Warping the Glass: Measurement Integrity in Measurement-Based Monetary Systems

Shehzad Ahmed*

June 2026

Abstract

A measurement-based monetary system replaces chosen policy rates (committees) with measured ones (sensors): credit intensities read from markets, money valued by measured backing, triggers fired by observed stress. This removes the committee as a capture target and creates a new one — when the policy variable is a measurement, corruption migrates to the measuring layer. The attacker no longer debases the coin or lobbies the committee; it warps the glass. We analyse the full measurement stack of such a system layer by layer and derive the calibration each layer needs to make warping unprofitable. Three results organise the paper. *First*, for stake-weighted-median oracles we derive the manipulation frontier in closed form and show implementable slashing requires a **depth rule**: the value the oracle may reprice per detection window must be capped relative to bonded stake (at our calibration, roughly 7%) — a leverage limit on the measurement itself. *Second*, for the social layer (registrars attesting real-asset backing) the same structure reappears: honesty requires collateral $\geq$ attested value / audit probability, which makes single-registrar attestation unscalable and yields a quorum rule under which safe attested value grows approximately quadratically in the attestation quorum. *Third*, local no-profit calibrations **do not compose**: we exhibit a mixed attack — an in-band oracle bias inflating the value of fraudulently attested collateral — whose cross-term is priced by neither layer's defense. Composition is restored by an end-to-end audit: a sequential (CUSUM) monitor of the divergence between system prices and an independent instrument bounds *total* distortion regardless of source. Calibrated on a real 14-year, 7-country panel of credit intensities against sovereign CDS (per-country correlations 0.989–0.998), the monitor detects a 25-basis- point warp at the first observation and caps the optimal slow attack at roughly 10 bp-years of repriceable notional. The residual trust assumption is identified exactly: at least one audit instrument must remain beyond the attacker's reach.

Keywords: oracle manipulation, measurement integrity, mechanism design, sequential detection, blockchain oracles, benchmark manipulation, registries. **JEL:** D82, G14, K42, L86.

1 Introduction

Monetary systems have always had a layer an attacker most wants to corrupt. In commodity-money systems it was the coin itself: clipping and debasement. In administered systems it is the chooser: committee capture and benchmark rigging — the LIBOR scandal being the canonical modern case, in which the measurement was a survey and the survey was the attack

*Independent University, Bangladesh. This paper extends the mechanism-design analysis of Ahmed (2026a) to the measurement layer of the interest-free Layer-1 specification of Ahmed (2026b); the problem is general to any system whose policy variables are measurements rather than choices.

surface (Duffie & Stein, 2015). A new class of designs — zero-emission and interest-free Layer-1 architectures in which credit intensities are read from live markets, money is valued by its measured backing, and stabilisation mechanisms fire on observed stress (Ahmed, 2026b) — removes the committee entirely. Every policy variable becomes a *measurement*.

This is usually presented, correctly, as an integrity gain: a stake-weighted median computed from observed trades cannot be moved by a phone call. But the attack surface does not vanish; it migrates. If money is a mirror of real value, the profitable crime is no longer to counterfeit the metal but to *warp the glass*: bias the sensors, corrupt the attestations that bind tokens to real assets, or grind the reference constants — each individually small, none visible in any single epoch, all cumulatively a wealth transfer. Manipulation of thin crypto markets is documented (Gandal et al., 2018); what is missing is a design theory: **what does each layer of the measurement stack cost to corrupt, and how must collateral, audits, and monitors be calibrated so that no warp — fast, slow, or mixed — is profitable?**

