Liveness under Macro Winters: Mechanism Design for Fee-Only Blockchains

Shehzad Ahmed*

June 2026

Abstract

Every blockchain whose consensus layer is financed purely by transaction fees — by design (zero-emission chains) or by destiny (Bitcoin as its block subsidy decays) — faces a liveness hazard that emission-funded chains conceal: in the aftermath of a persistent macroeconomic shock, transaction volume can remain below validator operating cost for months. We show the binding risk is the *duration* of such "winters," not their depth, and that the intuitive remedy — counter-cyclical fee multipliers — is both pro-cyclical for users and manipulable by the validators who produce the volatility measurements. We propose instead a pre-funded mutual support pool (a *liveness sleeve*) with three design commitments: a *conjunction trigger* (volume and a stress index, each sustained, so that the one signal validators control never suffices alone); a *shortfall payout* capped at a published reference operating cost; and a slow payout decay whose role, we show, is not to discourage idleness but to make a validator's receipts strictly increasing in its own honest fee income — without decay, censoring one's own fee base is payoff-neutral inside the subsidised state. Under the shortfall structure, censorship aimed at forcing the trigger is strictly dominated whenever the support floor lies below normal income (Proposition 1). Monte-Carlo experiments on six-year paths with heterogeneous validators locate the feasible design window — replacement ratio $\beta \in [1.2, 1.6]$ of reference cost with decay at most 4% per week — and a reinforcement-learning adversary (a 35%-stake cartel free to censor up to 75% of transactions) converges to near-zero censorship, never beating honesty on matched evaluation episodes. Survival of a thirty-validator set through a black-swan winter (volume −85%, half-life twelve months) rises from 56% to 77% with the sleeve in place. The mechanism requires no emission, no fee increases on distressed users, and no discretionary bailout authority.

Keywords: blockchain security budget, validator economics, mechanism design, fee markets, liveness, mutual insurance, zero-emission consensus. **JEL:** D82, G22, L86.

1 Introduction

The long-run security of a proof-of-stake or proof-of-work blockchain rests on a simple flow constraint: the consensus layer's income must cover its operating cost. Most chains today satisfy the constraint with predetermined issuance — inflationary block rewards that pay validators for the passage of time regardless of network activity. A growing class of designs rejects issuance, on monetary grounds (zero-emission "sound money" designs; interest-free substrates in which a predetermined time-based return is prohibited by construction (Ahmed, 2026b)) or has it taken away by schedule: Bitcoin's block subsidy halves toward zero, and

*Independent University, Bangladesh. The mechanism analysed here originates in the validator-economics design of an interest-free Layer-1 specification (Ahmed, 2026b); the liveness problem and the mechanism are general to any chain whose consensus rewards are funded by transaction fees alone.

the question of whether fees alone can fund its security — the *security budget problem* — is well posed in the literature (Carlsten et al., 2016; Budish, 2018; Easley et al., 2019).

That literature has concentrated on two failure modes: strategic deviations in block production when fee income is volatile at the timescale of blocks (Carlsten et al., 2016), and the adequacy of steady-state fee income against attack costs (Budish, 2018). This paper analyses a third, slower failure mode that emission-funded chains never exhibit and the fee-market literature has largely passed over: the **macro winter**. A severe macroeconomic shock produces, first, a burst of activity — liquidations, de-risking, repricing — during which fee income *rises*; and then a long aftermath in which volume settles far below its prior level and recovers only slowly. For a fee-only consensus layer, the acute phase is self-financing; the aftermath is the threat. If the winter outlasts validators' working capital, the chain loses operators precisely when the remaining operators' incentives are weakest, compounding censorship risk, liveness risk, and stake concentration.

Two intuitive remedies fail. *Counter-cyclical fee multipliers* — raising protocol fee floors during stress to keep validators whole — are pro-cyclical for users (they tax distressed transactors to preserve infrastructure margins) and create a perverse measurement problem: the validators who would benefit from the multiplier are the same agents who produce, order, and can distort the on-chain measurements of stress. *Discretionary treasuries* reintroduce the trusted bailout authority that the architecture was designed to exclude. We propose and analyse a third route: a rule-based, pre-funded mutual support pool — a *liveness sleeve* — filled counter-cyclically from fees in good times and disbursed under a conjunction trigger with a shortfall payout in certified winters.

