Token-Based Platform Governance and Agency Conflicts*

Michele Fabi† Valerie Laturnus‡ Romain Rossello§

July 16, 2026

Abstract

Proxy voting and delegation are often viewed as mechanisms that improve governance efficiency. We show that they can instead create instability when delegates strategically respond to endogenous changes in the governance environment. We develop a dynamic principal-agent model of decentralized governance in which tokenholders delegate control rights to informed representatives whose types are initially unknown. In equilibrium, malicious delegates exert high effort early to build reputation, expand the platform, and increase future gains from misconduct. Observing effort and growth, tokenholders rationally reduce monitoring, raising delegates' returns to opportunism and making platform failures most likely after periods of apparent success.

JEL classification: D72, D82, D86, G34, G23, L22

Keywords: Tokenomics, Decentralized Governance, Poisson Voting, Stochastic Games

*We thank Philip Dybvig, Evgeny Lyandres, Basil Williams (discussant), Hanna Halaburda, Andrea Barbon, Louis Bertucci, and all other participants at the 6th CBER Conference, Université Paris-Dauphine-PSL Tech 4 Finance on AI and Blockchain, and 2026 Frontiers of DeFi workshop for their helpful comments. We acknowledge financial support from the Institut Europlace de Finance. An earlier version of this paper was circulated under "Delegated Control, Agency Conflicts and Endogenous Governance Cycles in Token-Based Platforms".

†Telecom Paris-CREST (Institut Polytechnique de Paris), 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France. ✉:Michele.FABI@ensae.fr.

‡Durham University, Business School, Mill Hill Lane, Durham DH1 3LB, United Kingdom. ✉:valerie.laturnus@durham.ac.uk.

§HEC Paris, 1 Rue de la Libération, 78350 Jouy-en-Josas, France. ✉:romain.rossello@hec.edu.

1 Introduction

Digital platforms have become central to coordinating economic activity. They match users (buyers and sellers), set participation rules, and control key conditions of exchange. In traditional platforms, these decisions are typically made by a corporate entity whose authority is tied to equity ownership. Blockchain technology has introduced an alternative organizational form. Token-based systems, including blockchains, issue their own digital currencies or tokens that allocate control rights among their users.[1] Because tokens provide strong incentives for participation and growth, tokenised platforms have scaled much faster than traditional platforms (Catalini, 2017), with some of the most successful examples generating $63 million in daily revenue (as of July 2026).

Prior literature has recognized that tokenization fosters commitment and network effects, mitigates moral hazard in governance, and thereby finances platform development at launch (Cong et al., 2022; Chod et al., 2022; Li and Mann, 2025). This paper studies the dynamic process of platform building over time after tokens have been issued and control rights have been allocated. Tokenholders are often dispersed, uninformed, or inactive, and therefore delegate voting power to representatives. Similar to proxy voting in corporate governance, dispersed owners often face weak incentives to acquire information and participate directly in every decision, so voting influence shifts toward more informed or specialized agents (Bhandari et al., 2021; Malenko and Shen, 2016; Malenko and Malenko, 2019). As a result, it becomes easier for a small number of private actors to gain influence and control key economic decisions within the blockchain ecosystem (Ferreira et al., 2023). This centralization becomes a central monitoring problem once their incentives diverge from those of tokenholders.

The paper shows that delegated control in decentralized governance creates a dynamic agency problem, i.e, early reputation-building and platform growth make delegation valuable, but later concentrated control can undermine governance discipline. The question is relevant for market participants and regulators because token-governed systems often control economically important rules and resources

[1] Examples include decentralized exchanges such as Uniswap (Lehar and Parlour, 2025), lending platforms such as Aave and Compound (Lehar and Parlour, 2025), staking services that allow users to support blockchain infrastructure and security (Lehar et al., 2026), such as Lido, and settlement infrastructure such as Arbitrum and Optimism (Harvey et al., 2025).

without a single point of accountability.

The model considers a dynamic token-based platform populated by tokenholders and delegates. Tokenholders enter when the platform offers sufficiently high expected surplus and exit after liquidity shocks. They hold governance tokens and participate in elections that determine whether the current leading delegate remains in control or is replaced by a challenger. Delegates make decisions that affect platform quality and the treasury[2], exert costly effort, and accumulate reputation through histories of good governance participation. Some delegates are malicious and can attack the platform by extracting platform resources and reducing platform quality. Tokenholders do not observe delegate types and therefore use observed histories to update beliefs about whether a delegate is aligned with the platform. The interaction among entry, monitoring, voting, delegate effort, reputation, attacks, and market clearing defines a stationary Markov equilibrium of the model.

The first main result is that pseudonymous re-entry can undermine effort-based screening. In classical principal-agent models, performance-based screening requires a single-crossing condition (Spence, 1973; Rothschild and Stiglitz, 1976). The aligned type should have a higher marginal return to effort than the misaligned type. In our framework, pseudonymous re-entry can invert this condition. A malicious delegate can mimic the effort of an honest delegate while preserving the option to extract the treasury in the future. Because misconduct does not permanently exclude the delegate from future participation, platform growth increases not only continuation rents but also future extraction opportunities. As a result, the malicious delegate's marginal return to effort can exceed that of the honest delegate. Observed effort and performance need not separate types, and in the option-preserving region, malicious delegates may exert more effort than honest delegates. The standard single-crossing condition is restored only in an authenticated benchmark, where delegate identity is permanent and misconduct leads to exclusion.

The second main result is that reputation becomes an imperfect disciplining device. Longer activity without misconduct signals reliability because aligned delegates have a stronger reason to maintain the same identity and participate over a longer horizon. Under pseudonymous re-entry, however, this signal remains imperfect. A malicious delegate can build reputation strategically, use it to obtain or retain

[2]The pool of funds that the platform manages to finance its activities.

control, and attack once the current prize becomes sufficiently large relative to the value of continued office. Unlike standard reputation models (Holmström, 1979, 1999; Kreps and Wilson, 1982), we find that reputation disciplines behavior in some states, but can also help malicious delegates increase the value of future misconduct.

The third main result is that platform growth weakens monitoring through declining pivotality. Tokenholders do not receive a direct monetary payment for participating in governance. They acquire information only to the extent that better governance affects the value of their participation and token holdings. As the platform grows, each tokenholder becomes small relative to the aggregate voting population. Her probability of changing the governance outcome falls, and so does her private incentive to pay attention (Palfrey and Rosenthal, 1985). We formalize this logic by modeling delegate elections as a Myerson Poisson voting game (Myerson, 1998). Tokenholders can pay an attention cost to learn which delegate has the higher reputation. They do so only when the pivotality-weighted benefit of information exceeds the attention cost. In the model, pivotality decays with platform scale, so larger and more successful platforms are precisely those in which individual monitoring incentives are weakest.

Attack incentives connect the two frictions. A malicious delegate compares the current prize from expropriating the treasury with the franchise value of remaining in office. A larger treasury raises the value of attacking, while higher platform quality and stronger incumbent reputation raise the value of waiting. A stronger challenger reduces expected tenure and makes immediate extraction more attractive. The model therefore identifies a state-dependent attack frontier. Attacks are most likely when treasury assets are large, monitoring is weak, and the incumbent's continuation value has deteriorated. Pseudonymous re-entry makes misconduct less punishable, while declining pivotality makes it less likely that tokenholders discipline delegates before the attack region is reached.

These mechanisms generate a negative feedback loop between platform growth and governance quality. Higher reputation supports platform growth by increasing expected tenure, delegate effort, and the franchise value of office. Growth then expands the tokenholder base and the treasury, which lowers individual monitoring incentives and increases the prize from misconduct. Successful governance can therefore weaken the discipline that sustains it. In some parameter regions, this feedback generates

boom-bust dynamics in which reputation and quality rise, monitoring declines, attack incentives increase, a governance failure occurs, and the platform resets to a lower-quality state with lower reputation.

The model also links governance discipline to asset prices. Because the governance token capitalizes the value of future platform outcomes, weaker monitoring reduces expected governance discipline and puts downward pressure on token prices. Governance failures reduce value by lowering platform quality, draining the treasury, and resetting reputation. The model therefore predicts that platform failures are most likely after periods of apparent success, when platform growth has reduced monitoring and increased the private gains from misconduct.

We use the model to quantify the welfare cost of pseudonymous token-based governance. The relevant comparison is not a frictionless first best, but an authenticated benchmark in which delegate identities are permanent and malicious delegates can be excluded after misconduct. The welfare loss from pseudonymous governance has two components. First, attacks reduce future surplus by lowering quality, draining the treasury, and resetting reputation. Second, monitoring is costly and becomes inefficiently weak as the platform grows. These mechanisms jointly determine the invariant distribution of platform quality, treasury size, reputation, monitoring, and attacks, which we use to compute the deadweight loss of pseudonymous tokenized governance relative to authenticated governance.

We extend the baseline model in two directions. The first extension introduces token-denominated delegate compensation. Token pay ties continuation value to the long-run value of the platform and can discourage attacks when delegates are sufficiently forward-looking. Token compensation is not a panacea, however, because it makes token supply and delegate token holdings endogenous and creates a feedback between incentive design, dilution, and token prices. The second extension allows delegates to choose treasury-financed investment. Investment raises future quality but reduces current liquid resources. When the treasury is small, a malicious delegate may over-invest to build a larger future prize. When the treasury is large, she may under-invest and hoard liquidity because immediate extraction becomes more attractive. The same treasury can therefore be both productive capital and an object of opportunistic extraction.

The paper contributes to three literature streams. First, it contributes to the literature on delegated

monitoring. Diamond (1984) shows that delegating monitoring to an intermediary can economize on duplicated monitoring costs, but also creates a new agency problem because the intermediary must itself be incentivized. We study a setting in which delegation is not only a response to costly monitoring, but also a source of endogenous governance fragility.

Second, the paper contributes to the literature on reputation and dynamic moral hazard. Classic models show that agents may exert high effort or forgo short-run gains early on to build reputation and later benefit from established trust (see e.g., Klein and Leffler, 1981; Kreps and Wilson, 1982; Milgrom and Roberts, 1982; Shapiro, 1983). In our setting, reputation-building can also undermine governance discipline because malicious delegates may use reputation to expand the platform and increase the future prize from misconduct.

Third, the paper contributes to the literature on proxy voting and shareholder democracy in traditional settings, which partly mirror the blockchain environment through hidden incentives and opaque delegated control by proxy advisors (Malenko and Shen, 2016; Malenko and Malenko, 2019; Levit et al., 2024). Existing work often studies how governance outcomes depend on the preferences of voters and proxy advisors. We show that governance outcomes also depend on agents' incentives to an evolving governance environment.

The remainder of the paper is organized as follows. Section 2 provides institutional details. Section 3 introduces the model. Section 4 analyzes delegate and tokenholder incentives. Section 5 derives the governance implications. Section 6 extends the model with compensation and investment incentives. Section 7 concludes.

2 Institutional Details and Related Literature

2.1 Utility vs Governance Tokens

Utility tokens are platform-issued digital claims recorded on a blockchain that provide access to platform services, voting rights, or some combination of both. They do not provide cash-flow rights, so early

investors often hold them for speculative purposes (Fahlenbrach and Frattaroli, 2021; Gan et al., 2021) and later sell them to users who value them for future access to, and demand for, the platform's services (Prat et al., 2025). The literature shows that utility tokens can help finance platform growth (Howell et al., 2020), support adoption through network effects (Cong et al., 2022), commit platform owners to more competitive pricing (Goldstein et al., 2024), and be optimal relative to equity financing when agency conflicts are low (Gryglewicz et al., 2021).

As platforms mature, many issue or assign voting rights to token holders. These *governance* tokens give holders voting rights over platform policies (e.g., operating rules, risk management, incentive programs, and spending decisions). Governance tokens also avoid explicit dividend or cash-flow rights (partly to reduce securities law concerns). Their value therefore depends more on expected platform adoption, control, and secondary market trading.

2.2 Token-based Governance vs Vote Delegation

Token-based governance has several advantages over traditional corporate governance. First, token-based governance allows ownership and control rights to be created, exchanged, and verified cheaply. Because the relevant records are stored on a blockchain, ownership and voting information can be updated in real time, whereas traditional systems often fall short (Kahan and Rock, 2008; Yermack, 2017). Second, these records can make decision-making with tokens more inclusive and transparent, with far less secrecy and fewer back-door negotiations than in traditional governance.

An important limitation is that user pseudonymity and the ability to reset identity on the blockchain can still allow actors to obscure true ownership and control over the platform. Nonetheless, the transparency of the governance process itself allows stakeholders to trace decision-making and, in the worst case, exit (Bakos and Halaburda, 2022), or push back against poor and abusive platform governance (Sockin and Xiong, 2023; Abadi and Brunnermeier, 2024).

Token-based governance usually follows a one-token-one-vote rule. Larger token holders have more voting power and thus more skin in the game. Empirically, participation in token-based governance is

costly and concentrated (Appel and Grennan, 2023; Han et al., 2025). Voting requires time, research, and effort. Because votes are public, small voters may also be discouraged if large voters overturn outcomes at the last minute before the voting window closes (Rossello, 2024). To reduce participation costs, many major DeFi platforms have adopted vote delegation, a programmable version of *liquid democracy* (Blum and Zuber, 2016; Kahng et al., 2021; Berinsky et al., 2025). Token holders can either self-delegate, retaining voting power and voting directly, or delegate voting power to representatives who vote on their behalf. This delegation system creates, in effect, a liquid market for control rights (Jensen and Ruback, 1983; Grossman and Hart, 1988; Burkart et al., 2024).

2.3 Delegates and Incentives

Delegates fall into three broad groups. The first group consists of professional delegates, including governance advisory firms (e.g., Gaunlet), investment and trading firms (e.g., Wintermute), and technical development teams (see Internet Appendix for a description). Advisory firms compete for reputation and voting power, which allows them to sell consulting services. Investment and trading firms may value delegation because platform policies affect token prices and platform risk, and can create trading opportunities (Cong et al., 2025). Development teams participate because they have technical expertise and are often paid to maintain, upgrade, or integrate the platform. All of these actors often interact repeatedly with each other to discuss platform policies.

The second group consists of individuals and nonprofit organizations, including student groups from US Ivy League universities.[3] These delegates are often the largest group by address count but usually hold less aggregate voting power than professional delegates. Their participation motives are more intrinsic than financial. The third group consists of application-based delegates that seek to increase participation by rewarding voters or routing governance activity through a separate platform (see Laturnus and Lehar (2026) for a review).

Explicit delegate compensation remains limited but appears to be growing.[4] For example, Uniswap

[3] See, e.g., Blockchain at Columbia, Blockchain at Michigan, Blockchain at Berkeley, Blockchain at UCLA, and Penn Blockchain, FranklinDAO.

[4] See gov.uniswap.org, gov.optimism.io, polygon.technology for their delegate reward programs.

introduced a program that pays eligible delegates up to $6,000 per month in *UNI* tokens, conditional on participation and engagement. Optimism has paid active delegates 6,000 tokens, roughly $5,000 at the time of distribution, based on voting activity. Polygon's 2025 delegate incentive program paid fourteen active delegates quarterly compensation between $624 and $10,519 in the platform's native token.

2.4 Policy Implementation Process, Treasuries, and Misconduct

The governance process typically begins off-chain. Users and delegates discuss potential platform policies online in public forums such as Discord or Telegram. Many platforms then hold a nonbinding vote to test community support. If support is sufficient, the proposal moves to an on-chain vote, where token-weighted voting determines whether it becomes binding. An on-chain proposal therefore contains both a written description of the policy to be implemented and executable code. Voting windows are often a few days and require specified quorums to pass.