Contributions. We model the measurement stack of a measurement-based monetary system in four layers — market oracles, asset registries, reference constants, and an end-to-end audit — and derive, for each, the no-profitable-warp calibration. *First* (Section 3), for stake-weighted-median oracles we give the achievable bias of a λ-stake coalition in closed form, $b(\lambda) = \sigma\,\Phi^{-1}\!\left(\frac{1/2}{1-\lambda}\right)$ capped by the validity band, and derive the slashing calibration that makes manipulation unprofitable. Because slashing cannot exceed stake, implementability forces a **depth rule**: repriceable value per detection window capped relative to bonded stake (Proposition 2). *Second* (Section 4), we construct the end-to-end audit — a two-sided CUSUM (Page, 1954) on the divergence between the system's credit intensities and an independent instrument — and calibrate it on a real panel: fourteen years of sovereign credit intensities implied by sukuk spreads against independently quoted sovereign CDS for seven countries (Ahmed & Bhuyan, 2026). The monitor detects a 25bp warp at the first print; the optimal *slow* attack is bounded at ≈10 bp-years. *Third* (Section 5), for the social layer we apply the classic deterrence logic (Becker, 1968; Shapiro & Stiglitz, 1984) to registrars: honesty requires bond-plus-franchise value ≥ attested value / audit probability, which makes single-registrar attestation of large assets impossible and yields a k-of-n quorum rule with safe value growing $\sim k^2$. *Fourth* (Section 6), we show local calibrations do not compose — an explicit mixed attack with a cross-term priced by neither layer — and prove that the end-to-end monitor restores composition because final-price divergence is additive in total distortion regardless of its source (Theorem 1). The residual assumption — the independence of at least one audit instrument — is stated exactly, because it is the system's final trust anchor.

A unifying principle emerges that we did not impose: at every layer, the no-warp condition takes the same form,

$$\textbf{collateral at risk} \;\geq\; \frac{\textbf{value repriced by the measurement}}{\textbf{detection probability}},$$

a *leverage limit on measurement itself*. Systems that let measured values grow faster than the collateral and audit capacity behind their sensors are undercollateralised in a precise, calculable sense — whatever their balance sheets say.

2 The measurement stack and threat model

We consider a monetary system in which: (i) per-market credit/convergence intensities $\hat{\kappa}$ are computed each epoch as stake-weighted medians of validator-submitted observations of market state, with submissions outside a validity band rejected and provably deviant

submissions slashable ex post; (ii) instruments must reference registered real assets, with registrars attesting the binding between on-chain tokens and off-chain assets; (iii) mechanism parameters (reference costs, trigger thresholds) are fixed governance constants recalibrated at low frequency; and (iv) an independent instrument (e.g. sovereign CDS quoted in a separate market) prices overlapping risks. The attacker is a coalition with a finite budget that may simultaneously: acquire validator stake, bribe registrars, and sustain sub-threshold biases over time. Its objective is extracted wealth net of capital costs, bribes, and expected penalties.

Four layers, four attacks: **L1** oracle capture (move the median); **L2** registry fraud (attest fake backing); **L3** the grind (sustained small bias below per-epoch detection); **L4** mixed attacks across layers. We treat them in turn.

3 Oracle capture and the depth rule

Honest submissions in an epoch are distributed F around the true value with dispersion σ (basis noise); the attacker controls stake share $\lambda < \frac{1}{2}$ and submits one-sided at the edge of the validity band $\pm\tau$.

Proposition 1 (Achievable bias)**.** *The stake-weighted median under one-sided piling is the q-quantile of the honest distribution with $q = \frac{1/2}{1-\lambda}$; for Gaussian noise the achievable bias is*

$$b(\lambda) \;=\; \min\Big(\sigma\,\Phi^{-1}\Big(\tfrac{1/2}{1-\lambda}\Big),\ \tau\Big),$$

increasing and convex in λ, reaching the band cap as $\lambda \to \frac{1}{2}$.

This closed-form bias law is reproduced empirically in the companion simulation framework (Ahmed & Bhuyan, 2026b): a Monte-Carlo Byzantine oracle in which the adversary captures the largest stakers and submits a +500bp lie holds the stake-weighted median to 2–4bp of bias for all $\lambda \le 1/3$ and lets it run away only as $\lambda \to 1/2$ — exactly the convex $b(\lambda)$ above and its $1/2$ breakdown — while a stake-weighted *mean* is dragged by hundreds of basis points. The same experiment confirms the liveness companion of the depth rule: under a *withholding* attack the $\ge 2/3$ acceptance quorum converts any $\lambda > 1/3$ into a safe pause at the last good value, never an accepted bad $\hat{\kappa}$.