Contributions. *First*, we formalise the winter-liveness problem and show its binding dimension is shock persistence, not depth: the support pool a chain needs diverges as the volume process's autocorrelation approaches unity, mirroring (and inheriting the mathematics of) buffer-sizing results for macroeconomic stabilisation funds (Ahmed & Bhuyan, 2026). *Second*, we give the mechanism three commitments whose interaction is the design's substance: (i) a conjunction trigger over volume and an independent stress index, each sustained over a hysteresis window, with *fixed* governance constants rather than trailing quantiles (trailing windows are gameable by slow grinding); (ii) a shortfall payout $\max(0,\, \beta \cdot \mathrm{OpEx}_{\mathrm{ref}} - y_i)$ against a *published* reference cost, performance-weighted by consensus-measurable work; (iii) a slow payout decay. *Third*, we prove the central incentive result (Proposition 1): under the shortfall structure, volume censorship aimed at forcing or exploiting the trigger is strictly dominated whenever the support floor lies below normal income — and we show the decay parameter's true role is to break the payoff-neutrality of self-censorship *inside* the subsidised state, not to discourage idleness. *Fourth*, we quantify the design window by simulation (six-year paths, heterogeneous validators, joint volume-and-stress crises) and stress the incentive claim with a learned adversary: a PPO-trained 35%-stake censor cartel converges to near-zero censorship and never beats honesty on matched episodes. *Fifth*, we derive sizing guidance: the pool's cap should be set by the tail winter's *duration* (twelve months of reference cost in our calibration), and the replacement ratio $\beta > 1$ should be read as a dispersion allowance over the audited validator cost distribution, not a profit margin.

Related work. The security-budget line begins with Nakamoto (2008)'s subsidy schedule and its consequences: Carlsten et al. (2016) show fee-only mining is unstable at the block timescale; Budish (2018) bounds the steady-state security a fee flow can buy; Easley et al. (2019) document the economics of the fee market's emergence. The fee-mechanism line — EIP-1559 and its analyses (Buterin et al., 2019; Roughgarden, 2021), and equilibrium fee formation (Huberman et al., 2021) — treats congestion pricing within a period, not

income insurance across periods; our mechanism is complementary and operates at a different timescale. On the funding side, Saleh (2021) analyses proof-of-stake without wasteful issuance. The sleeve's mutual, donation-based structure has a long institutional pedigree in mutual insurance; its specific instantiation here originates in a takaful (mutual protection) pool specified for an interest-free Layer-1 (Ahmed, 2026b), where a predetermined validator return would violate the substrate's own monetary constraint — a case in which the incentive requirement and the monetary-design requirement coincide (Section 6).

2 The winter-liveness problem

Consider a chain whose consensus layer earns only transaction fees. Normalise aggregate fee income in the normal state to 1 per period, of which a fixed wakala-style share w (here $w = 0.6$) accrues to validators; let aggregate reference operating cost be $\mathrm{OpEx}_{\mathrm{ref}}$ (here 0.36, a 40% normal-state margin). Validator i bears cost c_i with $\sum_i c_i = \mathrm{OpEx}_{\mathrm{ref}}$, finances deficits from working capital up to a credit line (two months of own cost in our calibration), and exits when the line is exhausted. Volume follows a log-AR(1) with rare crisis jumps; a stress index κ_t (in our calibration, the chain's published credit-intensity state, with monthly autocorrelation 0.774 estimated from real instrument panels (Ahmed & Bhuyan, 2026)) jumps jointly with crises.

Two features structure the problem. *First*, the acute phase of a crash is self-financing: liquidations and de-risking raise fee income exactly when volatility peaks. The threat is the aftermath — volume 60–85% below normal with a recovery half-life of months. *Second*, the required support diverges in persistence: for a winter with depth D and recovery half-life h, the cumulative income shortfall scales like $D \times h$, so a pool adequate for moderate winters fails in persistent ones at any depth. It is always the length of the winter, not its severity, that exhausts a finite reserve. Both features are inherited by any insurance mechanism built on top, and both militate against financing the gap out of *contemporaneous* user fees.