Many governance proposals change how the platform operates or decide how platform funds are spent. These funds are held in the platform's *treasury*, a pool of assets controlled by governance. The treasury may include money raised through token sales, platform revenues, and the platform's own governance token. It differs from *total value locked* (TVL), which refers to user-supplied assets (similar to customer assets under management) that cannot be spent through the governance process.

Because the governance process is transparent but pseudonymous, malicious actors can try to pass proposals that redirect treasury funds to their accounts (see Feichtinger et al. (2024); Laturnus and Lehar (2026) and defillama.com for examples of governance attacks). These attacks usually occur when communities pay limited attention to governance and often appear in two steps. Attackers first propose changes to voting thresholds, administrative privileges, or enforcement rules. They can then use these governance implementation changes to pass a second proposal that extracts the treasury's assets. Other forms of misconduct or entrenchment may include classic empire-building and the adoption of platform policies that serve individual (or elite) interests at the expense of the platform's growth (see Internet Appendix for empirical evidence).

3 Model

We consider a token-based platform governed by a population of tokenholders. Time is continuous and we denote it by $t \in \mathbb{R}_t$. At each point in time, the size of the tokenholder community, $N_t \in \mathbb{N}$, is Poisson distributed with average $\bar{n}(S_t) \in \mathbb{R}_+$, where S_t is the platform sate. Each Tokenholder holds an amount $x \in \mathbb{R}_+$ of tradable governance tokens and delegate voting power to specialized *delegates* through a representative plutocracy.[5] Delegation arises because specialized delegates are better equipped to evaluate frequent and technically complex governance decisions. However, it reintroduces a separation between ownership and control, giving rise to a principal-agent problem (Jensen and Meckling, 1976; Tirole, 2006). In this environment, The principal–agent problem in this environment is shaped by two frictions. First, delegate identity is anonymous—pseudonymous to be precise—and thus not perfectly verifiable. Second, tokenholder oversight is costly, so individual voters may reduce monitoring effort when their perceived influence on the governance outcome is small.

At each date, two delegates are active. The incumbent, denoted I, controls the platform enacting decisions; the challenger, denoted C, is the next potential replacement. Tokenholders run elections to maintain the current incumbent or replace it with the challenger at governance proposals, which occur according to a Poisson process with rate $\lambda \in \mathbb{R}_{++}$. Delegates and users alike can also be hit by idiosyncratic liquidity shocks at rate $\delta \in \mathbb{R}_{++}$ and quit the platform. When a users exits she is immediately replaced by a new user. If the incumbent delegate exits, the challenger becomes incumbent and a new challenger enters. If the challenger exits, the incumbent remains in office and the challenger is replaced.

The platform state is $S_t = (Q_t, T_t, p_t, \tilde{p}_t)$. Here $Q_t > 0$ is platform quality, $T_t \geq 0$ is the value of assets in the platform treasury, and $(p_t, \tilde{p}_t) \in (0,1)^2$ are the reputations of the incumbent and challenger, respectively. Reputation is the tokenholders' belief that a delegate is aligned with them. These beliefs affect voting outcomes and delegates' dynamic incentives, as specified below.

[5] Plutocracy means rule by wealth; In token governance, this corresponds to "one token, one vote." Democracy on the other hand means rule by the people, often summarized as "one person, one vote." The terms derive from the Greek *demos* (people), *kratos* (rule), and *ploutos* (wealth).

Types, Reputation, and Attacks. Each delegate has type $k \in \{\mathrm{H}, \mathrm{M}\}$, which can be either Honest or Malicious. A delegate is honest, H, with prior probability $p_0 \in (0,1)$, and malicious, M, with initial prior probability $1 - p_0$. Honest delegates never divert treasury resources. Malicious delegates may initially behave like honest delegates, exert effort, and build reputation, but retain the option to attack the platform once doing so becomes sufficiently attractive.

An attack is a governance action that diverts the treasury to the incumbent and reduces platform quality. Because only the incumbent controls the platform, attacks can occur only when a malicious delegate obtains or retains office. We assume that attacks can be performed immediately after a governance proposal. Following an attack, the delegate loses current control and accumulated reputation, but anonymity allows re-entry under a new identity.

Tokenholders do not observe delegate types, but they update beliefs over them from governance histories using Bayes' rule. Since honest delegates never attack, each governance proposal at which the incumbent remains in control without attacking increases tokenholders' posterior belief that she is honest. Let $p \in (0,1)$ denote tokenholders' current prior belief that the incumbent is honest before a proposal, and let p^+ denote the posterior belief after no attack is observed. If a malicious incumbent attacks with probability $q(S)$, Bayes' rule implies the odds-ratio update

$$\frac{p^+}{1-p^+} = \frac{p}{1-p}\frac{1}{1-q(S)}, \qquad p_0 \in (0,1) \text{ given.} \tag{1}$$

Thus, no-attack histories raise reputation whenever $q(S) > 0$. If an attack is observed, the delegate loses her accumulated reputation and re-enters under a new identity, resetting her reputation to p_0.

Platform Fundamentals. While in office, the incumbent chooses effort $E_t \in \mathbb{R}_+$. Effort improves platform quality but is costly. Platform quality evolves as

$$dQ_t = Q_t\big[(\mu(E_t) - \eta)dt + \sigma_Q dW_t\big] - \varphi Q_{t^-} dJ_t, \qquad \varphi \in [0,1), \tag{2}$$

where $\mu'(E) > 0$, $\mu''(E) \leq 0$, $\eta > 0$ is a technological depreciation rate, and J_t is a Poisson counter which increases upon attacks, reducing quality by fraction φ.

The treasury T_t is the stock of liquid resources controlled by the platform, and evolves according to:

$$dT_t = \Big[(f(Q_t) + g(Q_t))\bar{n}(S_t) - w(S_t)\Big]dt - T_{t^-}dJ_t. \tag{3}$$

Platform activity generates two inflows. Users pay a per-user fee $f(Q_t) \geq 0$, so aggregate fee revenue is $f(Q_t)\bar{n}(S_t)$. In addition, higher platform quality and participation attract external capital inflows, denoted $g(Q_t)\bar{n}(S_t)$. The per-user rate of external inflow satisfies $g(Q_t) > 0$ and $g'(Q)_t > 0$. The treasury is depleted by delegate compensation $w(S_t)$ and by governance attacks, captured by the jump term.[6]

We focus on the case in which fee revenue is fully passed through to the incumbent delegate as compensation, so that $w(S_t) = f(Q_t)\bar{n}(S_t)$. Fees therefore affect user participation and delegate incentives, but they do not accumulate in the treasury. Taking this into account, the treasury evolves according to:

$$dT_t = g(Q_t)\bar{n}(S_t)dt - T_{t^-}dJ_t. \tag{4}$$

We close the adoption side of the model by assuming that the previously defined participation intensity $\bar{n}(S)$ is increasing in the utilization surplus generated by the platform:

$$\bar{n}(S) = \alpha\bar{U}(S)^+, \qquad \alpha > 0. \tag{5}$$

Here $\bar{U}(S)$ denotes the non-monetary component of tokenholder surplus, net of the value of token holdings. We define it formally below when introducing user payoffs.

Governance Voting. At governance proposal times, tokenholders vote between the re-electing the current incumbent or replacing her with the challenger. We model the vote using the Poisson voting framework of Myerson (1998), so that vote counts are independent Poisson random variables.

We also assume that reputation is not perfectly observable by tokenholders and monitoring is costly:

[6]The main insights also hold under partial diversion.

Let ℓ^* denote the delegate with higher reputation,

$$\ell^* \in \arg\max_{\ell \in \{I,C\}} p_\ell, \qquad p_I = p, \quad p_C = \tilde{p},$$

and let $\ell' \neq \ell^*$ denote the other delegate.[7] A fraction $\rho \in [0,1]$ of voters pays an attention cost $\kappa > 0$ to become informed and vote for ℓ^*. Uninformed voters, on the other hand, rely on a noisy signal of delegate reputation and select ℓ^* with probability $\theta \in (1/2, 1)$. If $\bar{n}(S)$ is the expected voting population, Poisson thinning implies that vote counts for the higher- and lower-reputation delegates are independent Poisson random variables with means

$$\mu_+ = \bar{n}(S)\left[\rho + (1-\rho)\theta\right], \qquad \mu_- = \bar{n}(S)(1-\rho)(1-\theta). \tag{6}$$

The higher-reputation delegate is elected whenever $N_{\ell^*} - N_{\ell'} > 0$, with probability

$$\pi_{\ell^*}(S,\rho) = \Pr(N_{\ell^*} - N_{\ell'} > 0). \tag{7}$$

Hence the probability that the incumbent retains control is

$$\pi_I(S,\rho) = \begin{cases} \pi_{\ell^*}(S,\rho), & \text{if } I = \ell^*, \\ 1 - \pi_{\ell^*}(S,\rho), & \text{if } C = \ell^*. \end{cases}$$

Tokenholders' incentive to acquire information will depend on the probability of being pivotal and on the continuation-value gain from electing the superior delegate. Holding aggregate monitoring ρ fixed, an individual tokenholder benefits from paying the attention cost κ only when her vote changes the election outcome. Since $N_{\ell^*} - N_{\ell'}$ is the difference between two independent Poisson random variables, it follows a Skellam distribution. The pivotality probability is therefore

$$\mathbb{P}_{\text{piv}}(S,\rho) = \Pr(N_{\ell^*} - N_{\ell'} = 0) = e^{-(\mu_+ + \mu_-)} I_0(2\sqrt{\mu_+ \mu_-}), \tag{8}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind.[8]

[7] The payoff-relevant ranking is the continuation-value ranking. In the value-accretive environment considered below, higher reputation lowers perceived attack risk and increases expected governance value, so the continuation-value ranking coincides with the reputation ranking.

[8] For $z \in \mathbb{R}$, the order-zero modified Bessel function of the first kind is $I_0(z) = \sum_{m=0}^{\infty} \frac{(z^2/4)^m}{(m!)^2}$.

Let $\Delta U(S, x)$, defined below, denote the surplus gain for a tokenholder with holdings x from electing ℓ^* rather than ℓ'. Because information raises the probability of voting for ℓ^* from θ to one, its accuracy gain is $1 - \theta$. The marginal benefit of information is

$$\mathrm{MBI}(S, \rho) = \mathbb{P}_{\mathrm{piv}}(S, \rho)(1 - \theta)\Delta U(S, x). \tag{9}$$

Given aggregate monitoring ρ, an individual tokenholder acquires information whenever $\mathrm{MBI}(S, \rho) \geq \kappa$. In a symmetric monitoring rule, the informed share is characterized by

$$\rho(S) \in \begin{cases} \{1\} & \text{if } \mathrm{MBI}(S, \rho) > \kappa, \\ [0, 1] & \text{if } \mathrm{MBI}(S, \rho) = \kappa, \\ \{0\} & \text{if } \mathrm{MBI}(S, \rho) < \kappa. \end{cases} \tag{10}$$

State Transitions. Let $S_{t^-} = (Q, T, p, \tilde{p})$ denote the platform state right before a governance proposal. The possible post-proposal states are summarized in Table 1. If the incumbent is retained, her reputation

Table 1: Governance transitions at proposal times

Event	Post-proposal state
Incumbent retained, no attack	$S^+ = (Q, T, p^+, \tilde{p})$
Incumbent retained, attack	$S^- = \big((1-\varphi)Q, 0, p_0, \tilde{p}\big)$
Challenger elected, no attack	$S_R^+ = (Q, T, \tilde{p}^+, p)$
Challenger elected, attack	$S_R^- = \big((1-\varphi)Q, 0, p_0, p\big)$
Incumbent exits	$S_R = (Q, T, \tilde{p}, p)$

updates to p^+ after no attack; if she attacks, quality falls to $(1 - \varphi)Q$, the treasury is drained, and her reputation is reset to p_0. If the challenger is elected, roles rotate, so the challenger's reputation becomes the incumbent reputation and the former incumbent's reputation becomes the challenger reputation. The same attack consequences apply if the newly elected challenger attacks. Finally, if the incumbent exits exogenously, control passes to the challenger without an attack or reputation update.

Token Supply and Market Clearing. Let X denote the total supply of governance tokens. In the baseline model, total supply is fixed and normalized to one, $X = 1$. Since the expected mass of active

tokenholders is $\bar{n}(S)$, symmetric token holdings satisfy the market-clearing condition $X = \bar{n}(S)x(S)$, or

$$x(S) = \frac{1}{\bar{n}(S)}. \tag{11}$$

We keep X in the notation because token supply becomes a state variable in the token-compensation extension in Section 6.

Tokenholders' Payoff. Platform fundamentals and governance determine tokenholders' surplus from participation. A participating user obtains flow surplus $u(Q_t) - f(Q_t)$, where $u'(Q) > 0$ is the utility from platform services and $f(Q_t)$ is the participation fee. Users also hold governance tokens. Let x_t denote an individual user's token holdings and let $P(S_t)$ denote the token price. Users exit the platform after an idiosyncratic liquidity shock, at which point they liquidate their token holdings at market value.

The monitoring decision introduced above enters the user's payoff through the proposal-time information choice. At each proposal τ_k, the user chooses $m_k \in \{0, 1\}$, where $m_k = 1$ means that she pays the attention cost κ and becomes informed. Given an initial state $S_t = S$ and initial token holdings $x_t = x$, the user's value is

$$U(S, x) = \max_{\{x_s, m_k\}} \mathbb{E}_t \left[\int_t^\tau e^{-r(s-t)} \big(u(Q_s) - f(Q_s)\big)\, ds - \sum_{k:\, t<\tau_k\leq\tau} e^{-r(\tau_k - t)} \kappa\, m_k + e^{-r(\tau-t)} x_\tau P(S_\tau) \right], \tag{12}$$

where τ is the user's liquidity-shock exit time. The maximization is over token holdings and proposal-time information acquisition, taking the aggregate monitoring rule $\rho(S)$ and the induced voting probabilities as given.

Because preferences are quasi-linear in token holdings, user value can be decomposed as $U(S, x) = \bar{U}(S) + P(S)x$. Equivalently,

$$\bar{U}(S) = U(S, x) - P(S)x. \tag{13}$$

The term $\bar{U}(S)$ is the utilization surplus entering the adoption equation (5); it is the component of tokenholder surplus net of the market value of token holdings.

Delegates' Payoff. Delegates value control because office provides wage rents and, for malicious delegates, access to the treasury. While in office, a delegate receives wage flow $\omega(Q_t)$ and chooses effort E_t, which entails cost $c(E_t)$, with $c(0) = 0$, $c'(E) > 0$, and $c''(E) \geq 0$ for all $E > 0$. Honest delegates choose effort only. Malicious delegates also choose whether to attack after obtaining or retaining control at a governance proposal. The one-shot attack payoffs are $\Pi^M(S) = T$ and $\Pi^H(S) = 0$. For a type $k \in \{\mathrm{H}, \mathrm{M}\}$ incumbent, the value of participation in governance is

$$V_I^k(S) = \max_{\{E_s, a_j\}} \mathbb{E}_t \left[\int_t^\infty e^{-r(s-t)} \iota_s \big(\omega(Q_s) - c(E_s)\big)\, ds + \sum_{j:\, \tau_j \geq t} e^{-r(\tau_j - t)} w_j a_j \Pi^k(S_{\tau_j^-}) \right], \tag{14}$$

where ι_s indicates that the delegate is in office, w_j indicates that she controls the platform after the vote at τ_j, and $a_j \in \{0, 1\}$ indicates whether she attacks. For honest delegates the attack term is inactive.