The attacker's one-window profit is $b \cdot D \cdot H$, where D is the value repriced per unit of $\hat{\kappa}$ (open interest and collateral marked to the oracle) and H the epochs before ex-post audit. Its expected cost is $\lambda S \cdot s \cdot p_{\mathrm{det}}(b)$: stake at risk times slash fraction times the probability a deviation of size b is flagged against realized data. The no-profit condition gives the required slash fraction $s^*(\lambda) = b(\lambda)\,D\,H/(\lambda S\,p_{\mathrm{det}})$, plotted in Figure 1. Since $s^* \le 1$ (one cannot slash more than the stake), implementability is a constraint on the *market*, not the penalty:

Proposition 2 (The depth rule)**.** *The slashing mechanism can deter all coalitions $\lambda < \frac{1}{2}$ only if*

$$\frac{D}{S} \;\le\; \min_{\lambda}\ \frac{\lambda\,p_{\mathrm{det}}(b(\lambda))}{b(\lambda)\,H},$$

i.e. the value the oracle reprices per detection window is capped relative to bonded stake. At our calibration (σ equal to the empirical stationary dispersion of the intensity series, a $\pm 200bp$ band, daily audit) the cap binds at $\lambda^* \approx 0.42$ and equals $D/S \approx 0.07$.

Remark 1. *This is a leverage limit on the measurement itself, and it is a new design constraint for oracle-centric architectures: open interest and oracle-marked collateral must scale* with bonded stake*, not merely with demand. Where $s^* > 1$ no slashing schedule deters capture; the remedies are smaller D/S, faster audits (smaller H), or better detection — which is what Section 4 supplies.*

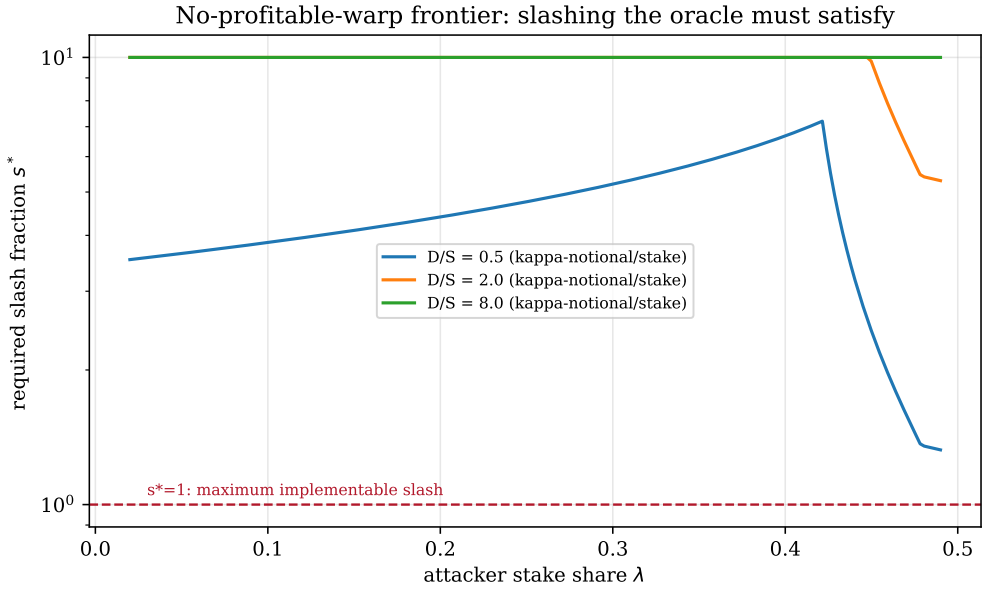

Figure 1: The no-profitable-warp frontier: required slash fraction $s^*(\lambda)$ by market depth D/S. Above $s^* = 1$ (dashed) no implementable slashing deters the coalition — the regime the depth rule excludes.