3 The mechanism

Definition 1 (Liveness sleeve)**.** *A liveness sleeve is a protocol-owned pool S_t with the following rules.* ***Fill****: in any period with volume above its long-run median and the sleeve inactive, a skim s (here 10%) of fees accrues to S_t, stopping at cap $\bar{S} = C$ months of* $\mathrm{OpEx}_{\mathrm{ref}}$*; a genesis allocation seeds part of the cap.* ***Trigger****: the sleeve activates only if measured volume is below $\underline{V}$ for K consecutive periods* and *the stress index exceeds a fixed constant $\bar{\kappa}_{p99}$ for K consecutive periods.* ***Payout****: while active, validator i with fee income $y_{i,t}$ receives*

$$\pi_{i,t} \;=\; \omega_{i,t}\,\max\!\big(0,\;\beta\,c_i^{\mathrm{ref}} - y_{i,t}\big)\,(1-d)^{\,\tau_t},$$

where c_i^{ref} is the published reference cost, $\omega_{i,t} \in [0,1]$ a performance weight built from consensus-measurable work (blocks signed, vote participation, oracle submissions within tolerance of the stake-weighted median, uptime), d a weekly decay, and τ_t weeks since activation. ***Exit****: deactivation requires volume above $1.3\,\underline{V}$ sustained one week, then a one-month cooldown.* ***Governance****: $(\underline{V}, \bar{\kappa}_{p99}, \beta, d, C)$ are fixed constants amendable by ordinary governance within board-bounded invariants; the reference cost is set by published methodology, not validator self-report.*

Three commitments deserve comment before the analysis. The *conjunction* exists because measured volume is the one trigger input validators control: a cartel can censor transactions to suppress it. An independent, manipulation- expensive stress leg (here, a stake-weighted-median oracle state whose falsification requires majority collusion and is slashable ex post)

must therefore co-sign every activation. The *fixed constants* exist because trailing-quantile thresholds — the adaptive alternative — can be ground down slowly by sustained, mild suppression. The *shortfall* form of the payout is the heart of the mechanism, as the next section shows.

4 Censorship is strictly dominated

Fix a cartel controlling stake share λ and consider any strategy that suppresses a fraction $\chi_t \in [0,1]$ of network transactions in period t. Censored transactions do not execute, so the cartel's fee income falls one-for-one with the fee base it destroys. Let $y_t(\chi)$ denote aggregate validator fee income under suppression χ, strictly decreasing in χ.

Proposition 1 (Shortfall dominance)**.** *Suppose the support floor lies below normal income, $\beta\,\mathrm{OpEx}_{\mathrm{ref}} < y^{\mathrm{normal}}$, and the decay satisfies $d > 0$. Then for every period and state: (i) outside the active state, any $\chi_t > 0$ strictly reduces the cartel's payoff; (ii) inside the active state, the cartel's receipts $y_t(\chi) + \max(0,\ \beta\,\mathrm{OpEx}_{\mathrm{ref}} - y_t(\chi))(1-d)^{\tau}$ are strictly increasing in $y_t(\chi)$, hence strictly decreasing in χ_t; and (iii) any strategy that incurs censorship costs to* cause *activation is strictly dominated by honesty, since the payout in the activated state is capped below the income the cartel destroys to reach it.*

Proof. (i) Outside activation there is no payout; receipts equal fee income, which is strictly decreasing in χ. (ii) Inside activation, receipts as a function of own income y are $R(y) = y + (\beta\,\mathrm{OpEx} - y)\,\delta$ on $y < \beta\,\mathrm{OpEx}$ with $\delta = (1-d)^{\tau} < 1$, so $R'(y) = 1 - \delta > 0$; and $R(y) = y$ above the floor. Receipts are therefore strictly increasing in income everywhere, and censorship, which only destroys income, strictly reduces them. (iii) To move the system from inactive to active, the cartel must hold measured volume below $\underline{V}$ for K periods, destroying at least $\lambda\, w\,(1-\underline{V})\, y^{\mathrm{normal}} K$ of its own income with zero payout during the approach (the sleeve is inactive), to reach a state whose per-period receipts are bounded by $\beta\,\mathrm{OpEx}_{\mathrm{ref}} < y^{\mathrm{normal}}$ and declining in τ. Every such path is pointwise dominated by the honest path. □