The payoff function in (14) pins down the attack probabilities that enter belief updating in equation (1). There, $q(S)$ denotes the attack probability of the delegate whose reputation is being updated. At a governance proposal, however, this probability depends on whether control is retained by the incumbent or transferred to the challenger. We therefore write $q_I(S)$ for the attack probability of a malicious incumbent who retains control, and $q_C(S)$ for the attack probability of a malicious challenger who have just obtained control. Using the transition states in Table 1, the attack margins are

$$\begin{aligned} \Delta_I(S) &\equiv T + V_C^M(S^-) - V_I^M(S^+), \\ \Delta_C(S) &\equiv T + V_C^M(S_R^-) - V_I^M(S_R^+). \end{aligned} \tag{15}$$

Thus the attack rule is characterized by

$$q_j(S) \in \begin{cases} \{1\} & \text{if } \Delta_j(S) > 0, \\ [0, 1] & \text{if } \Delta_j(S) = 0, \\ \{0\} & \text{if } \Delta_j(S) < 0, \end{cases} \qquad j \in \{I, C\}. \tag{16}$$

An attack occurs when the current extraction payoff plus the post-attack continuation value exceeds the value of preserving control.

Welfare. Aggregate welfare is the expected present value of platform surplus net of cost and disutilities. Transfers between users, delegates, and the treasury are internal to the platform and enter welfare only through their effect on incentives and future states. The real costs are delegate effort and tokenholder information acquisition. Given an effort path, monitoring policy, and induced state process, welfare is

$$\mathcal{W}(S_t) = \mathbb{E}_t\left[\int_t^\infty e^{-r(s-t)}\left(\bar{n}(S_s)u(Q_s) - c(E_s)\right)ds - \sum_{k:\tau_k \geq t} e^{-r(\tau_k - t)}\bar{n}(S_{\tau_k^-})\rho(S_{\tau_k^-})\kappa\right]. \tag{17}$$

Attacks affect welfare through their effect on the continuation state. For example, an incumbent attack moves the platform from $S^+ = (Q, T, p^+, \tilde{p})$ to $S^- = \big((1-\varphi)Q, 0, p_0, \tilde{p}\big)$, lowering future surplus, adoption, effort incentives, and governance discipline. Thus the welfare cost of an attack is captured by the induced continuation-welfare loss, $\mathcal{W}(S^+) - \mathcal{W}(S^-)$.

3.1 Recursive Formulation

The previous section defines the state variables, governance technology, and payoff primitives. To define and characterize Markov perfect equilibria (MPEs) of the model, it is necessary to express these agents' problems recursively. This subsection develops that recursive representation.

The first step to do so is to define the generator governing the continuous evolution of platform fundamentals between discrete governance events. During these periods, only Q and T evolve continuously, while beliefs change at proposal and exit events. For any function $F \in \mathcal{C}^2$, the diffusion generator under incumbent effort E is[9]

$$(\mathcal{L}^E F)(S) = b_Q(S; E)\partial_Q F + b_T(S)\partial_T F + \frac{1}{2}\sigma_Q^2 Q^2 \partial_{QQ} F, \tag{18}$$

where $b_Q(S; E) = Q(\mu(E) - \eta)$, and $b_T(S) = f(Q)\bar{n}(S) - \omega(Q)$ denote drift terms.

The jump terms in the Bellman equations below are induced by the transition states in Table 1 and by the voting, monitoring, and attack rules defined above.

[9]The symbol $\mathcal{C}^2$ stands for twice-continuously differentiable.

Tokenholders. Let $U(S,x)$ denote the continuation value of a tokenholder with holdings x. Using the decomposition in (13), $U(S,x) = \bar{U}(S) + P(S)x$. Given incumbent effort and the monitoring rule in (10), tokenholder value satisfies

$$rU(S,x) = \max_{x} \left\{ u(Q) - f(Q) + \mathcal{L}^E U(S,x) + \delta\big[xP(S) - U(S,x)\big] + \lambda\big[\mathcal{E}_{\text{vote}}(S,x,\rho) - U(S,x)\big] \right\}. \tag{19}$$

The liquidity-shock term reflects that an exiting user liquidates her token holdings at market value. The proposal term uses the continuation value

$$\mathcal{E}_{\text{vote}}(S,x,\rho) = \pi_I(S,\rho)\mathcal{U}_I(S,x) + \big(1 - \pi_I(S,\rho)\big)\mathcal{U}_C(S,x) - \rho(S)\kappa. \tag{20}$$

Here $\pi_I(S,\rho)$ is the incumbent-retention probability defined in the voting stage. The continuation value from delegate $\ell \in \{I, C\}$ controlling the platform is

$$\mathcal{U}_\ell(S,x) = (1 - \hat{q}_\ell(S))U(S_\ell^+, x) + \hat{q}_\ell(S)U(S_\ell^-, x), \tag{21}$$

where

$$\hat{q}_I(S) = (1-p)q_I(S), \quad \hat{q}_C(S) = (1-\tilde{p})q_C(S) \tag{22}$$

are the attack probabilities perceived by tokenholders. The states S_ℓ^+ and S_ℓ^- are those in Table 1: $(S_I^+, S_I^-) = (S^+, S^-)$ and $(S_C^+, S_C^-) = (S_R^+, S_R^-)$. The continuation-value gain from electing the higher-reputation delegate is

$$\Delta U(S,x) = \mathcal{U}_{\ell^*}(S,x) - \mathcal{U}_{\ell'}(S,x). \tag{23}$$

Together with pivotality, this is the object entering the marginal benefit of information in (9). Competitive token pricing implies that the token price equals the marginal value of token holdings:

$$P(S) = \partial_x U(S,x). \tag{24}$$

Delegates. Let $V_I^k(S)$ and $V_C^k(S)$ denote the continuation values of a type-$k \in \{\text{H}, \text{M}\}$ delegate as incumbent and challenger. The incumbent controls platform quality and receives wage flow $\omega(Q)$. The

challenger receives no wage and has no control over effort, but may obtain control at a future governance event or after incumbent exit. It is useful to collect proposal and incumbent-exit events into a single event rate

$$\zeta \equiv \lambda + \delta, \qquad \chi \equiv \frac{\delta}{\lambda + \delta}. \tag{25}$$

At such an event, incumbent exit occurs with probability χ, while a governance proposal occurs with probability $1 - \chi$.

For $j \in \{I, C\}$, define $(S_I^+, S_I^-) = (S^+, S^-)$, $(S_C^+, S_C^-) = (S_R^+, S_R^-)$. Conditional on controlling the platform after the proposal, a type-k delegate's continuation value is

$$\mathcal{A}_j^k(S) = \begin{cases} V_I^H(S_j^+), & k = \mathrm{H}, \\ (1 - q_j(S))V_I^M(S_j^+) + q_j(S)\big[T + V_C^M(S_j^-)\big], & k = \mathrm{M}. \end{cases} \tag{26}$$

This expression uses the attack probabilities characterized in (16): after obtaining or retaining control, a malicious delegate either preserves the franchise value of office or attacks and re-enters under a new identity.

The value of being out of office after a proposal is

$$\begin{aligned} \mathcal{B}_I^k(S) &= (1 - \hat{q}_C(S))V_C^k(S_R^+) + \hat{q}_C(S)V_C^k(S_R^-), \\ \mathcal{B}_C^k(S) &= (1 - \hat{q}_I(S))V_C^k(S^+) + \hat{q}_I(S)V_I^k(S^-). \end{aligned} \tag{27}$$

The first line applies when the incumbent is replaced by the challenger. The second applies when the challenger remains out of office, except when an incumbent attack transfers control in the damaged state. The event-continuation values are therefore

$$\begin{aligned} \mathcal{E}_I^k(S) &= \chi V_C^k(S_R) + (1 - \chi)\left[\pi_I(S, \rho)\mathcal{A}_I^k(S) + \big(1 - \pi_I(S, \rho)\big)\mathcal{B}_I^k(S)\right], \\ \mathcal{E}_C^k(S) &= \chi V_I^k(S_R) + (1 - \chi)\left[\big(1 - \pi_I(S, \rho)\big)\mathcal{A}_C^k(S) + \pi_I(S, \rho)\mathcal{B}_C^k(S)\right]. \end{aligned} \tag{28}$$

The incumbent and the challenger value solve

$$rV_I^k(S) = \max_{E \geq 0} \left\{ \omega(Q) - c(E) + \mathcal{L}^E V_I^k(S) + \zeta \left[\mathcal{E}_I^k(S) - V_I^k(S) \right] \right\}. \tag{29}$$

$$rV_C^k(S) = \mathcal{L}^E V_C^k(S) + \zeta \left[\mathcal{E}_C^k(S) - V_C^k(S) \right], \tag{30}$$

where E is the effort chosen by the current incumbent.

The attack probabilities entering $\mathcal{A}_j^k$ are pinned down by the attack margins defined in (15) and the attack rule in (16). Finally, if the incumbent effort choice is interior, the first-order condition from (29) is

$$c'(E_k(S)) = Q\mu'(E_k(S))\, \partial_Q V_I^k(S). \tag{31}$$

Effort is therefore increasing in the marginal value of platform quality to the incumbent.

3.2 Equilibrium

We now define a stationary Markov Perfect Equilibrium. The equilibrium is a fixed point of the recursive economy described above. Tokenholders choose token holdings and monitoring optimally, taking delegate strategies, prices, and aggregate monitoring as given. Delegates choose effort and attack policies optimally, taking user monitoring and election probabilities as given. Token prices are consistent with users' marginal value of token holdings, beliefs are updated according to Bayes' rule, and the induced state process admits an invariant distribution.

Definition 1 (Stationary Markov Perfect Equilibrium)**.** A *Stationary Markov Perfect Equilibrium* consists of value functions $\{U, V_I^H, V_I^M, V_C^H, V_C^M\}$, Markov policies $\{x^*, \rho^*, E^*, q_I^*, q_C^*\}$, a token price $P(S)$, adoption $\bar{n}(S)$, beliefs $(p, \tilde{p})$, and an invariant probability measure Φ such that:

1. **Tokenholder optimality.** Given prices, delegate strategies, and aggregate monitoring, tokenholders solve (19). The token price satisfies $P(S) = \partial_x U(S, x)$, as in (24); token holdings are given by $x^*(S)$; monitoring $\rho^*(S)$ satisfies (10); and voting probabilities are induced by (6)–(7).
2. **Delegate optimality.** Given monitoring and election probabilities, delegate values solve (29)–(30). Incumbent effort $E^*(S)$ satisfies (31) when interior, and attack policies $q_I^*(S)$ and $q_C^*(S)$ satisfy

(16), with margins $\Delta_I(S)$ and $\Delta_C(S)$ defined in (15).

3. **Belief consistency.** Perceived attack probabilities satisfy $\hat{q}_I(S) = (1-p)q_I^*(S)$ and $\hat{q}_C(S) = (1-\tilde{p})q_C^*(S)$. Beliefs update according to (1); post-proposal states follow Table 1; and continuation values are those in (20)–(28).

4. **Adoption and market clearing.** Adoption satisfies $\bar{n}(S) = \alpha \bar{U}(S)^+$, as in (5), and token holdings clear the market through $x^*(S) = 1/\bar{n}(S)$, as in (11).

5. **Stationarity.** Given equilibrium policies, the state process induced by $\mathcal{L}^{E^*}$ in (18), the event structure (ζ, χ) in (25), and Table 1 admits invariant probability measure Φ.

Refinement and selection. The threshold rules for monitoring and attacks may generate knife-edge states in which tokenholders are exactly indifferent about acquiring information or malicious delegates are exactly indifferent about attacking. To avoid selecting equilibria supported only by such indifferences, we focus on stable limits of perturbed equilibria. Let $\Sigma(S) = \big(\rho(S), q_I(S), q_C(S)\big)$ denote the vector of strategic probabilities. In an ε-perturbed game, each component of $\Sigma(S)$ is restricted to the interval $[\varepsilon, 1-\varepsilon]$. Let $\Sigma(\cdot;\varepsilon)$ denote an equilibrium strategic profile of the perturbed game, and let Φ_ε denote an invariant probability measure of the state process induced by that profile. A strategy profile Σ^* is stable if there exists a sequence $\varepsilon_n \downarrow 0$ and corresponding perturbed equilibria $\Sigma(\cdot;\varepsilon_n)$ such that

$$\lim_{n\to\infty} \int_{\mathcal{S}} \|\Sigma(S;\varepsilon_n) - \Sigma^*(S)\|_1 \, d\Phi_{\varepsilon_n}(S) = 0. \tag{32}$$

This refinement selects equilibria that can be approached by fully mixed perturbations of the monitoring and attack subgames. It rules out behavior supported only by exact indifference at isolated threshold states.

We also restrict attention to environments in which higher platform quality raises the expected treasury drift.

Assumption 1: Quality-accretive treasury inflows Treasury inflows are locally increasing in platform

quality:

$$\partial_Q b_T(S) = g'(Q)\bar{n}(S) + g(Q)\partial_Q \bar{n}(S) > 0. \tag{33}$$

Assumption 1 states that higher platform quality raises expected treasury accumulation through the external-inflow channel. This condition captures the model's core tension: quality growth makes the platform more valuable for users and attracts more resources under platform control, but it also increases the stock of resources exposed to expropriation.

Definition 2 (Stable-accretive Markov Perfect Equilibrium)**.** A *stable value-accretive Markov Perfect Equilibrium* is a Stationary Markov Perfect Equilibrium in the sense of Definition 1 whose strategic profile $\Sigma^*(S) = \big(\rho^*(S), q_I^*(S), q_C^*(S)\big)$ satisfies the stability condition in (32) and whose compensation and fee policy satisfies Assumption 1.

Equilibrium existence. The following result establishes existence of a stable value-accretive equilibrium and local uniqueness of the strategic threshold policies used in the analysis. The proof is in Appendix A.

Proposition 1 (Existence and local uniqueness of strategic policies)**.** *Assume that:*

(i) $\mu(E)$ *is increasing and concave,* $c(E)$ *is increasing and convex, and* $\lim_{E\to\infty} \mu(E)/c(E) = 0$*;*

(ii) quality dynamics are non-explosive under admissible effort policies;

(iii) user flow surplus is bounded above;

(iv) Assumption 1 holds.

Then there exists at least one stable value-accretive Markov Perfect Equilibrium.

Moreover, for any fixed state S *and continuation-value system, the monitoring policy* $\rho^*(S)$ *is unique except at the indifference threshold* $\mathrm{MBI}(S, \rho) = \kappa$*, and the attack policies* $q_I^*(S)$ *and* $q_C^*(S)$ *are unique except on the attack-threshold surfaces* $\Delta_I(S) = 0$ *and* $\Delta_C(S) = 0$*. The perturbation-stability refinement in* (32) *selects behavior at these indifference states.*

Proposition 1 provides the equilibrium foundation for the analysis below. It establishes existence and pins down the local threshold structure of monitoring and attack policies. It does not require global

uniqueness of the stationary distribution: multiple stationary fixed points may arise because beliefs, attack risk, continuation values, and adoption are jointly determined.

4 Incentive Analysis

We now derive the main incentive implications of the equilibrium characterized above. The analysis proceeds in three steps. First, we study delegates' incentives and show that anonymity reverses the usual single-crossing property from screening models (Mirrlees, 1971; Spence, 1973): malicious delegates may have stronger incentives than honest delegates to build platform value. Second, we analyze tokenholders' monitoring incentives and show that platform growth generates rational apathy through declining pivotality. Third, we characterize the attack margin that determines when accumulated treasury value overwhelms the franchise value of continued office.