4 The glass monitor: an end-to-end sequential audit

Per-epoch slashing sees only per-epoch deviations: a sustained bias *within* the flagging threshold — the grind — is invisible to it and cumulatively large. The defense is sequential: monitor the divergence between the system's measured intensities and an *independent* instrument pricing overlapping risk, and alarm on persistent drift.

We construct the monitor on real data: the monthly credit intensities implied by sovereign sukuk spreads (the system leg) against independently quoted sovereign CDS (the audit leg) for seven countries over fourteen years (Ahmed & Bhuyan, 2026). Per-country correlations between the two legs are 0.989–0.998 — the two instruments price the same risk from different markets, which is precisely what makes the divergence informative. The divergence series $d_t = s^*_t - \mathrm{CDS}_t$ is standardized on a training window and fed to a two-sided CUSUM (Page, 1954).

Calibrating the alarm threshold on the clean historical series and then injecting synthetic warps into the *real* data from a fixed date gives the detection-lag curve (Figure 2): a 25bp sustained bias is detected at the *first* subsequent observation in every country; a 10bp bias within one observation in 71% of countries. The audit leg in our panel is annual (the CDS source publishes yearly), so lags are reported in years; a production monitor on daily feeds shrinks them proportionally. Two of seven clean country-series trip the aggressive threshold — on inspection these are real dislocation episodes, i.e. the monitor also detects genuine market stress; in production those alarms route to the stabilisation mechanism rather than the slashing module.

Proposition 3 (The grind bound)**.** *Against a monitor with detection-lag curve $\tau(b)$, the maximum extraction by any sustained bias is* $\max_b b \cdot D \cdot \tau(b)$*. On the real panel calibration this is* ≈ 10 *bp-years of repriceable notional, attained by the smallest detectable bias — the attacker's optimal warp is the slow one, and the monitor's threshold is what caps it.*

The monitor above is calibrated and stress-tested on the historical sukuk–CDS panel. Its companion validation runs the *same* construction inside a live agent-based protocol:

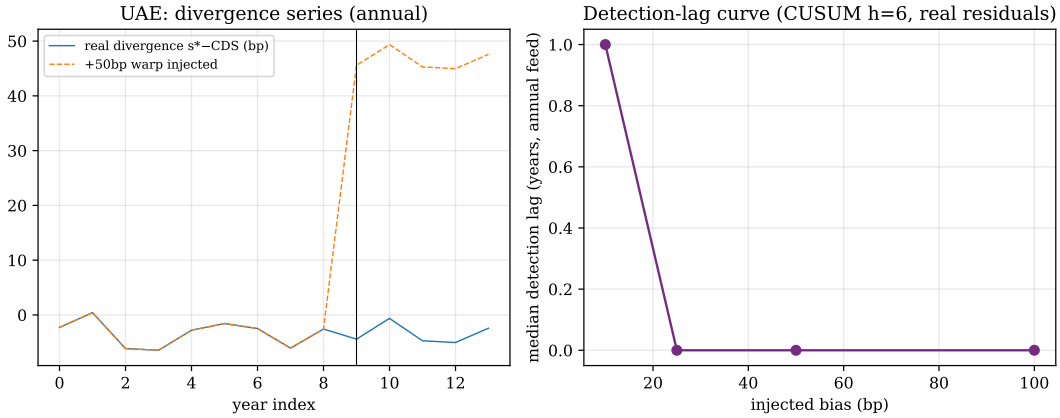

Figure 2: The glass monitor on real data. Left: a real divergence series with a +50bp warp injected (dashed). Right: median detection lag by injected bias (CUSUM on annual observations; daily feeds scale lags down proportionally).