Remark 1 (The decay's true role)**.** *At $d = 0$, part (ii) fails at the margin: $R'(y) = 0$ below the floor, so a cartel already inside the active state is* indifferent *to destroying its own fee base — censorship becomes costless at the margin, and the mechanism loses its strict incentive to keep producing blocks that carry fees. The decay's purpose is precisely to make receipts strictly increasing in honest income inside the subsidised state. Its anti-idleness framing is secondary: in our simulations the trigger's false-positive rate against ordinary downturns is zero, so the "lazy validator in a minor dip" never reaches the sleeve at all. This inverts the Bagehot intuition with which we began the design: penal pricing of the support is neither necessary (the conjunction gate and the performance weights carry the selection burden) nor sufficient (aggressive decay guts protection in long winters); what is essential is that support never exceed the floor and that receipts always rise with honest work.*

5 Quantitative design

5.1 Reference model and calibration

We simulate six-year paths at hourly resolution (and replicate at daily resolution inside a cadCAD engine (Ahmed, 2026a)) with thirty validators of lognormally dispersed cost, the volume and stress processes of Section 2, and Poisson crises: minor (volume −35%, half-life one month, roughly annual) and major — the winters — in three severities: moderate (−65%, half-life six months), severe (−75%, nine months), black swan (−85%, twelve months).

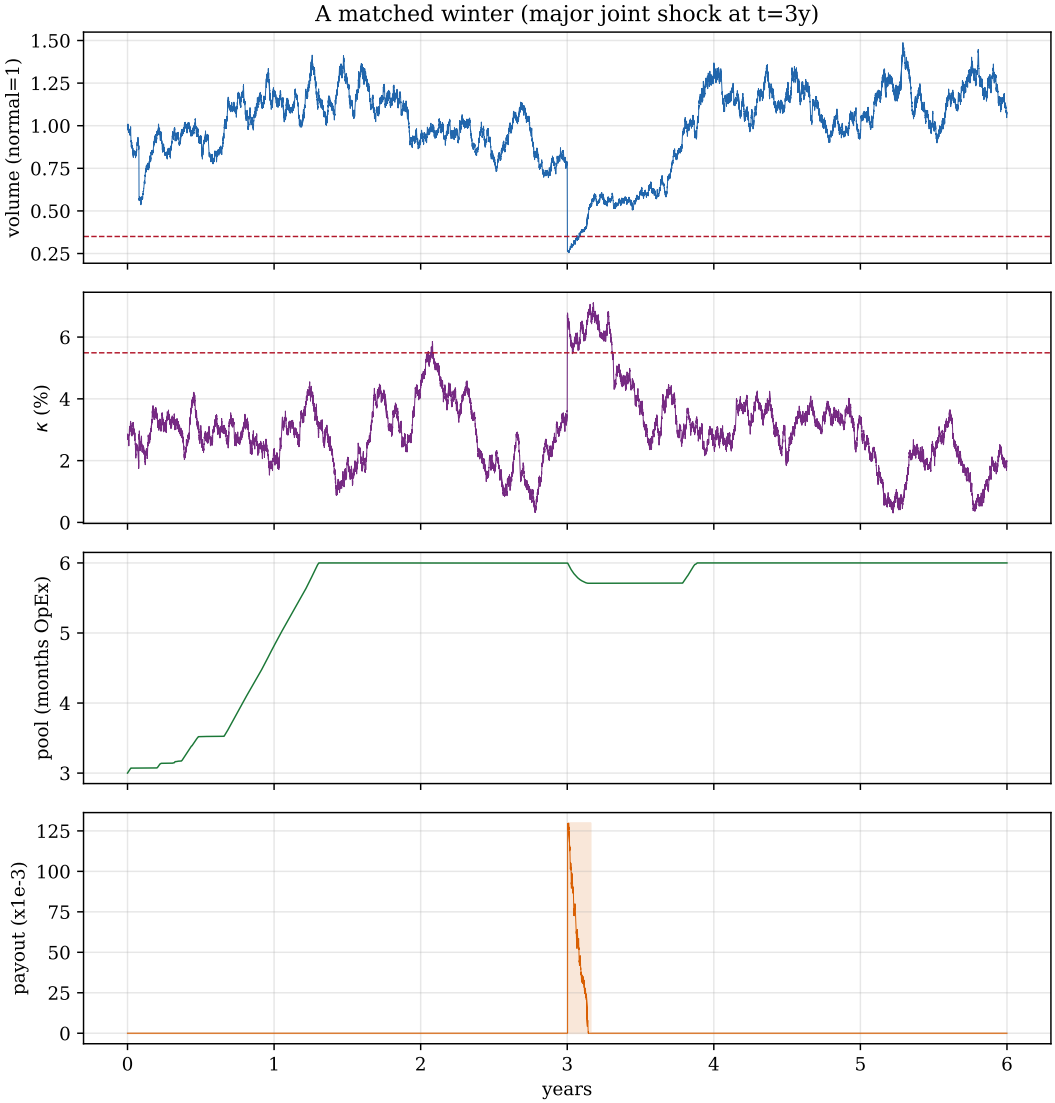

Figure 1: A matched severe winter. Top to bottom: volume against the trigger floor $\underline{V}$; the stress index against its $p99$ constant; the pool in months of reference cost; payouts during the activation episode.