4.1 Contracting under Anonymous Governance

A central feature of the model is an *incentive inversion*. The honest delegate values platform quality through its effect on continuation in office and wage flows. The malicious delegate values the same continuation flows, but also internalizes the option value of attacking a larger treasury. As a result, in states where the attack option is relevant but not yet exercised, the malicious type has a strictly higher marginal value of platform quality.

Lemma 1 (Inverted single crossing)**.** *In any state S in which the malicious incumbent preserves the attack option—formally, for $q_I^*(S) \in [0, 1)$—the malicious incumbent has a strictly higher marginal value of platform quality than the honest incumbent:*

$$\partial_Q V_I^M(S) > \partial_Q V_I^H(S).$$

Lemma 1 implies an *impossibility of performance-based screening*. Efficiency wages (Shapiro and Stiglitz, 1984) can raise effort, but under anonymity they do not separate types. Any wage schedule that makes high performance attractive to an honest delegate is weakly more attractive to a malicious delegate,

because the same improvement in platform fundamentals also raises the value of future expropriation. Hence observed effort and performance histories are pooling: a malicious delegate can mimic the honest type's performance path, and in the option-preserving region may even choose a higher-effort path.

The Competence Paradox

This incentive inversion creates a "competence paradox" regarding delegates' actions: the delegate with the most detrimental long-term intentions can have the strongest short-run incentives to grow the platform. The reason is that platform growth raises not only the franchise value of office, but also the value of the malicious delegate's embedded attack option.

Proposition 2 (Adverse incentives and the competence paradox)**.** *Suppose the incumbent effort choice is interior. In every state S in the option-preserving region characterized in Lemma 1, the malicious incumbent exerts strictly higher effort than the honest incumbent:*

$$E_M^*(S) > E_H^*(S).$$

Weakly, outside this region, the effort ranking follows the corresponding marginal-value ranking:

$$E_M^*(S) \geq E_H^*(S) \quad \Longleftrightarrow \quad \partial_Q V_I^M(S) \geq \partial_Q V_I^H(S).$$

The result follows directly from the effort first-order condition. Both types face the same marginal cost of effort. The difference is that, in the option-preserving region, the malicious type has a higher marginal value of platform quality. Thus the malicious delegate may look more competent precisely because she has stronger incentives to build the platform before extracting from it.

Figure A.1 about here

The Authenticated Benchmark

To isolate the role of anonymity, consider an authenticated counterfactual in which delegate identities are permanent and publicly verifiable. Authentication changes the continuation value of misconduct.

If a delegate attacks, she can be excluded from future governance, so the post-attack continuation value associated with anonymous re-entry is eliminated. Formally, the malicious delegate's post-attack continuation value is set to zero, $V_C^M(S^-) = 0$. We also allow for a penalty $\psi \geq 0$, so that the net private payoff from expropriation is $T - \psi$.

Authentication removes the re-entry component of the attack payoff. Under anonymity, a malicious delegate who attacks obtains the current treasury T and retains the continuation value of future re-entry under a new identity, $V_C^M(S^-)$. Conditional on retaining control, she therefore chooses between preserving office and taking the payoff $T + V_C^M(S^-)$. Under authentication, exclusion sets the post-attack continuation value to zero, and any off-chain sanction $\psi \geq 0$ further reduces the private payoff. The corresponding choice becomes

$$\max\left\{V_I^M(S^+),\, T - \psi\right\}. \tag{34}$$

The honest delegate's problem is otherwise unchanged. This change removes the source of the incentive inversion and restore the traditional single-crossing condition.

Proposition 3 (Restored Single-Crossing and Effort Ranking)**.** *In the authenticated regime ($V_C^M(\mathbf{S}^-) = 0$), under any quality-dependent compensation $\omega(Q)$ with $\partial_Q \omega > 0$, the standard single-crossing property holds globally:*

$$\partial_Q V_I^H(\mathbf{S}) > \partial_Q V_I^M(\mathbf{S}) \tag{35}$$

Consequently, the honest delegate exerts strictly higher optimal effort than the malicious delegate ($e_H^*(\mathbf{S}) > e_M^*(\mathbf{S})$), fully reversing the competence paradox.

Because authentication restores the standard monotonic relationship between agent alignment with the platform and marginal effort, tokenholders recover the ability to utilize dynamic contracting to separate types.

Corollary 1 (Separating Equilibrium via Performance Pay)**.** *Given the strict single-crossing property $\partial_Q V_I^H > \partial_Q V_I^M$, there exists a cutoff wage schedule $\bar{\omega}(Q)$ such that tokenholders can construct a menu of contracts $\{\omega^1(Q), \omega^2(Q)\}$ with $\partial_Q \omega^1 > \partial_Q \bar{\omega} > \partial_Q \omega^2$ that separates the types. Specifically, let $V_I^k(\omega)$*

denote the expected continuation value of type k under contract ω. The menu ensures the incentive compatibility (IC) constraints hold such that:

$$V_I^H(\omega^1) \geq V_I^H(\omega^2) \quad \textit{and} \quad V_I^M(\omega^2) \geq V_I^M(\omega^1), \tag{36}$$

Furthermore, if we assume that delegates have an outside option (yielding a participation constraint $V_I^k \geq 0$), tokenholders can use the contract to completely exclude the malicious type.

Corollary 2 (Exclusion via Participation Constraints)**.** *There exists a critical discount rate $\bar{r} > 0$ such that for all $r < \bar{r}$, tokenholders can perfectly exclude malicious actors ex ante. Specifically, there exists a single, high-powered take-it-or-leave-it contract $\omega^*(Q)$ that satisfies the honest delegate's participation constraint ($V_I^H(\omega^*) \geq 0$) while strictly violating the malicious delegate's participation constraint ($V_I^M(\omega^*) < 0$).*

This establishes that the impossibility of screening is not an intrinsic feature of delegated governance, but a direct consequence of *anonymity*, which removes the ability to punish via exclusion.

4.2 Monitoring incentives and Rational Apathy

We next analyze the voting stage. The central friction is that governance quality depends on dispersed tokenholders acquiring costly information, yet the private return to becoming informed is inherently a *pivotality* return. A user pays the attention cost only if her information is likely to change the electoral outcome and if the resulting change in governance quality is sufficiently valuable. This creates a fundamental tension in token-based platforms: the same growth that raises the economic stakes of governance also enlarges the voting population, thereby diluting the influence of any one tokenholder. We first show that this force generates rational apathy in large communities, and then show how the resulting deterioration in governance discipline is capitalized into token prices.

Marginal benefit of information. The user's monitoring condition is defined in (9)–(10). Equation (9) shows that the marginal benefit of information is the product of pivotality, the accuracy gain $1 - \theta$, and the continuation-value gain $\Delta U(S, x)$ from selecting the superior delegate. This expression isolates two

forces. First, $\Delta U(S, x)$ is a *governance-stakes term*: it is larger when fundamentals, treasury exposure, or differences in perceived delegate quality make the electoral choice more consequential. Second, $\mathbb{P}_{\text{piv}}(S, \rho)$ is a *congestion term*: it captures how much influence any one user retains once voting power is aggregated across the population. Governance apathy emerges when the second force collapses faster than the first rises.

Skellam pivotality and the scale effect. The Poisson voting technology in Section 3 implies that the vote margin $\Delta N = N_{\ell^*} - N_{\ell'}$ is Skellam. With the voting intensities μ_+ and μ_- defined there, pivotality is the tie probability in (8). Write

$$a(\rho, \theta) = \rho + (1 - \rho)\theta, \qquad b(\rho, \theta) = (1 - \rho)(1 - \theta),$$

so that $\mu_+ = \bar{n}(S)a(\rho, \theta)$, $\mu_- = \bar{n}(S)b(\rho, \theta)$, and $\mu_+ + \mu_- = \bar{n}(S)$. Standard large-argument asymptotics for I_0 imply[10]

$$\mathbb{P}_{\text{piv}}(S, \rho) = O\big(\bar{n}(S)^{-1/2} e^{-\xi(\rho,\theta)\bar{n}(S)}\big), \qquad \xi(\rho, \theta) \equiv \Big(\sqrt{a(\rho, \theta)} - \sqrt{b(\rho, \theta)}\Big)^2. \tag{37}$$

If heuristic voters are even slightly informative, $\theta > 1/2$, then $a(\rho, \theta) > b(\rho, \theta)$ for every $\rho \in [0, 1]$. Hence $\xi(\rho, \theta) > 0$, and pivotality decays exponentially fast in the size of the voting population, up to the usual $\bar{n}^{-1/2}$ prefactor. The familiar $\Theta(\bar{n}^{-1/2})$ decay arises only in the knife-edge balanced case in which the election remains asymptotically tied.

Rational apathy and the paradox of successful governance. The asymptotic form in (37) sharpens the logic of rational apathy. As the platform scales, two things happen simultaneously. Higher platform quality attracts more users and increases treasury inflows, making governance more valuable in the aggregate. At the same time, larger scale makes the electoral outcome less contestable by any individual voter. The latter force is mechanical and arises even if users remain fully rational and forward-looking. In sufficiently large systems, the private return to information acquisition becomes negligible relative to

[10] The more precise asymptotic approximation is $\mathbb{P}_{\text{piv}}(S, \rho) \sim \frac{1}{\sqrt{4\pi\bar{n}(S)\sqrt{a(\rho,\theta)b(\rho,\theta)}}} \exp\left[-\bar{n}(S)\Big(\sqrt{a(\rho, \theta)} - \sqrt{b(\rho, \theta)}\Big)^2\right]$.

its cost, despite governance being socially important.

This creates a paradox of growth. The success of the platform generates a larger governance problem—more users, a larger treasury, and more at stake—while simultaneously weakening the private incentives needed to discipline delegates. Governance becomes more important in levels and less valuable at the margin for any individual participant.

Heuristics and information substitution. The parameter θ governs the quality of heuristic voting. When θ is high, the crowd already tends to select the superior delegate even without private investigation. The value of becoming informed is then small for two reasons: the informational improvement $1-\theta$ is smaller, and the election is less likely to be close, which lowers pivotality. This makes heuristic accuracy and private monitoring strategic substitutes.

By contrast, when θ is close to $1/2$, heuristic voting is nearly uninformative. In that case, the value of information can be substantial, but only if the platform has not yet reached a scale at which pivotality is already negligible. Private monitoring is therefore most sustainable in intermediate environments: elections must be consequential enough that $\Delta U(S,x)$ is large, but not so large in scale that individual influence has already disappeared.

The following proposition collects these implications: pivotality vanishes with scale, the marginal benefit of information inherits this decay, and monitoring collapses once the voting population is sufficiently large.

Proposition 4 (Pivotality decay and rational apathy)**.** *Fix a state S and suppose $\theta > 1/2$. Then:*

(i) Pivotality decays with platform scale at the rate in (37), with $\xi(\rho,\theta) > 0$.

(ii) Since $\Delta U(S,x)$ is bounded under the bounded-value conditions used for equilibrium existence,[11] the marginal benefit of information satisfies the same order bound.

(iii) For each fixed state S, $\mathrm{MBI}(S,\rho)$ is strictly decreasing in ρ. Hence the voting subgame admits a unique equilibrium monitoring choice $\rho^*(S) \in [0,1]$, except at the knife-edge indifference cases

[11]The user value U is bounded under the maintained bounded-flow and discounted HJB assumptions. Since each $\mathcal{U}_\ell(S,x)$ is a convex combination of values of U, $|\mathcal{U}_\ell(S,x)| \le \|U\|_\infty$. Hence $|\Delta U(S,x)| \le 2\|U\|_\infty$.

selected by the perturbation refinement.

(iv) There exists $\bar{N} < \infty$ such that whenever $\bar{n}(S) > \bar{N}$, the unique best response is $\rho^*(S) = 0$.

Proposition 4 implies that governance participation is fragile not only because monitoring is costly, but because it is a public good built on an increasingly vanishing private margin. In large platforms, governance therefore becomes endogenously reliant on heuristics, reputation, and a small residual set of highly motivated participants. This is the voting-side analogue of the broader governance cycle developed in the paper: growth increases both adoption and the treasury, but it also reduces individual incentives to monitor, making the platform more exposed to opportunistic delegates.

Uniqueness within states and multiple thresholds across states. A remark on uniqueness of monitoring thresholds is worth making. For a fixed state S, holding continuation values fixed, $\Delta U(S, x)$ does not vary with an individual monitoring deviation, while pivotality is strictly decreasing in the aggregate informed share ρ.[12] Hence $\mathrm{MBI}(S, \rho)$ is strictly decreasing in ρ, so the within-state participation game has at most one interior mixing condition and otherwise selects a corner solution through (10).

Across states, however, monitoring need not be monotone. The governance-stakes term $\Delta U(S, x)$ may rise with treasury size, delegate quality differentials, or perceived attack risk, while pivotality falls with platform scale through $\bar{n}(S)$. Because these forces move in opposite directions, the set

$$\{S \in \mathcal{S} : \mathrm{MBI}(S, \rho^*(S)) \geq \kappa\}$$

need not be connected. Monitoring may therefore be active in some regions, absent in others, and re-emerge when governance stakes rise sufficiently relative to scale.

Impact of Apathy on Token Pricing

The monitoring friction also affects token prices. Since the token price is the marginal value of token holdings, governance outcomes enter prices through the post-proposal distribution of future states. When

[12] Since $a_\rho(\rho, \theta) = 1 - \theta$ and $b_\rho(\rho, \theta) = -(1 - \theta)$, $\partial_\rho[a(\rho, \theta)b(\rho, \theta)] = (1 - \theta)[b(\rho, \theta) - a(\rho, \theta)] < 0$ whenever $\theta > 1/2$. Hence the Skellam tie probability in (8) is strictly decreasing in ρ.

monitoring is weak, the election is less likely to select the higher-reputation delegate, perceived attack risk rises, and future platform states deteriorate. These effects are capitalized into the current token price.

The token-price law of motion follows from differentiating the user's HJB in (19) with respect to token holdings, using the pricing condition in (24). The derivation is in Appendix A.3.

Lemma 2 (Token-price law of motion)**.** *In a stationary Markov Perfect Equilibrium, the competitive token price satisfies*

$$rP(S) = \mathcal{L}^{E^*} P(S) + \lambda \left[\mathbb{E}[P(S') \mid S, \rho^*(S)] - P(S) + \rho^*(S) \partial_x \mathrm{MBI}(S, \rho^*(S)) \right]. \tag{38}$$

Here S' is the post-proposal state induced by the voting outcome and perceived attack risk.

Equation (38) embeds governance risk into token prices through two channels. First, governance outcomes affect the distribution of future platform states: retaining the incumbent or electing the challenger changes expected continuation values, attack risk, and therefore future prices $P(S')$. Second, monitoring incentives enter through the pivotality-adjusted value of selecting the superior delegate, captured by the term $\rho^*(S)\partial_x \mathrm{MBI}(S, \rho^*(S))$. Thus token prices internalize both platform fundamentals and governance frictions.

Proposition 5 (Rational apathy and token prices)**.** *Fix S and suppose $\theta > 1/2$. If $\partial_x \Delta U(S, x)$ is bounded, then the monitoring component of the pricing equation decays at the same rate of pivotality. Formally,*

$$\rho^*(S)\partial_x \mathrm{MBI}(S, \rho^*(S)) = O\big(\bar{n}(S)^{-1/2} e^{-\xi(\rho^*(S),\theta)\bar{n}(S)}\big).$$

Proposition 5 connects the voting friction to asset pricing. The collapse of pivotality does not merely reduce participation; it changes the state-contingent value of the token. When the community becomes large, individual tokenholders stop acquiring information, the election relies more heavily on noisy reputation signals, and expected governance discipline weakens. The resulting increase in governance risk lowers the expected continuation price $\mathbb{E}[P(S') \mid S, \rho^*(S)]$ relative to an attentive governance benchmark. Hence rational apathy generates a governance discount in token prices.