a cadCAD model of the κ-substrate (Ahmed & Bhuyan, 2026b) embeds this two-sided CUSUM (reference slack $k = 0.5\sigma$, threshold calibrated to the 25bp warp) as the chain's on-line measurement-integrity monitor, with a frequent-batch-auction matching layer, a stake-weighted-median index, and the depth rule $D/S \leq 0.07$ of Section 3. Against sustained warps of 10–100bp it reproduces the detection-lag curve of Figure 2 dynamically (100% detection in 3–8 blocks, faster for larger warps) and confirms the grind bound empirically: with the depth rule capping per-block repriceable notional, attacker extraction before detection stays below \$21k for every detectable warp, while the naive last-trade mark — lacking the monitor — lets the warp run the full horizon. The analytic bound here and that live simulation are thus two views of one result: the slow warp is the only one that pays, and the monitor's threshold is what bounds it.

4.1 The cross-sectional layer: catching targeted warps from within

The independent-instrument leg catches *global* warps — biases applied to the whole published surface — at the audit instrument's frequency. But the profitable targeted attack is different: bias *one* issuer's measurements, to mis-mark one borrower's collateral for liquidation or cheap acquisition. For this attack the system audits itself. From the full-universe history panel (273 bonds with daily G-spreads, aggregated to 37 issuer-month $\hat{\kappa}$ series over 2015–2026) we build a *cross-sectional* monitor: each issuer's level residual from its fitted relationship to the panel median, standardized on a training window and fed to a per-issuer CUSUM. The panel median is immune to single-issuer manipulation, so it serves as an internal audit instrument that updates *monthly*. Injecting sustained targeted warps into the real series: a 50bp single-issuer warp is detected in a median of 5 months, 25bp in 9.5, and even 10bp in 11.5 — with 100% detection in every case (Figure 3). Six of thirty-seven issuers alarm on the *clean* history; inspection shows real idiosyncratic episodes (restructurings, sanctions), which a production system routes to review rather than slashing. The division of labour is now three-way: the depth rule and slashing govern fast attacks; the cross-sectional monitor catches targeted grinds from within, monthly; and the external independent instrument is needed *only* against globally uniform warps — which materially shrinks the residual trust assumption of Section 6.

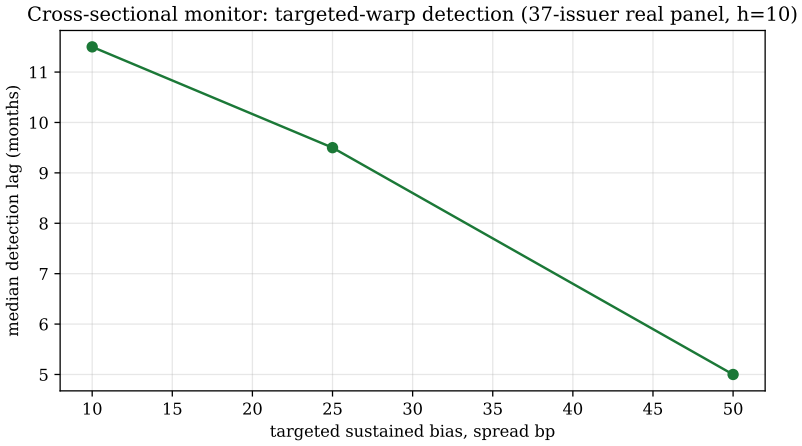

Figure 3: The cross-sectional monitor on the real 37-issuer panel: median detection lag for sustained single-issuer warps; detection rate 100% at all tested sizes.

5 The social layer: registry incentive design

Asset-backing rules (the system's existential-risk gate) bottom out in registrars: institutions attesting that a warehouse receipt, a sukuk certificate, a title corresponds to a real, deliverable asset. This layer is social — no cryptography verifies wheat — so its integrity must be purchased with the classic deterrence instruments (Becker, 1968; Shapiro & Stiglitz, 1984): bonds, random physical audits, and franchise value.