The stress index is calibrated to the full-universe panel-median intensity series (273 bonds aggregated to a 127-month index, 2015–2026): measured monthly AR(1) $\rho = 0.921$ — an *8.4-month* half-life, substantially more persistent than single-issuer estimates, confirming that winters are the relevant regime — with the $p95/p99$ governance constants taken from the real distribution. The real decade contains two stress episodes: the 2020 episode (peak jump +305bp of spread, four months above $p95$, measured decay half-life three months from peak) and a 2016 episode (+200bp, slower decay). The measured 2020 episode is, in intensity units, roughly 2.5× *larger* than the scenario severe winter (+762bp vs. +300bp of κ) but decays much faster (three months vs. nine); both are run below. Volume remains a scenario input, as no fee-only chain history yet exists. Figure 1 shows one matched winter: the joint shock, the pool's drawdown, and the payout episode.

5.2 Trigger discrimination and survival

Across all simulated paths without a true winter, the conjunction trigger fired **zero** times: ordinary downturns never reach the sleeve, so no minor-downturn moral hazard arises at the trigger stage. History gives the same answer: on the real 127-month index, the κ gate was open (sustained above $p99$) in just **two months of the decade** (1.6%; seven months at $p95$) — outside those windows no volume manipulation, however aggressive, could have fired the sleeve. Through forced winters at the validated parameters ($\beta = 1.3$, $d = 2\%$/week, cap twelve months, genesis seed four months), survival of the validator set improves monotonically in severity — the sleeve is a tail instrument. The table reports the daily cadCAD implementation; the hourly reference model yields the same ordering with survival several points higher in every cell (daily aggregation deepens effective drawdowns against a fixed credit line), and all qualitative conclusions — the monotone severity gradient, the near-exhaustion of the pool only in the black-swan case, and the location of the feasible window — are common to both:

Winter	without sleeve	with sleeve	Δ
moderate (−65%, hl 6mo)	85.1%	90.2%	+5.1pp
severe (−75%, hl 9mo)	72.6%	84.4%	+11.8pp
black swan (−85%, hl 12mo)	56.3%	76.5%	+20.2pp

The pool approaches exhaustion only in the black-swan case, which is what sizes the cap: a six-month cap — adequate for every moderate winter — fails exactly in the persistent tail, the quantitative form of the persistence-divergence point of Section 2. The *measured* 2020 episode makes the same point with real data: replayed through the model (+762bp measured jump, measured three-month decay), it is nearly harmless — 96% of validators survive *without any sleeve* — despite a stress jump 2.5× the scenario severe winter's. A huge fast shock is survivable; the dangerous winter is the slow one. Persistence, not magnitude, is the binding risk — now an observation, not only a simulation.

5.3 The feasible window

Sweeping (β, d) on severe winters under the joint criterion *survival* $\geq 90\%$ *and subsidised income* $\leq 85\%$ *of normal* locates the feasible window at $\beta \in [1.2, 1.6]$, $d \leq 4\%$/week (Figure 2). Two readings matter. First, our own initial calibration — $\beta = 0.8$ with 10%/week decay, the Bagehot-flavoured "penalty rate" — lies outside the window: it is incentive-safe but protects almost nothing in long winters, improving severe-winter survival by under two percentage points. Second, $\beta > 1$ is not profit: the reference cost is an aggregate, and the upper tail of a lognormal cost distribution needs 1.2–1.4× the reference to reach its own floor. The binding constraint throughout the window is protection, not incentives — subsidised income runs at 71–77% of normal at $\beta = 1.3$, so the preference inversion (no validator prefers the subsidised state) holds with room to spare.

