lies outside the model, and risk metrics manufacture the confidence that makes it ruinous. The core objection the argument must survive, and the reason it must never present itself as a measure of the tail. Verified.

2. **Mark Spitznagel, Safe Haven: Investing for Financial Storms, Wiley, 2021 (foreword by N. N. Taleb).** Risk mitigation is worth having only if it is cost-effective and addresses the risks that matter. The test the signal passes by acting on concealment, and the warning that mitigating the wrong uncertainty is worse than none. Verified.

3. **Frank H. Knight, Risk, Uncertainty and Profit, 1921.** The distinction between measurable risk and unmeasurable uncertainty. The signal acts on the reducible, epistemic part; the practitioners in this note are the custodians of the irreducible part, which is why the two do not collide. Verified.

4. **Nassim Nicholas Taleb, Antifragile, 2012.** The via negativa: robustness comes from removing harm, not forecasting it. The principle under which verifiable evidence, which subtracts an opacity rather than adding a prediction, counts as sound risk reduction. Verified.

5. **Warren Buffett, Columbia Business School, 1993, reported in the Omaha World-Herald, January 1994, and in R. Hagstrom, The Warren Buffett Way, 1994.** “Risk comes from not knowing what you are doing.” The plainest statement that risk is epistemic before it is statistical. Verified to the earliest reliable published source.

6. **Benjamin Graham and David L. Dodd, Security Analysis, 1934; Benjamin Graham, The Intelligent Investor, 1949.** The margin of safety: a discount demanded because the investor’s knowledge is incomplete, stated first in Security Analysis and carried to the general reader in The Intelligent Investor. The uncertainty loading in an investor’s language, and the quantity verifiable evidence narrows. Verified.

7. **Howard Marks, The Most Important Thing, 2011.** Risk is the probability of permanent loss, not volatility; it cannot be objectively measured and is highest when least perceived. The source of the argument’s concession that it does not repeal the cycle and must never claim risk is removed. Verified.

8. **Charles T. Munger, in Poor Charlie’s Almanack, ed. P. Kaufman, 2005.** Jacobi’s “invert, always invert,” and incentive-caused bias. Inversion yields the penalty-for-opacity reading; incentive-caused bias explains why self-report and issuer-paid assessment separate no one. Verified.

9. **George Soros, The Alchemy of Finance, 1987.** Reflexivity: participants’ picture of reality feeds back into the reality it describes, so a priced signal changes the behaviour it measures. The sixth test, and the failure mode the signal’s no-score, decomposable, symmetric design exists to survive. Verified.

10. **Charles Goodhart, “Problems of Monetary Management: The UK Experience,” in Papers in Monetary Economics, Reserve Bank of Australia, 1975.** The origin of Goodhart’s law: any observed statistical regularity tends to collapse once pressure is placed upon it for control purposes. The popular form, that a measure which becomes a target ceases to be a good measure, is Marilyn Strathern’s 1997 paraphrase. The special case of reflexivity the signal must be built against. Verified.

Production Note

Standard: Arkaya Production Standard v1.
Drafting tools: Claude (Anthropic). Research synthesis assisted; all assertions verified against primary sources.
Verification: Primary sources named in the numbered references; each reference carries substantive content. Citation audit complete pre-delivery.
Human approver: David J McKibbin, Founder, Arkaya Risk.
Version: v4. Approved for library publication by the founder, 3 July 2026.
AKR AIPS · v4