4.3 Attack Incentives

As tokenholders become apathetic, the malicious incumbent faces the opposite problem: rather than deciding whether to monitor, he chooses whether to preserve the franchise value of office or to exercise the one-shot option to expropriate the treasury. In the interior region, the equilibrium attack probability is pinned down by the indifference condition already embedded in equation (15): conditional on reelection, the malicious delegate must be indifferent between continuing in office and attacking immediately.

The prize versus the franchise value. The attack decision trades off the current *prize* of expropriation against the *franchise value* of future incumbency. A larger treasury raises the immediate gain from attack. By contrast, stronger platform fundamentals raise the value of remaining in office, because continued incumbency becomes more valuable when future platform quality and continuation payoffs are higher. The attack option is therefore most attractive when short-run extractable value is high relative to the continuation value of governance.

Reputation and tenure. Beliefs affect attack incentives through expected tenure. A higher belief that the incumbent is honest improves reelection prospects and enlarges the shadow value of future office, which discourages attack. By contrast, a stronger challenger reduces the incumbent's expected tenure and compresses the value of waiting, which makes immediate expropriation more attractive. Thus, reputation is disciplining only insofar as it preserves future rents.

Proposition 6 (Comparative statics of the attack margin)**.** *In a stable value-accretive Markov Perfect Equilibrium, suppose that the malicious incumbent's continuation value is increasing in the franchise value of office and that the post-attack state does not inherit the incumbent's current reputation. Then, in any regular neighborhood of the attack threshold,*

$$\partial_T \Delta_I(S) > 0, \qquad \partial_Q \Delta_I(S) < 0, \qquad \partial_p \Delta_I(S) < 0, \qquad \partial_{\tilde{p}} \Delta_I(S) > 0.$$

Consequently, under any stable smooth perturbation of the threshold rule in (16)*, the selected attack probability is increasing in T, decreasing in Q, decreasing in p, and increasing in $\tilde{p}$.*

Proposition 6 identifies the platform's window of vulnerability: attack incentives are strongest when the treasury is large, but the incumbent's franchise value is weak—either because fundamentals have deteriorated or because expected tenure has fallen. In that region, the immediate gain from expropriation dominates the value of preserving office.

No-attack region. The attack-margin comparative statics characterize the regions of the state space in which no-attack discipline can be sustained. From the threshold rule in (16), a malicious incumbent preserves the platform whenever the continuation value of remaining in office exceeds the value of immediate extraction. Equivalently, no attack is sustained when $\Delta_I(S) < 0$, while attacks occur when $\Delta_I(S) > 0$. The boundary $\Delta_I(S) = 0$ is the attack frontier. The relationship between state evolution and the no-attack region can be immediately signed by the previous propostion.

Corollary 3 (Determinants of the attack region)**.** *Under the conditions of Proposition 6, the incumbent no-attack region expands with platform quality Q and incumbent reputation p, and shrinks with treasury size T and challenger reputation $\tilde{p}$. Equivalently, the attack region is reached when the treasury prize is large relative to the incumbent's franchise value of office.*

Corollary 3 shows that no-attack discipline is state contingent. The same platform can be disciplined when reputation and expected tenure are high, but vulnerable when treasury assets accumulate or when the incumbent's effective franchise value falls. Governance attacks are therefore endogenous threshold events: they occur when the state crosses the attack frontier.

5 Governance Dynamics

The preceding section characterized the local incentives of tokenholders and delegates. We now study how these equilibrium policies shape the law of motion of the platform state. The central result of this section is that anonymous delegated governance creates a negative feedback between platform growth and governance discipline: Reputation raises growth by increasing the incumbent's expected tenure, effort, and franchise value. Yet growth weakens governance discipline because higher quality expands

adoption, lowers individual pivotality, and reduces tokenholders' incentives to monitor. The platform is therefore most difficult to discipline precisely when it has become more valuable.

This section derives that feedback formally. We first write the expected law of motion induced by equilibrium monitoring, attack policies, and the belief dynamics for delegate reputation. We then show that the reduced system in platform quality and incumbent reputation has a negative feedback structure. Finally, we identify the economic conditions under which governance fragility is absorbed through stable adjustment path and those under which it produces recurrent boom–bust dynamics.

Equilibrium State Dynamics

We begin by collecting the equilibrium laws of motion for the platform state. The only new object needed for this aggregation is the endogenous attack hazard. Conditional on a governance proposal, an attack can occur either if the incumbent is retained and attacks, or if the challenger is elected and attacks. Hence the total attack hazard is

$$h_A(S) = \lambda \left[\pi_I(S,\rho^*)\hat{q}_I(S) + \big(1-\pi_I(S,\rho^*)\big)\hat{q}_C(S)\right]. \tag{39}$$

This hazard converts the proposal-time attack probabilities into a continuous-time exposure to governance failure.

Using $h_A(S)$, the expected laws of motion for platform quality and treasury assets are

$$\dot{Q}_t = Q_t\big(\mu(E_t^*) - \kappa\big) - \varphi Q_t h_A(S_t), \tag{40}$$

$$\dot{T}_t = f(Q_t)\bar{n}(S_t) - \omega(Q_t) - T_t h_A(S_t). \tag{41}$$

The incumbent's reputation evolves through the proposal and turnover events described above. Its law of motion is

$$\begin{aligned}\dot{p}_t = {} & \lambda\pi_I(S_t,\rho^*)\left[(1-\hat{q}_I(S_t))(p^+ - p_t) + \hat{q}_I(S_t)(p_0 - p_t)\right] \\ & + \lambda\big(1-\pi_I(S_t,\rho^*)\big)(\tilde{p}^+ - p_t) + \delta(\tilde{p}_t - p_t).\end{aligned} \tag{42}$$

The first term captures reputation updating when the incumbent is retained, the second captures replacement through voting, and the third captures replacement through exogenous turnover.

Equations (40)–(42) define a system of ordinary differential equations (ODEs) for the evolution of the platform state under equilibrium policies. To study these dynamics, consider an interior stationary point

$$\bar{S} = (\bar{Q}, \bar{T}, \bar{p}, \tilde{\bar{p}})$$

at which the expected drift of the state is zero. The local behavior of the economy around $\bar{S}$ is governed by the Jacobian of the ODE system:

$$J_{(Q,T,p)} = \begin{pmatrix} \frac{\partial \dot{Q}}{\partial Q} & \frac{\partial \dot{Q}}{\partial T} & \frac{\partial \dot{Q}}{\partial p} \\ \frac{\partial \dot{T}}{\partial Q} & \frac{\partial \dot{T}}{\partial T} & \frac{\partial \dot{T}}{\partial p} \\ \frac{\partial \dot{p}}{\partial Q} & \frac{\partial \dot{p}}{\partial T} & \frac{\partial \dot{p}}{\partial p} \end{pmatrix}_{\bar{S}}. \tag{43}$$

The entries of this matrix describe how a local change in one state variable affects the expected drift of the others. The next subsection uses this Jacobian to isolate the central feedback loop of the model: reputation supports growth and treasury accumulation, while growth and treasury accumulation weaken monitoring incentives and weakens governance.

5.1 The Negative Growth–Governance Feedback Loop

We now use the Jacobian in (43) to isolate the main feedback mechanism of the model. The relevant interaction is not only between platform quality and reputation, but between platform fundamentals more broadly and governance discipline. Platform success has two state dimensions: quality Q, which determines user surplus and adoption, and treasury assets T, which measure the liquid resources exposed to opportunistic extraction. Both dimensions are affected by incumbent reputation, and both feed back into the evolution of reputation.

The first direction of the feedback runs from reputation to fundamentals. A higher incumbent reputation raises expected tenure and increases the franchise value of office. This strengthens effort incentives and lowers the equilibrium attack hazard. As a result, higher reputation increases expected

quality growth and protects treasury accumulation from expected depletion. In terms of the Jacobian,

$$J_{13} = \frac{\partial \dot{Q}}{\partial p} > 0, \qquad J_{23} = \frac{\partial \dot{T}}{\partial p} > 0.$$

Reputation is therefore productive: it supports both the growth of platform quality and the preservation of platform resources.

The second direction runs from fundamentals back to reputation. Higher quality expands adoption and therefore increases the expected voting population. Since individual pivotality declines with the size of the voting population, tokenholders' incentives to monitor fall as the platform grows. This weakens electoral discipline and reduces the expected drift of incumbent reputation. Similarly, a larger treasury raises the prize from opportunistic extraction. By increasing the attack margin, treasury accumulation makes reputation more fragile because attacks reset beliefs to the prior. Hence

$$J_{31} = \frac{\partial \dot{p}}{\partial Q} < 0, \qquad J_{32} = \frac{\partial \dot{p}}{\partial T} < 0.$$

Platform success therefore weakens the governance discipline that allowed success to accumulate in the first place.

The model thus features two negative feedback channels. The quality–reputation channel captures the scale effect: quality raises adoption, adoption lowers pivotality, and lower pivotality weakens monitoring. The treasury–reputation channel captures the extraction-prize effect: treasury accumulation raises the value of attacking, and attacks destroy reputation. Together, these channels imply that anonymous delegated governance becomes harder to discipline precisely when the platform has become more valuable.

Proposition 7 (Negative feedback between fundamentals and governance discipline)**.** *Consider an interior stationary point $\bar{S}$ of the ODE system* (40)–(42)*. Suppose the equilibrium features an interior effort choice, locally positive monitoring incentives, and locally differentiable equilibrium policies. Suppose further that the attack-margin comparative statics in Proposition 6 hold in a regular neighborhood of $\bar{S}$. Then the Jacobian in* (43) *satisfies*

$$J_{13} > 0, \qquad J_{23} > 0, \qquad J_{31} < 0, \qquad J_{32} < 0.$$

Consequently, both feedback channels between platform growth and its governance are negative:

$$J_{13}J_{31} < 0, \qquad J_{23}J_{32} < 0.$$

The proof is provided in Appendix A.5. The proposition formalizes the paper's central dynamic mechanism. Reputation increases the expected growth and preservation of platform fundamentals, but stronger fundamentals undermine the incentives that sustain reputation. Quality weakens governance discipline through rational apathy, while treasury accumulation weakens discipline by increasing the value of opportunistic extraction. Anonymous delegated governance therefore contains an internal source of fragility: the same success that raises the value of the platform also erodes the private incentives required to protect it.

5.2 Dynamic Regimes

The negative feedback characterized in Proposition 7 does not, by itself, imply cyclical dynamics. It implies that platform fundamentals and governance discipline move against each other locally: reputation supports platform success, while platform success weakens the discipline that sustains reputation. Whether this feedback is absorbed smoothly or generates recurrent boom–bust dynamics depends on its strength relative to the direct stabilizing forces in the system.

To make this precise, consider a reduced two-dimensional subsystem consisting of one platform fundamental $Z \in \{Q, T\}$ and incumbent reputation p. The local Jacobian of this subsystem is

$$J_{(Z,p)} = \begin{pmatrix} J_{ZZ} & J_{Zp} \\ J_{pZ} & J_{pp} \end{pmatrix}_{\bar{S}}, \tag{44}$$

where $Z = Q$ gives $J_{Zp} = J_{13}$ and $J_{pZ} = J_{31}$, while $Z = T$ gives $J_{Zp} = J_{23}$ and $J_{pZ} = J_{32}$. By Proposition 7,

$$J_{Zp} > 0, \qquad J_{pZ} < 0, \qquad J_{Zp}J_{pZ} < 0.$$

Thus, each reduced subsystem contains the same negative feedback structure: reputation raises the drift

of the fundamental, while the fundamental lowers the drift of reputation. The eigenvalues of $J_{(Z,p)}$ solve

$$\xi^2 - \mathrm{tr}(J_{(Z,p)})\xi + \det(J_{(Z,p)}) = 0.$$

The associated discriminant is:

$$\Delta_Z = (J_{ZZ} - J_{pp})^2 + 4J_{Zp}J_{pZ}. \tag{45}$$

Since $J_{Zp}J_{pZ} < 0$, the feedback term lowers the discriminant. The first term, $(J_{ZZ} - J_{pp})^2$, captures the difference in the direct adjustment speeds of the fundamental and reputation. The second term, $4J_{Zp}J_{pZ}$, captures the strength of the negative feedback between platform success and governance discipline.

Lemma 3 (Local dynamic regimes)**.** *Consider the reduced subsystem (Z, p), with $Z \in \{Q, T\}$, around an interior stationary point $\bar{S}$. Suppose direct adjustment is locally stabilizing, $J_{ZZ} < 0$ and $J_{pp} < 0$, and suppose the feedback signs in Proposition 7 hold. Then the stationary point of the reduced subsystem is locally stable. Moreover:*

1. *If $(J_{ZZ} - J_{pp})^2 > 4|J_{Zp}J_{pZ}|$, then $\Delta_Z > 0$, the eigenvalues are real, and local adjustment is monotone.*
2. *If $(J_{ZZ} - J_{pp})^2 < 4|J_{Zp}J_{pZ}|$, then $\Delta_Z < 0$, the eigenvalues are complex conjugates, and local adjustment is oscillatory.*

Lemma 3 gives the precise sense in which the model generates governance cycles. Cycles arise when the feedback between platform success and governance discipline is strong relative to direct mean reversion. In that case, reputation and fundamentals do not return smoothly to the stationary point after a perturbation. Instead, the system overshoots. Reputation supports quality growth and treasury accumulation; stronger fundamentals reduce monitoring incentives and increase extraction incentives; weaker governance discipline raises attack exposure; attacks and reputation resets then reduce fundamentals; and lower fundamentals restore monitoring incentives, restarting the cycle.

When the feedback is weaker, the same economic forces remain present but are absorbed through stable adjustment. The platform still suffers a governance drag, because attack risk and weak monitoring

reduce the level of fundamentals relative to a frictionless benchmark. However, the adjustment path does not display recurrent boom–bust dynamics.

Determinants of the Dynamic Regime

Lemma 3 shows that the local regime is governed by a simple comparison. Direct dampening is summarized by $(J_{ZZ} - J_{pp})^2$, while the strength of the feedback loop is summarized by $|J_{Zp}J_{pZ}|$. The economy is more likely to display monotone adjustment when direct dampening dominates, and more likely to display oscillatory adjustment when the feedback between fundamentals and reputation is strong.

The model primitives shift this comparison through the four Jacobian terms. Parameters that increase J_{Zp} strengthen the effect of reputation on fundamentals. Parameters that increase $|J_{pZ}|$ strengthen the reverse effect of fundamentals on governance discipline. By contrast, parameters that make J_{ZZ} and J_{pp} more negative increase direct mean reversion and dampen the feedback.

For the quality–reputation subsystem, the feedback is strongest when reputation has a large effect on effort and attack risk, and when quality has a large effect on adoption and monitoring incentives. The first channel follows from the incumbent effort condition,

$$Q\mu'(E^*(S))\partial_Q V_I(S) = 1.$$

When effort is more responsive to the franchise value of office, a change in reputation translates into a larger change in quality growth. The second channel follows from the adoption closure,

$$\bar{n}(S) = \alpha \bar{U}(S)^+,$$

and from Proposition 4: when adoption is highly responsive to quality, a rise in Q produces a larger voting population, lowers pivotality more sharply, and weakens monitoring more strongly.