5.4 A learned adversary

Proposition 1 covers censorship strategies; simulations can stress its premises against strategies we did not enumerate. We train a PPO agent (Schulman et al., 2017) playing a 35%-stake cartel that chooses, daily, a censorship intensity in $\{0, 25\%, 50\%, 75\%\}$, observing measured volume, the stress ratio, the sleeve state, and the pool level, and rewarded with its stake share of total validator income (fees plus payouts). Episodes run three years; half contain a severe winter, so exploitable stress states, if any, appear in training. After 150,000 steps the learned policy converges to mean suppression 1.4% — effectively honesty — and on thirty matched evaluation episodes it beats the zero-censorship policy **never** (mean per-episode

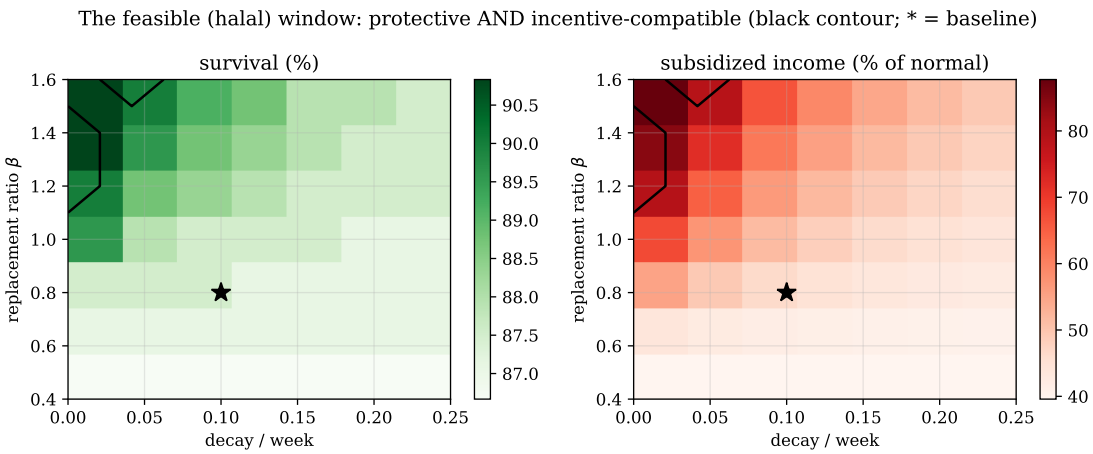

Figure 2: The feasible window on severe winters: survival (left) and subsidised-income ratio (right) over (β, d); the contour bounds the jointly feasible region; the star marks the initial Bagehot-flavoured calibration, which fails the protection criterion.

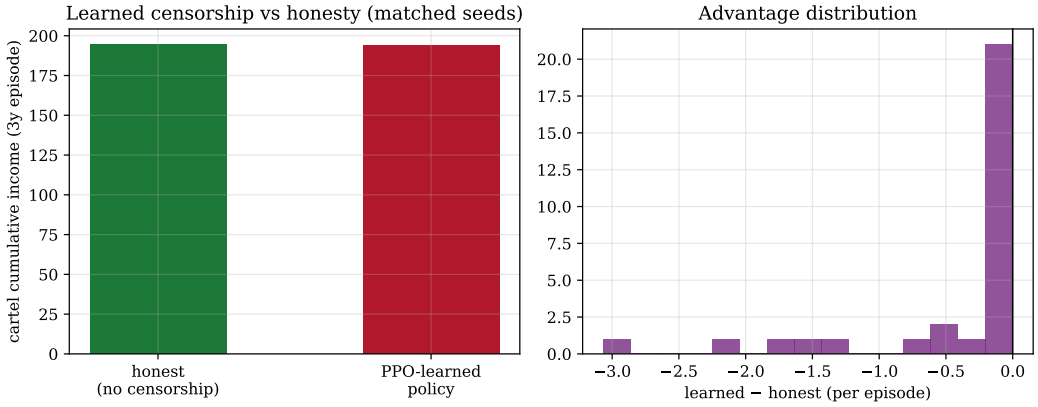

Figure 3: The learned censor-cartel versus honesty on matched seeds: cumulative cartel income (left) and the distribution of the learned policy's advantage (right) — negative in every episode.

disadvantage −0.39 income units; Figure 3). The agent's verdict agrees with the proposition: the trigger can sometimes be forced, but forcing it never pays.

5.5 Reproducibility

The hourly reference model, the daily cadCAD implementation (partial state update blocks), the adversary environment and PPO training script, the trained policy, the parameter grids behind Figure 2, and all figure generators are included in the replication package accompanying this paper; every number in this section is reproducible from a single script invocation per experiment.