A registrar earning per-period franchise profit π_r (continuation value $C = \pi_r/(1-\beta_r)$, an equity claim — this is an interest-free system, but time preference exists) posts bond W and faces per-period audit probability p_a. An attacker offers a bribe up to the fraudulent value V it unlocks. Honesty requires $p_a(W + C) \geq V$:

Proposition 4 (Bond schedule and concentration cap)**.** *Single-registrar safety requires $W(V) \geq V/p_a - C$; equivalently, for any fixed bond, the attested value must satisfy $V \leq p_a(W + C)$. At realistic audit rates ($p_a \in [0.05, 0.10]$) the required bond is an order of magnitude or more above the attested value (the ratio is approximately $1/p_a$): single-registrar attestation does not scale. Under k-of-n quorum attestation (the attacker must bribe k registrars; any single audit kills the attestation), safety requires $k\,[1-(1-p_a)^k]\,(W+C) \geq V$, so safe attested value grows approximately as k^2 for small p_a.*

The design consequences for the architecture's whitelist: a per-registrar attested-value cap (the social layer's depth rule), quorum attestation for large assets in place of large bonds, and audit intensity proportional to attested value. Note the structural identity with Proposition 2: collateral $\geq$ value / detection, again.

6 Composition: local defenses do not add up — the monitor does

Suppose both preceding layers are calibrated exactly to their no-profit bounds. Local safety does not imply global safety:

Remark 2 (A super-additive mixed attack)**.** *Let the attacker run an in-band oracle bias (locally net-zero by Proposition 2's calibration)* and *attest fake backing at the registry cap (locally net-zero by Proposition 4). The warped intensity inflates the market value of the*

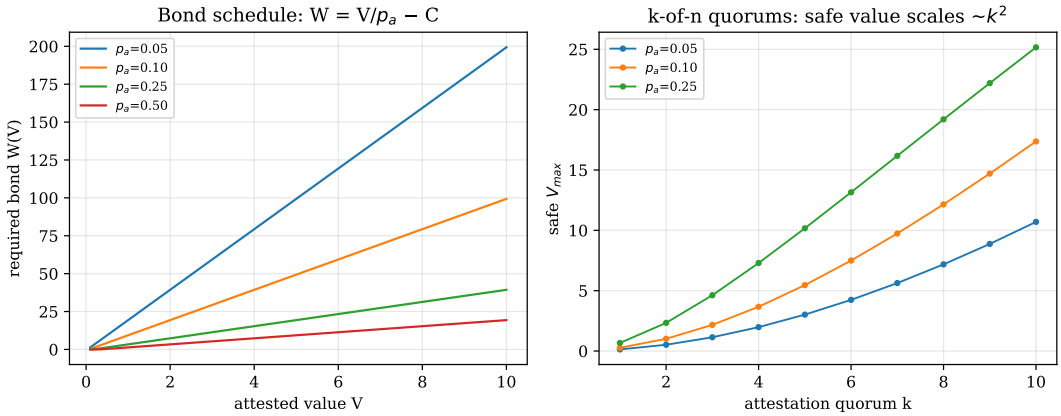

Figure 4: Registry design. Left: required bond against attested value by audit probability. Right: safe attested value under k-of-n quorums — quorums, not bonds, are how large assets scale.

fraudulently attested collateral: the attacker borrows or sells against assets that are both fake and *overpriced. The cross-term — relative price inflation times fake value — appears in neither layer's deterrence condition. At an illustrative calibration (a within-band bias inflating prices* 10%*, fake backing at the cap), the mixed attack nets strictly positive profit while every local condition holds. Layer-by-layer security composes sub-additively in defense, super-additively in attack.*

What restores composition is the end-to-end property of the monitor:

Assumption 1 (Independent instrument)**.** *At least one audit instrument prices the overlapping risk in a market the attacker cannot move at relevant scale.*