For the treasury–reputation subsystem, the feedback is strongest when treasury accumulation sharply raises the value of opportunistic extraction. Since the attack payoff is $\Pi(S) = T$, Proposition 6 implies that a larger treasury raises the attack margin. Hence a higher sensitivity of attack incentives to treasury size increases $|J_{pT}|$. Conversely, higher incumbent reputation lowers the attack hazard and protects the

treasury from expected depletion, increasing J_{Tp}.

These observations imply the following comparative-static statement.

Proposition 8 (Economic determinants of dynamic regimes)**.** *Fix an interior stationary point $\bar{S}$ and suppose the conditions of Lemma 3 hold. The economy is more likely to be in the oscillatory regime when model primitives increase the strength of the feedback term $|J_{Zp}J_{pZ}|$ relative to direct dampening $(J_{ZZ} - J_{pp})^2$. In particular:*

1. *higher adoption responsiveness α strengthens the quality-to-reputation channel by increasing the effect of Q on the voting population and therefore on rational apathy;*
2. *greater effort responsiveness to the franchise value of office strengthens the reputation-to-quality channel by increasing the effect of p on equilibrium effort;*
3. *greater sensitivity of attack incentives to treasury size strengthens the treasury-to-reputation channel by increasing the effect of T on the attack hazard;*
4. *a larger attack damage parameter φ amplifies the effect of governance failures on fundamentals;*
5. *stronger direct mean reversion in fundamentals or reputation, for example through higher depreciation κ or higher exogenous turnover δ, makes monotone adjustment more likely.*

Proposition 8 gives the economic interpretation of the local-regime condition. Oscillatory governance dynamics are most likely in platforms where user adoption is highly elastic to quality, delegate effort is highly sensitive to reputation, treasury accumulation strongly increases extraction incentives, and attacks impose large losses on fundamentals. These are precisely the environments in which platform success quickly feeds back into weaker discipline. By contrast, when adoption is less elastic, effort is less responsive, depreciation is stronger, or turnover limits the duration of incumbency, the same governance friction is absorbed through monotone adjustment rather than recurrent boom–bust dynamics.

Figure A.2 about here

5.3 Welfare Loss from Tokenized Governance Fragility

The previous subsections characterize how anonymous delegated governance generates a negative feedback between platform growth and governance discipline. We now quantify the welfare cost of this mechanism. The comparison is not with a first-best planner, but with the authenticated counterfactual introduced above. This is the relevant institutional benchmark for the model: authentication makes delegate identity permanent, allows exclusion after misconduct, and removes the renewable attack option generated by anonymous re-entry.

Let $R \in \{A, N\}$ index the governance regime, where A denotes the anonymous baseline and N denotes authenticated governance. Under anonymous governance, malicious delegates may build reputation, attack, and later re-enter under a new identity. Under authenticated governance, attackers are permanently excluded. We therefore treat the authenticated counterfactual as a no-attack regime in which tokenholders do not need to acquire costly information to discipline anonymous re-entry. Thus, relative to the anonymous baseline, authentication removes both the attack-and-reentry channel and the monitoring burden generated by anonymity.

Welfare calculation. For each regime R, let Φ^R denote the invariant distribution induced by the corresponding equilibrium state process. Stationary flow welfare is

$$g^R(S) = \bar{n}^R(S)u(Q) - E_R^*(S) - \lambda \bar{n}^R(S)\rho_R^*(S)c. \tag{46}$$

The first term is aggregate platform surplus. The second and third terms are real resource costs: delegate effort and tokenholder information acquisition. Transfers between users, delegates, and the treasury are excluded from welfare. Treasury diversion matters only through its effect on future payoff-relevant states.

Long-run average welfare under regime R is

$$\overline{W}^R = \int_{\mathcal{S}} g^R(S)\, d\Phi^R(S). \tag{47}$$

Equivalently, if the economy is initialized from its invariant distribution, discounted welfare is $\overline{W}^R/r$. Since the discount factor is common across regimes, the welfare comparison can be expressed in terms

of average flow welfare.

Sources of inefficiency. The welfare loss from anonymous governance has two sources. The first is the attack-and-reentry mechanism. Under anonymity, reputation can be used instrumentally: a malicious delegate may behave well, increase platform quality, accumulate control over a larger treasury, and then attack while preserving the possibility of future re-entry. The attack is costly not because treasury diversion is itself a social loss, but because it changes the payoff-relevant state of the platform. When an attack occurs, the economy moves from the no-attack continuation state $S^+ = (Q, T, p^+, \tilde{p})$ to the attack state $S^- = \big((1-\varphi)Q, 0, p_0, \tilde{p}\big)$. This jump lowers future welfare by reducing quality, adoption, reputation, effort incentives, and governance stability. The attack loss is therefore the endogenous continuation-welfare loss $\mathcal{W}^A(S^+) - \mathcal{W}^A(S^-)$, not an additional primitive.

The second source is rational apathy. Because individual pivotality declines with the size of the voting population, tokenholders underinvest in information as the platform grows. Anonymous governance therefore creates a real monitoring cost whenever tokenholders acquire information, and a disciplinary failure whenever monitoring becomes too weak to constrain opportunistic delegates. These two effects are jointly determined: costly monitoring affects the probability of attacks, while attack risk changes the stationary distribution of platform fundamentals.

Deadweight loss measurement. We measure the deadweight loss from anonymous governance as the percentage reduction in stationary welfare relative to authenticated governance:

$$\mathcal{L} = 100 \times \frac{\overline{W}^N - \overline{W}^A}{\overline{W}^N}. \tag{48}$$

This object asks how much long-run platform surplus is lost because delegate identity is not permanent and misconduct does not lead to exclusion.

In the simulations, we also compute related quantities that help interpret the magnitude of $\mathcal{L}$. The first is the stationary monitoring burden,

$$\int_{\mathcal{S}} \lambda \bar{n}^A(S) \rho_A^*(S) c \, d\Phi^A(S),$$

which measures the average real resource cost of information acquisition under anonymous governance. The second is the attack frequency under the stationary distribution, obtained by averaging the anonymous-regime attack hazard over Φ^A. The third is the average continuation-welfare loss conditional on attack, measured by $\mathcal{W}^A(S^+) - \mathcal{W}^A(S^-)$ at attack states. Together, these quantities distinguish costly monitoring from the state damage created by attacks.

These diagnostic quantities are not additive components of $\mathcal{L}$. Monitoring, attacks, effort, adoption, and reputation jointly determine the invariant distribution Φ^A. The quantities are therefore used to interpret the simulations rather than to mechanically decompose welfare. They show whether the welfare loss is driven mainly by costly information acquisition, by frequent attacks, by large continuation losses conditional on attack, or by the interaction between these forces.

Figure A.3 about here

6 Extensions

In this section, we extend the baseline model to explore two critical design features of token-based platforms: token-denominated compensation and endogenous treasury investment. To maintain narrative focus, we present the modified state spaces, laws of motion, and key marginal conditions below, deferring the full Hamilton-Jacobi-Bellman (HJB) system formulations to the Appendix.

6.1 Token Compensation of Delegates

We first extend the baseline model by allowing delegate compensation to include a token grant $x_D > 0$ issued at governance proposal times, in addition to the numeraire wage flow $\omega(Q)$. The purpose of this extension is to tie the delegate's continuation payoff to the long-run value of the platform, thereby raising the opportunity cost of opportunistic attacks.

The aggregate state expands to include total token supply X_t and the token stake held by the current

incumbent delegate D_t:

$$S_t = (Q_t, T_t, X_t, D_t, p_t, \tilde{p}_t). \tag{49}$$

While the continuous continuous-time laws of motion for Q_t and T_t remain unchanged, the token variables evolve discretely at governance events. If the incumbent is retained and does not attack, the token supply and the incumbent's stake update according to:

$$X^+ = X + x_D, \qquad D^+ = D + x_D. \tag{50}$$

If the incumbent executes an attack, she is assumed to forfeit her unvested token stake ($D^- = 0$), though the total supply remains X.

Token compensation alters the malicious delegate's attack incentive by embedding the token price $P(S)$ into her continuation value. Let $\Psi^k(S; D) \equiv DP(S) + V_C^k(S)$ denote the continuation value of a delegate out of office holding D tokens. Conditional on retaining control, the malicious incumbent's modified attack surplus becomes:

$$\Delta_I^D(S) \equiv \Pi(S) + \Psi^M(S^-; 0) - V_I^M(S^+; D + x_D). \tag{51}$$

Relative to the baseline model, token compensation lowers the attack surplus whenever the retained token stake raises the continuation value of staying in office ($\partial_D V_I^M > 0$). This directly shifts the attack threshold.

Proposition 9 (Primacy of Token Compensation)**.** *Suppose token prices are strictly increasing in platform quality. By forfeiting her token stake after an attack, increasing the vested token grant x_D strictly relaxes the malicious delegate's no-attack constraint ($\Delta_I^D(S) \leq 0$). Moreover, for sufficiently patient delegates (low discount rate r), token compensation is a more cost-effective margin than current numeraire wages for increasing the continuation value of not attacking.*

Proposition 9 rationalizes the empirical dominance of token-based pay in DAOs. Cash wages are short-duration flow transfers, whereas tokens are long-duration claims on the future sequence of platform fundamentals. However, because token grants are issued rather than paid out of the treasury, this creates

a pricing feedback loop: continuous issuance dilutes X, meaning optimal compensation design must dynamically balance disciplinary stakes against aggregate token dilution.

6.2 Endogenous Treasury Investment

We now endogenize the platform's capital allocation by allowing the leading delegate to choose, alongside effort e_t, a treasury-financed investment level $I_t \in [0, T_t]$. Investment acts as a productive input that increases future platform quality but depletes current liquid resources.

This dual role fundamentally alters the continuous-time laws of motion. Between proposal times, the generator $\mathcal{L}_{e,I}$ is driven by the modified drifts:

$$\dot{Q}_t = Q_t\big(\mu(e_t, I_t) - \kappa\big), \tag{52}$$

$$\dot{T}_t = f(\bar{n}(Q_t)) - \omega(Q_t) - I_t. \tag{53}$$

If investment is unobservable to tokenholders, the delegate chooses I_t to maximize her private continuation value. Let $F^k(I; S)$ denote the type-k delegate's marginal investment payoff, which balances the gain from expanding future surplus through higher quality against the loss of immediate liquid resources:

$$F^k(I; S) \equiv Q\mu_I(e_k^*(S), I)\partial_Q V_I^k(S) - \partial_T V_I^k(S), \qquad k \in \{H, M\}. \tag{54}$$

The socially aligned honest delegate sets $F^H(I_H^*(S); S) = 0$. The malicious delegate, however, values both the growth channel (which increases the future extractable prize) and the liquidity-hoarding channel (which preserves immediate treasury funds). To formalize this distortion, we evaluate the malicious delegate's marginal incentive at the honest delegate's optimal investment schedule, defining the malicious wedge:

$$G(S) \equiv F^M(I_H^*(S); S). \tag{55}$$

By the strict concavity of the investment problem, $G(S) > 0 \implies I_M^*(S) > I_H^*(S)$ (over-investment), and $G(S) < 0 \implies I_M^*(S) < I_H^*(S)$ (under-investment).

Proposition 10 (Build-to-Steal vs. Liquidity Hoarding)**.** *Assume investment choices are unobservable. There exists a unique treasury threshold $T^\dagger(Q, p, \tilde{p}) > 0$ such that the sign of the malicious wedge $G(S)$ flips, yielding a state-contingent investment distortion:*

1. ***Small Treasury Asymptotics (Front-loading):*** *For $T < T^\dagger$, the wedge is strictly positive ($G(S) > 0$). The malicious delegate strictly over-invests relative to the honest type ($I_M^* > I_H^*$) to rapidly scale the platform and inflate the future extractable prize.*
2. ***Large Treasury Sign (Hoarding):*** *For $T > T^\dagger$, the wedge is strictly negative ($G(S) < 0$). Expropriation is already highly attractive, causing the liquidity-hoarding channel to dominate. The malicious delegate under-invests ($I_M^* < I_H^*$) to preserve immediate extractable value.*

Unlike standard corporate finance models where managers universally over-invest due to empire-building or under-invest due to financing frictions, the anonymity friction in tokenized governance generates a non-monotonic distortion. A malicious delegate may initially appear highly visionary—deploying treasury funds aggressively to build infrastructure—only to abruptly pivot to starving the protocol of productive capital once the treasury crosses the critical threshold $T^\dagger$.

7 Conclusion

This paper develops a dynamic principal–agent model of delegated governance in token-based platforms. In this environment, we show how endogenous reputation shapes both tokenholders' willingness to monitor and delegates' incentives to preserve future rents, generating a governance subgame which equilibrium outcomes feed back into platform fundamentals and token values.

We first show that honest delegates lack the option value of attacking the platform and therefore have weaker incentives to mimic high early effort. This generates a competence paradox: in states where the attack option is relevant, the malicious delegate may exert more effort than the honest one.

Our main result is that anonymous delegated governance can generate a reputation–monitoring cycle as the effort provided by a malicious delegate builds both a higher continuation value of delegates'

wages and the size of the treasury it may choose to steal away. A delegate who is malicious exerts effort, improves platform quality, and accumulates reputation. As its reputation rises and the platforms grows, tokenholders' incentives to monitor the delegate as well as their pivotality in the aggregate monitoring level fall. The resulting reduced oversight increases the attractiveness of expropriation as the expected tenure and hence the expected continuation value of wages falls, allowing governance failures to emerge endogenously after periods of apparently successful performance. Our model thus provides microfoundations for recurrent governance failures and boom–bust dynamics in token prices.

A key implication of our model is that, under anonymity, proposal frequency becomes a central determinant of governance stability through its effect on reputation dynamics. A higher proposal frequency accelerates the rate at which reputation is accumulated. Another implication of the model is that malicious delegates may rationally overinvest in their initial performance to inflate the value of the platform and their own credibility. This accelerates the rational decrease of monitoring by tokenholders and paves the way to attacks.

We further show that when delegates are sufficiently patient, token compensation is the most cost-effective way to create continuation rents. But token pay is not a panacea. As it operates through token supply and future valuations, excessive issuance can dilute incentives and weaken discipline. We finally show that when delegates control treasury-financed investment, a malicious delegate may overinvest when the treasury is small in order to build a larger future prize, but underinvest and hoard liquidity when the treasury is already large and extraction is immediately attractive.

Our model provides a clear normative implication that enhancing transparency about voting delegates can strengthen intertemporal discipline in tokenized governance by restoring the link between reputation and long-term accountability.

Finally, our analysis relies on a governance environment with two delegates and Poisson voting intensities, which ensures tractability but abstracts from richer institutional features such as multiple competing delegates, coalition formation, and strategic coordination among tokenholders. Extending the framework along these dimensions would allow future research to study how competition and coordination frictions interact with reputation and monitoring incentives.

References

Abadi, J. and Brunnermeier, M. (2024). Token-based platform governance. *Journal of Financial Economics*, 162:103951.

Appel, I. and Grennan, J. (2023). Control of decentralized autonomous organizations. In *AEA Papers and Proceedings*, volume 113, pages 182–185. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Bakos, Y. and Halaburda, H. (2022). Will blockchains disintermediate platforms? The problem of credible decentralization in DAOs. Working paper.

Berinsky, A. J., Halpern, D., Halpern, J. Y., Jadbabaie, A., Mossel, E., Procaccia, A. D., and Revel, M. (2025). Tracking truth with liquid democracy. *Management Science*, 71(9):7821–7839.

Bhandari, T., Iliev, P., and Kalodimos, J. (2021). Governance changes through shareholder initiatives: The case of proxy access. *Journal of Financial and Quantitative Analysis*, 56(5):1590–1621.

Blum, C. and Zuber, C. I. (2016). Liquid democracy: Potentials, problems, and perspectives. *Journal of Political Philosophy*, 24(2):162–182.