6 When incentive design and monetary design coincide

Run the mechanism's failure mode in reverse. A badly parameterised sleeve — a generous floor, no decay, stake-weighted rather than performance-weighted disbursal — pays validators a predetermined amount for the passage of time, regardless of work done. That is, structurally,

an inflationary block reward funded from a pool rather than from issuance: exactly the predetermined time-based return that zero-emission designs exist to exclude, and that interest-free monetary substrates prohibit as a matter of construction (Ahmed, 2026b). The conditions that keep the sleeve incentive-compatible — payments that are outcome-contingent (shortfall only), earned (performance-weighted), bounded (capped at audited cost, decaying), and exceptional (conjunction-triggered) — are the same conditions that keep it from being a disguised fixed return. We do not claim this coincidence is deep in general; we observe that in this mechanism the incentive-compatible point and the no-predetermined-return point are the same point, which suggests that "returns must track contemporaneous real work" is doing the load-bearing work in both arguments.

7 Limitations

These results are design-stage. The volume process is a scenario calibration — no fee-only chain has yet lived through the winters we simulate — though the stress side is now fully measured: the index dynamics, the governance constants, and the major-winter magnitude are taken from the 127-month full-universe record rather than assumed. Validator cost dispersion is assumed lognormal; the reference-cost audit process is modelled as exact. The adversary's strategy class is censorship intensity; selective censorship, timing games against the cooldown and hysteresis windows, oracle-leg manipulation by majority coalitions, and collusion with off-chain positions are future work, as is an implementation in a production consensus engine. Proposition 1 is robust to these extensions in one direction only: any strategy whose effect on the trigger passes through destroying the cartel's own fee base inherits the dominance argument; strategies that move the stress leg without touching fee income do not, which is why the stress leg must remain majority-collusion-expensive and slashable.

8 Conclusion

Fee-only consensus layers — whether by zero-emission design or by subsidy decay — face a winter-liveness problem whose binding dimension is persistence. A pre-funded liveness sleeve with a conjunction trigger, a shortfall payout against a published reference cost, and a slow decay protects the validator set through tail winters (black-swan survival +20 percentage points in our experiments) while making censorship strictly dominated and teaching a learned adversary to be honest. The design inverts the Bagehot instinct: the support should be *adequate and strictly increasing in honest work*, not punitive; selection belongs to the trigger, not the price. The cap belongs to the tail winter's duration, and the replacement ratio to the cost distribution's dispersion. None of it requires issuance, discretion, or taxing distressed users — which is what a consensus layer built to exclude predetermined returns requires, and, we suspect, what Bitcoin's fee-only future will eventually need as well.

References

Ahmed, S. (2026a). An integrated simulation framework for decentralised finance protocols: cadCAD, reinforcement learning, game theory, and mechanism design. *SSRN Working Paper* 6323618.

Ahmed, S. (2026b). The κ-Chain: an interest-free ($\iota = 0$) Layer-1 monetary substrate. Technical whitepaper, v0.7.6 (design specification).

Ahmed, S., & Bhuyan, R. (2026). Interest-free monetary equilibrium: a general-equilibrium foundation for the κ-Standard. Working paper.

Budish, E. (2018). The economic limits of Bitcoin and the blockchain. *NBER Working Paper* 24717.

Buterin, V., Conner, E., Dudley, R., Slipper, M., Norden, I., & Bakhta, A. (2019). EIP-1559: Fee market change for ETH 1.0 chain. Ethereum Improvement Proposals.

Carlsten, M., Kalodner, H., Weinberg, S. M., & Narayanan, A. (2016). On the instability of Bitcoin without the block reward. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 154–167).

Easley, D., O'Hara, M., & Basu, S. (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, 134(1), 91–109.

Huberman, G., Leshno, J. D., & Moallemi, C. (2021). Monopoly without a monopolist: An economic analysis of the Bitcoin payment system. *Review of Economic Studies*, 88(6), 3011–3040.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. White paper.

Roughgarden, T. (2021). Transaction fee mechanism design. In *Proceedings of the 22nd ACM Conference on Economics and Computation* (p. 792).

Saleh, F. (2021). Blockchain without waste: Proof-of-stake. *Review of Financial Studies*, 34(3), 1156–1190.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv:1707.06347*.