Theorem 1 (End-to-end composition bound)**.** *Under Assumption 1, the divergence between system prices and the independent instrument reflects the* total *distortion — the sum of oracle bias, registry-induced mispricing, and any cross-terms — regardless of which layers produced it. Consequently the monitor's grind bound (Proposition 3) applies to the* mixed *attack: total undetected extraction is bounded by* $\max_b b\,D\,\tau(b)$ *plus the one-shot pre-detection caps of the local layers. Local calibrations bound per-layer profits; the end-to-end monitor bounds their composition.*

Proof sketch. Every layer's distortion, to reach the attacker's payoff, must move a system price: the oracle bias moves $\hat{\kappa}$ directly; registry fraud moves the priced collateral pool; cross-terms move both. The independent leg, by Assumption 1, does not move. Hence the divergence statistic accumulates the sum of all price-relevant distortions, and the CUSUM bound on sustained divergence applies to that sum. One-shot components are bounded by the per-layer caps already derived. □

Remark 3 (The residual trust anchor)**.** *Assumption 1 is irreducible: a system all of whose reference instruments the attacker can move is unauditable from within. The practical defense is* diversity *— multiple audit legs (conventional bond bases, CDS from distinct venues, realized default records) so that moving all of them simultaneously exceeds any plausible budget — and the assumption's scope is narrower than it first appears: the cross-sectional monitor of Section 4 polices targeted distortions from within, so external independence is required only against* globally uniform *warps. This is the precise, non-rhetorical sense in which a measurement-based monetary system still rests on something it must trust: not a committee, but the continued existence of at least one honest mirror.*

7 Limitations

The oracle model treats honest noise as Gaussian and one-sided piling as the optimal coalition strategy; richer submission strategies (timed, correlated with volatility) deserve adversarial search of the kind applied to the liveness mechanism in Ahmed (2026a). The monitor's audit leg in our panel is annual; production calibration on daily feeds remains to be done, and alarm-routing (slashing vs stabilisation) on real dislocations needs design. Registrar parameters (franchise values, audit costs) are normalized, not estimated. The composition theorem bounds price-mediated extraction only: attacks whose payoff never routes through a system price (pure exfiltration of audit data, governance capture for non-pecuniary ends) are outside the model. The independence assumption is stated, not proven — it is the residual trust anchor, and we prefer to name it than to hide it.

8 Conclusion

When the policy variable is a measurement, the attack surface is the sensor. We have priced that surface layer by layer: a depth rule capping what an oracle may reprice relative to the stake behind it; a bond-and-quorum rule making registry fraud cost more than it unlocks; a sequential end-to-end monitor — calibrated here on fourteen years of real cross-instrument data — that detects fast warps at the first print and caps the slow ones; and a composition theorem showing why the monitor, not the local defenses, is what makes the layers add up. One principle runs through every layer: collateral at risk must exceed value-at-measurement over detection probability. Ancient monetary law made the integrity of the measure the sovereign's first duty; in a measurement-based system that duty becomes a calibration — and, like all calibrations, it can be checked.

References

Ahmed, S. (2026a). Liveness under macro winters: Mechanism design for fee-only blockchains. SSRN Working Paper 6924459.

Ahmed, S. (2026b). The κ-Chain: A sovereign Islamic financial infrastructure layer. Technical whitepaper v0.7.6, SSRN 6924600.

Ahmed, S., & Bhuyan, R. (2026). An Islamic credit default swap on smart-contract infrastructure: Design, pricing, and regulatory pathway. *SSRN Working Paper* 6323519.

Ahmed, S., & Bhuyan, R. (2026b). A simulation framework for the κ-substrate: agent-based and mechanism-design validation of the on-chain convergence intensity. *SSRN Working Paper* 6323618.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169–217.

Duffie, D., & Stein, J. C. (2015). Reforming LIBOR and other financial market benchmarks. *Journal of Economic Perspectives*, 29(2), 191–212.

Gandal, N., Hamrick, J. T., Moore, T., & Oberman, T. (2018). Price manipulation in the Bitcoin ecosystem. *Journal of Monetary Economics*, 95, 86–96.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2), 100–115.

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74(3), 433–444.