Burkart, M., Lee, S., and Voss, P. (2024). The evolution of the market for corporate control. *European Corporate Governance Institute–Finance Working Paper*, (956):24–05.

Catalini, C. (2017). How blockchain technology will impact the digital economy. *Blockchains Smart Contracts Internet Things*, 4:2292–2303.

Chod, J., Trichakis, N., and Yang, S. A. (2022). Platform tokenization: Financing, governance, and moral hazard. *Management Science*, 68(9):6411–6433.

Cong, L. W., Li, Y., and Wang, N. (2022). Token-based platform finance. *Journal of Financial Economics*, 144(3):972–991.

Cong, L. W., Rabetti, D., Wang, C. C., and Yan, Y. (2025). Centralized governance in decentralized organizations. Working paper.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3):393–414.

Fahlenbrach, R. and Frattaroli, M. (2021). ICO investors. *Financial Markets and Portfolio Management*, 35(1):1–59.

Feichtinger, R., Fritsch, R., Heimbach, L., Vonlanthen, Y., and Wattenhofer, R. (2024). SoK: Attacks on DAOs. Working paper.

Ferreira, D., Li, J., and Nikolowa, R. (2023). Corporate capture of blockchain governance. *Review of Financial Studies*, 36(4):1364–1407.

Gan, J., Tsoukalas, G., and Netessine, S. (2021). Initial coin offerings, speculation, and asset tokenization. *Management Science*, 67(2):914–931.

Goldstein, I., Gupta, D., and Sverchkov, R. (2024). Utility tokens as a commitment to competition. *Journal of Finance*, 79(6):4197–4246.

Grossman, S. J. and Hart, O. D. (1988). One share-one vote and the market for corporate control. *Journal of Financial Economics*, 20:175–202.

Gryglewicz, S., Mayer, S., and Morellec, E. (2021). Optimal financing with tokens. *Journal of Financial Economics*, 142(3):1038–1067.

Han, J., Lee, J., and Li, T. (2025). A review of DAO governance: Recent literature and emerging trends. *Journal of Corporate Finance*, page 102734.

Harvey, C. R., Saleh, F., and Sverchkov, R. (2025). An economic model of the L1-L2 interaction. Working paper.

Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, pages 74–91.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182.

Howell, S. T., Niessner, M., and Yermack, D. (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. *Review of Financial Studies*, 33(9):3925–3974.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Jensen, M. C. and Ruback, R. S. (1983). The market for corporate control: The scientific evidence. *Journal of Financial Economics*, 11(1-4):5–50.

Kahan, M. and Rock, E. (2008). The hanging chads of corporate voting. *Georgetown Law Journal*, 96(4):1227.

Kahng, A., Mackenzie, S., and Procaccia, A. (2021). Liquid democracy: An algorithmic perspective. *Journal of Artificial Intelligence Research*, 70:1223–1252.

Klein, B. and Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89(4):615–641.

Kreps, D. M. and Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253–279.

Laturnus, V. and Lehar, A. (2026). Decentralized autonomous organizations. Working paper.

Lehar, A. and Parlour, C. (2025). Decentralized exchange: The Uniswap automated market maker. *Journal of Finance*, 80(1):321–374.

Lehar, A., Parlour, C. A., and Yuan, K. (2026). Liquid staking. Technical report, Working Paper.

Levit, D., Malenko, N., and Maug, E. (2024). Trading and shareholder democracy. *Journal of Finance*, 79(1):257–304.

Li, J. and Mann, W. (2025). Digital tokens and platform building. *The Review of Financial Studies*, 38(7):1921–1954.

Malenko, N. and Malenko, A. (2019). Proxy advisory firms: The economics of selling information to voters. *Journal of Finance*, 74(5):2441–2490.

Malenko, N. and Shen, Y. (2016). The role of proxy advisory firms: Evidence from a regression-discontinuity design. *Review of Financial Studies*, 29(12):3394–3427.

Milgrom, P. and Roberts, J. (1982). Predation, reputation, and entry deterrence. *Journal of Economic Theory*, 27(2):280–312.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208.

Myerson, R. B. (1998). Population uncertainty and poisson games. *International Journal of Game Theory*, 27(3):375–392.

Palfrey, T. R. and Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review*, 79(1):62–78.

Prat, J., Danos, V., and Marcassa, S. (2025). Fundamental pricing of utility tokens. *Management Science*.

Rossello, R. (2024). Blockholders and strategic voting in defi daos' governance. *Available at SSRN 4706759*.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90(4):629–649.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *Quarterly Journal of Economics*, 98(4):659–679.

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

Sockin, M. and Xiong, W. (2023). Decentralization through tokenization. *Journal of Finance*, 78(1):247–299.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374.

Tirole, J. (2006). *The Theory of Corporate Finance*. Princeton University Press, Princeton, NJ.

Yermack, D. (2017). Corporate governance and blockchains. *Review of Finance*, 21(1):7–31.

A Proofs

A.1 Equilibrium Existence

Preliminaries

Throughout the appendix, the aggregate state is $\mathbf{S} = (Q, T, p, \tilde{p}) \in \mathcal{S} = \mathbb{R}^2_+ \times [0,1]^2$ as in Section 3. Between proposal times, the controlled generator acting on a function $F \in C^2(\mathcal{S})$ is the diffusion generator $\mathcal{L}^e$ defined in Equation 18. Proposal times arrive at Poisson rate λ and generate jumps according to the transition maps $\mathbf{S} \mapsto \mathbf{S}^+$, $\mathbf{S} \mapsto \mathbf{S}^-$, and $\mathbf{S} \mapsto \mathbf{S}_R$.

We impose the standard regularity conditions required for discounted infinite-horizon continuous-time control with jumps:

(i) Primitives are measurable and locally Lipschitz in $\mathbf{S}$ on $\mathrm{int}(\mathcal{S})$.

(ii) Flow payoffs and utility $u(Q)$ are bounded.

(iii) At the boundary $T = 0$, the treasury drift is nonnegative, ensuring the process does not exit $\mathcal{S}$ through $T < 0$.

Auxiliary definition (State-Constrained Viscosity Solution). A function V solves a state-constrained HJB on $\mathcal{S}$ if it is a viscosity solution on $\mathrm{int}(\mathcal{S})$ and satisfies the viscosity inequalities at boundary points $T = 0$ using only test functions consistent with admissible trajectories (no outward-pointing test functions).

Lemma 4 (Auxiliary Lyapunov condition for positive recurrence)**.** *Suppose there exists a twice continuously differentiable function $V : \mathcal{S} \to \mathbb{R}_+$ and constants $c_1, c_2 > 0$, $R < \infty$ such that $V(\mathbf{S}) \to \infty$ as $\|\mathbf{S}\| \to \infty$ and, for the controlled generator $\mathcal{G}$ of the induced jump-diffusion under stationary Markov policies,*

$$\mathcal{G}V(\mathbf{S}) \le -c_1 \quad \textit{for all } \|\mathbf{S}\| \ge R, \qquad \sup_{\|\mathbf{S}\| \le R} \mathcal{G}V(\mathbf{S}) \le c_2.$$

Then the induced Markov process is positive recurrent and admits at least one invariant probability

measure on $\mathcal{S}$.

Proof of Proposition 1

Proof. We prove existence by formalizing the fixed-point mapping between individual optimization and aggregate state distributions through four steps: (i) HJB well-posedness, (ii) coercivity and existence of optimal Markov controls, (iii) bounding the state space via an endogenous reflecting boundary to find an invariant measure, and (iv) applying a topological fixed-point theorem.

Step 1: Conditional HJBs and Viscosity Solutions. Fix an arbitrary profile of stationary Markov policies $\Sigma = (e(\mathbf{S}), q(\mathbf{S}), \rho(\mathbf{S}))$ and perceived attack rules. Because proposals arrive at a bounded Poisson rate λ, each Bellman equation (Equations 19, 29, and 30) takes the form of a discounted continuous-time HJB with a bounded non-local jump operator. Under our regularity conditions and strictly positive discounting $r > 0$, standard dynamic programming arguments for jump-diffusions guarantee that the conditional HJBs admit unique, bounded, and continuous state-constrained viscosity solutions $V_I^k, V_C^k, U \in C_b(\mathcal{S})$.

Step 2: Coercivity and Existence of Optimal Markov Controls. Given the continuous value functions from Step 1, optimal controls are selected pointwise. The action spaces for monitoring $\rho \in [0, 1]$ and attack probability $q \in [0, 1]$ are trivially compact.

For the effort choice $e \geq 0$, the incumbent's objective is to maximize:

$$\max_{e \geq 0} \Big\{ -c(e) + Q\mu(e)\partial_Q V_I(\mathbf{S}) \Big\} \tag{1}$$

Since $V_I(\mathbf{S})$ is bounded and locally Lipschitz, its generalized gradient $\partial_Q V_I$ is bounded by some constant K. By the Inada condition (Assumption 1(i)), $\lim_{e\to\infty} \mu'(e) = 0$ and $\lim_{e\to\infty} c'(e) = \infty$. Therefore, there exists an upper bound $\bar{e} < \infty$ such that the marginal benefit of effort $Q\mu'(e)\partial_Q V_I(\mathbf{S})$ is strictly less than the marginal cost $c'(e)$ for all $e > \bar{e}$.

This strict coercivity allows us to restrict the effort control space without loss of generality to the compact set $[0, \bar{e}]$. Because the control sets are compact and the Hamiltonian is continuous, Berge's Maximum Theorem ensures the argmax correspondence is non-empty, compact-valued, and upper hemicontinuous. The Measurable Selection Theorem thus guarantees the existence of measurable

stationary Markov policies $e^*(\mathbf{S}), q^*(\mathbf{S}), \rho^*(\mathbf{S})$.

Step 3: Endogenous Reflecting Boundary and Invariant Measure. Let $\mathcal{G}$ denote the extended generator of the piecewise-deterministic Markov process induced by the policies Σ. To prove the existence of an invariant measure Φ, we apply Lemma 4 by constructing the Foster-Lyapunov function $L(\mathbf{S}) = Q + T$. We must show that $\mathcal{G}L(\mathbf{S}) \leq -c_1 L(\mathbf{S}) + c_2$ outside a compact set.

1. **Quality amental Drift:** For large Q, the strict concavity of $\mu(e)$ bounded by $\mu(\bar{e})$ and the depreciation $\kappa > 0$ imply the drift $b_Q = Q(\mu(e) - \kappa)$ becomes strictly negative (Assumption 1(ii)).
2. **Treasury Drift and Electoral Collapse:** As T and Q grow, the mean-field adoption $\bar{n}(\mathbf{S}) \to \infty$. By the scale decay property, voter pivotality vanishes, causing $\rho^*(\mathbf{S}) \to 0$.
3. **The Attack Boundary:** With $\rho^* \to 0$, electoral discipline fails. The value of expropriation $\Pi(T)$ scales with T. For sufficiently large T, $\Pi(T) + V_C^M(\mathbf{S}^-) > V_I^M(\mathbf{S}^+)$, making the attack strictly optimal ($q^* \to 1$).
4. **The Jump Operator:** The expected drift from the jump operator at a proposal time is $\lambda(1 - \chi)\pi_I(\mathbf{S})q^*(\mathbf{S})[L(\mathbf{S}^-) - L(\mathbf{S})]$. Because an attack drains the treasury completely ($T^- = 0$), $L(\mathbf{S}^-) - L(\mathbf{S}) \approx -T$.

As $\|\mathbf{S}\| \to \infty$, the negative jump drift $-\lambda(1 - \chi)T$ linearly dominates the bounded continuous inflows $f\bar{n}(\mathbf{S})$. Thus, the system is strictly mean-reverting. The inevitable attack acts as an endogenous reflecting boundary, confining the process. By Lemma 4, this guarantees the existence of at least one invariant probability measure $\Phi \in \mathcal{P}(\mathcal{S})$.

Step 4: Fixed Point. We employ the ε-perturbation refinement, bounding $\Sigma \in [\varepsilon, 1 - \varepsilon]$ for strategic probabilities, which guarantees the transition kernel has full support and the invariant measure mapping is continuous in the weak topology.

Define the mapping Γ from the space of candidate policies and invariant measures to itself. By Step 2, the best-response mapping from measures to policies is upper hemicontinuous and convex-valued (allowing for mixed strategies at indifference thresholds). By Step 3, the mapping from policies to invariant measures is well-defined and continuous. The space of probability measures $\mathcal{P}(\mathcal{S})$ satisfying

our Lyapunov condition is tight, and thus relatively compact by Prohorov's theorem.

Applying the Kakutani-Fan-Glicksberg fixed-point theorem to the composite mapping Γ yields the existence of an ε-equilibrium. Taking the limit sequence $\varepsilon_n \downarrow 0$, the stability condition (Equation **??**) ensures the existence of a Stable Markov Perfect Equilibrium Σ^*. □

A.2 Delegate Effort Rankings and Screening

Proof of Lemma 1

(i) Mixing region. Consider a state S in which $q_I^*(S) \in (0,1)$. By (**??**), interior mixing implies

$$\Delta_I(S) = 0, \qquad \Delta_I(S) \equiv \Pi(S) + V_C^M(S^-) - V_I^M(S^+).$$

Using (26), the malicious incumbent's value conditional on retaining control can be written as

$$\mathcal{A}_I^M(S) = (1 - q_I^*(S))V_I^M(S^+) + q_I^*(S)\big[\Pi(S) + V_C^M(S^-)\big] = V_I^M(S^+) + q_I^*(S)\Delta_I(S).$$

Thus the attack option enters $\mathcal{E}_I^M(S)$ through $(1-\chi)\pi_I(S,\rho^*)q_I^*(S)\Delta_I(S)$. Differentiating this term with respect to Q gives

$$\partial_Q[(1-\chi)\pi_I(S,\rho^*)q_I^*(S)\Delta_I(S)] = (1-\chi)\pi_I(S,\rho^*)q_I^*(S)\partial_Q\Delta_I(S),$$

because the terms involving $\partial_Q\pi_I$ and $\partial_Q q_I^*$ are multiplied by $\Delta_I(S) = 0$. Hence the attack option has zero level value in the mixed region but a non-zero marginal value.

The primitive source of this marginal value is the treasury channel. In the model, $\Pi(S) = T$, an attack resets the treasury to zero, and honest delegates cannot appropriate treasury resources. Moreover, for any smooth F,

$$\partial_Q(\mathcal{L}^E F)(S) = \mathcal{L}^E(\partial_Q F)(S) + (\mu(E) - \kappa)\partial_Q F(S) + \partial_Q b_T(S)\partial_T F(S) + \sigma_Q^2 Q\,\partial_{QQ}F(S).$$

By Assumption 1, $\partial_Q b_T(S) = f'(Q)\bar{n}(S) + f(Q)\partial_Q\bar{n}(S) - \omega'(Q) > 0$. Thus higher Q raises the expected treasury drift, and treasury is appropriable only by a malicious incumbent.

Applying the envelope theorem to (29) gives, for $k \in \{H, M\}$,

$$(r+\zeta)\partial_Q V_I^k(S) = \omega'(Q) + \partial_Q\left(\mathcal{L}^{E_k^*} V_I^k\right)(S) + \zeta \partial_Q \mathcal{E}_I^k(S).$$

The wage term is common across types. The malicious equation contains the additional mixed-attack source $\zeta(1-\chi)\pi_I(S,\rho^*)q_I^*(S)\partial_Q\Delta_I(S)$, which is absent from the honest equation. Since $1-\chi > 0$, $\pi_I(S,\rho^*) > 0$, $q_I^*(S) > 0$, $\Pi(S) = T$, and $\partial_Q b_T(S) > 0$, this source is strictly positive.

Let $D(S) \equiv \partial_Q V_I^M(S) - \partial_Q V_I^H(S)$. Subtracting the two differentiated HJBs gives a linear discounted equation for D with a strictly positive source term $\mathcal{S}$ in the mixed region. Therefore, by the Feynman–Kac representation,

$$D(S) = \mathbb{E}_S\left[\int_0^\infty e^{-(r+\zeta)t}\mathcal{S}(S_t)\,dt\right] > 0.$$

Hence $\partial_Q V_I^M(S) > \partial_Q V_I^H(S)$.

(ii) No-attack region. Consider a state S such that $q_I^*(S) = 0$ and $\Delta_I(S) < 0$. Let $\tau_A \equiv \inf\{t \geq 0 : q_I^*(S_t) > 0\}$. Define $V_I^{M,0}$ as the value of a malicious incumbent constrained never to attack. Under this constraint, the malicious incumbent has the same payoff and feasible effort set as the honest incumbent, so $V_I^{M,0} = V_I^H$. Hence $V_I^M(S) - V_I^H(S)$ is exactly the value of the future attack option. Since this option is not exercised before τ_A,

$$V_I^M(S) - V_I^H(S) = \mathbb{E}_S\left[e^{-(r+\zeta)\tau_A}\left(V_I^M(S_{\tau_A}) - V_I^H(S_{\tau_A})\right)\mathbf{1}_{\{\tau_A<\infty\}}\right].$$

By Assumption 1, $\partial_Q b_T(S) > 0$. Thus, for $S^\varepsilon = (Q+\varepsilon, T, p, \tilde{p})$, under the same Brownian path and event realizations, the no-attack treasury paths satisfy $T_t^\varepsilon \geq T_t$ for $t < \tau_A$, with strict inequality on a set of positive probability. Since $\Pi(S) = T$, the boundary payoff $V_I^M(S_{\tau_A}) - V_I^H(S_{\tau_A})$ is increasing in the treasury coordinate. Therefore,

$$V_I^M(S_{\tau_A}^\varepsilon) - V_I^H(S_{\tau_A}^\varepsilon) \geq V_I^M(S_{\tau_A}) - V_I^H(S_{\tau_A}),$$

with strict inequality with positive probability whenever $\Pr_S(\tau_A < \infty) > 0$. Taking expectations,

dividing by ε, and letting $\varepsilon \downarrow 0$ yields

$$\partial_Q \big[V_I^M(S) - V_I^H(S)\big] > 0.$$

Hence $\partial_Q V_I^M(S) > \partial_Q V_I^H(S)$.

(iii) Attack region. Consider a state S such that $q_I^*(S) = 1$ and $\Delta_I(S) > 0$. Then, conditional on retaining control, the malicious incumbent strictly attacks, so $\mathcal{A}_I^M(S) = \Pi(S) + V_C^M(S^-)$. Equivalently, since $\mathcal{A}_I^M(S;q) = V_I^M(S^+) + q\Delta_I(S)$, at $q = 1$,

$$\partial_Q \mathcal{A}_I^M(S) = \partial_Q V_I^M(S^+) + \partial_Q \Delta_I(S) = (1-\varphi)\partial_Q V_C^M(S^-),$$

where the last equality uses $\Pi(S) = T$ and $S^- = ((1-\varphi)Q, 0, p_0, \tilde{p})$. The honest incumbent never attacks, so $\partial_Q \mathcal{A}_I^H(S) = \partial_Q V_I^H(S^+)$. Therefore, in the strict attack region, the inverted single-crossing condition reduces to

$$(1-\varphi)\partial_Q V_C^M(S^-) > \partial_Q V_I^H(S^+).$$

This condition may hold when anonymous re-entry preserves enough future option value, but it is not implied by $\Delta_I(S) > 0$ alone. Hence the ranking is automatic in the option-preserving region $q_I^*(S) < 1$, while in the strict attack region it requires the post-attack continuation comparison above.

Proof of Proposition 2

Proof. For a type $k \in \{H, M\}$ incumbent, effort affects the HJB in (29) only through the linear effort cost and the quality drift term in the generator: $-E + Q(\mu(E) - \kappa)\partial_Q V_I^k(S)$. At an interior optimum, the first-order condition is therefore $-1 + Q\mu'(E_k^*(S))\partial_Q V_I^k(S) = 0$, or equivalently,

$$Q\mu'(E_k^*(S))\partial_Q V_I^k(S) = 1.$$

Since $\mu'(E) > 0$ and $\mu''(E) < 0$, the function $\mu'(E)$ is strictly decreasing. Hence the solution $E_k^*(S)$ is strictly increasing in the marginal value $\partial_Q V_I^k(S)$. By Lemma 1, in the option-preserving region, $\partial_Q V_I^M(S) > \partial_Q V_I^H(S)$. Therefore, $E_M^*(S) > E_H^*(S)$.

More generally, whenever $\partial_Q V_I^M(S) \geq \partial_Q V_I^H(S)$, the same monotonicity of the first-order condition

implies $E_M^*(S) \geq E_H^*(S)$. □

Proof of Proposition 3

Proof. In the authenticated regime, an attack is terminal:

$$\mathcal{A}_I^M(S) = \max\{V_I^M(S^+), \Pi(S) - \psi\}.$$

Let τ be the malicious delegate's stopping time to attack, and let Λ_t denote the equilibrium displacement hazard. By the envelope theorem and Feynman–Kac representation,

$$\partial_Q V_I^H(S) = \mathbb{E}_S\left[\int_0^\infty e^{-\int_0^t (r+\Lambda_s)\,ds} \omega_Q(Q_t)\,dt\right], \tag{2}$$

$$\partial_Q V_I^M(S) = \mathbb{E}_S\left[\int_0^\tau e^{-\int_0^t (r+\Lambda_s)\,ds} \omega_Q(Q_t)\,dt + e^{-\int_0^\tau (r+\Lambda_s)\,ds} \partial_Q(\Pi(S_\tau) - \psi)\right]. \tag{3}$$

Since $\Pi(S) = T$ and ψ is independent of Q, $\partial_Q(\Pi(S_\tau) - \psi) = 0$. Hence

$$\partial_Q V_I^H(S) - \partial_Q V_I^M(S) = \mathbb{E}_S\left[\int_\tau^\infty e^{-\int_0^t (r+\Lambda_s)\,ds} \omega_Q(Q_t)\,dt\right].$$

Because $\omega_Q(Q) > 0$, the right-hand side is strictly positive whenever $\Pr_S(\tau < \infty) > 0$. Thus

$$\partial_Q V_I^H(S) > \partial_Q V_I^M(S).$$

Finally, the interior effort condition (31) is $Q\mu'(E_k^*(S))\partial_Q V_I^k(S) = 1$. Since $\mu'' < 0$, $E_k^*(S)$ is strictly increasing in $\partial_Q V_I^k(S)$. Therefore $E_H^*(S) > E_M^*(S)$. □

A.3 Voting incentives

Proof of Proposition 4

Proof. Parts (i) and (ii) follow from the Skellam expression for pivotality in (8) and the large-$\bar{n}(S)$ asymptotic in (37). If $\theta > 1/2$, then

$$a(\rho, \theta) = \rho + (1-\rho)\theta > b(\rho, \theta) = (1-\rho)(1-\theta)$$

for every $\rho \in [0,1]$. Hence

$$\kappa(\rho,\theta) = \left(\sqrt{a(\rho,\theta)} - \sqrt{b(\rho,\theta)}\right)^2 > 0,$$

and pivotality decays at the rate stated in part (i). Since $\mathrm{MBI}(S,\rho) = \mathbb{P}_{\mathrm{piv}}(S,\rho)(1-\theta)\Delta U(S,x)$ by (9), and $\Delta U(S,x)$ is bounded, the same order bound applies to $\mathrm{MBI}(S,\rho)$. This proves part (ii).

For part (iii), fix S. Holding continuation values fixed, $\Delta U(S,x)$ does not vary with ρ. Moreover,

$$\partial_\rho[a(\rho,\theta)b(\rho,\theta)] = (1-\theta)\big[b(\rho,\theta) - a(\rho,\theta)\big] < 0,$$

because $a(\rho,\theta) > b(\rho,\theta)$. Since the Skellam tie probability in (8) is increasing in $\sqrt{a(\rho,\theta)b(\rho,\theta)}$, pivotality is strictly decreasing in ρ. Therefore $\mathrm{MBI}(S,\rho)$ is strictly decreasing in ρ, so the equation $\mathrm{MBI}(S,\rho) = c$ has at most one solution. Together with the threshold rule in (10), this gives uniqueness of the monitoring choice, up to the knife-edge indifference cases selected by the stability refinement.

For part (iv), since $\kappa(\rho,\theta)$ is continuous on $[0,1]$ and strictly positive for $\theta > 1/2$, there exists

$$\underline{\kappa}(\theta) \equiv \min_{\rho\in[0,1]} \kappa(\rho,\theta) > 0.$$

Using the boundedness of $\Delta U(S,x)$, part (ii) implies that $\sup_{\rho\in[0,1]} \mathrm{MBI}(S,\rho) \to 0$ as $\bar{n}(S) \to \infty$. Hence, for any $c > 0$, there exists $\bar{N} < \infty$ such that if $\bar{n}(S) > \bar{N}$, then $\mathrm{MBI}(S,\rho) < c$ for all $\rho \in [0,1]$. By (10), the unique equilibrium monitoring choice is then $\rho^*(S) = 0$. □

Proof of Lemma 2.

At equilibrium monitoring $\rho^*(S)$, the user's HJB is

$$rU(S,x) = u(Q) - f(Q) + \mathcal{L}^{E^*}U(S,x) + \delta\big[xP(S) - U(S,x)\big] + \lambda\big[\mathcal{E}_{\mathrm{vote}}(S,x,\rho^*) - U(S,x)\big]. \quad (4)$$

Competitive pricing implies $P(S) = \partial_x U(S,x)$. Differentiating the HJB with respect to x, the left-hand side becomes $rP(S)$. The flow payoff $u(Q) - f(Q)$ does not depend on x, and the generator commutes with ∂_x, since $\mathcal{L}^{E^*}$ acts on the state variables rather than on token holdings:

$$\partial_x\big[\mathcal{L}^{E^*}U(S,x)\big] = \mathcal{L}^{E^*}\partial_x U(S,x) = \mathcal{L}^{E^*}P(S).$$

The liquidity-shock term drops out because an exiting user receives the market value of her token holdings:

$$\partial_x \left\{ \delta \big[xP(S) - U(S,x) \big] \right\} = \delta \big[P(S) - \partial_x U(S,x) \big] = 0.$$

Therefore,

$$rP(S) = \mathcal{L}^{E^*} P(S) + \lambda \left[\partial_x \mathcal{E}_{\text{vote}}(S, x, \rho^*) - P(S) \right]. \tag{5}$$

It remains to compute $\partial_x \mathcal{E}_{\text{vote}}(S, x, \rho^*)$. From (20),

$$\partial_x \mathcal{E}_{\text{vote}}(S, x, \rho^*) = \pi_I(S, \rho^*) \partial_x \mathcal{U}_I(S, x) + \big(1 - \pi_I(S, \rho^*)\big) \partial_x \mathcal{U}_C(S, x) + \rho^*(S) \partial_x \text{MBI}(S, \rho^*),$$

where the equilibrium monitoring policy is held fixed by the envelope argument. For each delegate $\ell \in \{I, C\}$, (21) implies

$$\partial_x \mathcal{U}_\ell(S, x) = (1 - \hat{q}_\ell(S)) P(S_\ell^+) + \hat{q}_\ell(S) P(S_\ell^-).$$

Thus,

$$\pi_I(S, \rho^*) \partial_x \mathcal{U}_I(S, x) + \big(1 - \pi_I(S, \rho^*)\big) \partial_x \mathcal{U}_C(S, x) = \mathbb{E}[P(S') \mid S, \rho^*(S)] \,.$$

Combining the previous two expressions gives

$$\partial_x \mathcal{E}_{\text{vote}}(S, x, \rho^*) = \mathbb{E}[P(S') \mid S, \rho^*(S)] + \rho^*(S) \partial_x \text{MBI}(S, \rho^*(S)).$$

Substituting this expression into the differentiated HJB yields

$$rP(S) = \mathcal{L}^{E^*} P(S) + \lambda \left[\mathbb{E}[P(S') \mid S, \rho^*(S)] + \rho^*(S) \partial_x \text{MBI}(S, \rho^*(S)) - P(S) \right],$$

which rearranges to equation (38). □

A.4 Attack incentives

Proof of Proposition 6

Proof. The malicious incumbent's attack margin is

$$\Delta_I(S) = \Pi(S) + V_C^M(S^-) - V_I^M(S^+).$$

The attack policy in (**??**) is increasing in this margin. Hence it is enough to sign the derivatives of Δ_I.

For the treasury, $\Pi(S) = T$ and an attack resets the treasury in S^-. Therefore

$$\partial_T \Delta_I(S) = 1 + \partial_T V_C^M(S^-) - \partial_T V_I^M(S^+).$$

Since S^- has zero treasury and the current treasury enters the attack payoff one-for-one, while the continuation value of office does not inherit the current treasury one-for-one, the direct extraction effect dominates in a regular neighborhood of the attack threshold. Hence $\partial_T \Delta_I(S) > 0$.

For quality,

$$\partial_Q \Delta_I(S) = \partial_Q \Pi(S) + (1-\varphi)\partial_Q V_C^M(S^-) - \partial_Q V_I^M(S^+).$$

Because $\Pi(S) = T$, $\partial_Q \Pi(S) = 0$. Higher quality raises the franchise value of continuing in office, whereas an attack moves the platform to the damaged state $S^- = ((1-\varphi)Q, 0, p_0, \tilde{p})$. A marginal increase in Q reduces attack incentives whenever it raises the franchise value of continued incumbency more than the value of post-attack re-entry:

$$\partial_Q V_I^M(S^+) > (1-\varphi)\partial_Q V_C^M(S^-).$$

For beliefs, the attack payoff $\Pi(S) = T$ is independent of p and $\tilde{p}$. A higher incumbent reputation p raises expected tenure and therefore the value of continuing in office relative to attacking. Thus $\partial_p V_I^M(S^+) > \partial_p V_C^M(S^-)$, which gives $\partial_p \Delta_I(S) < 0$. Conversely, a higher challenger reputation $\tilde{p}$ weakens the incumbent's expected tenure and improves the value of re-entry after an attack relative to the value of preserving incumbency. Hence $\partial_{\tilde{p}} \Delta_I(S) > 0$.

Finally, under the stable perturbation refinement, the selected attack rule can be represented locally

as a smooth increasing function of the attack margin, $q_I^{\varepsilon}(S) = h_\varepsilon(\Delta_I(S))$ with $h'_\varepsilon > 0$. Therefore the selected attack probability inherits the signs of the attack-margin derivatives. Taking the stable limit yields the stated comparative statics away from knife-edge discontinuities. □

A.5 Governance Cycles

Proof of Proposition 7

Proof. By (22) and Proposition 6, higher incumbent reputation lowers perceived attack risk, $\partial_p \hat{q}_I(S) < 0$. From (21), this raises the user's continuation value from retaining the incumbent and therefore increases the retention probability, $\partial_p \pi_I(S, \rho^*) > 0$. The implied expected tenure,

$$\mathcal{T}_I(S) = \left[\delta + \lambda(1 - \pi_I(S, \rho^*))\right]^{-1},$$

is therefore increasing in p. Since $\mathcal{E}_I^k(S)$ in (28) weights the value of remaining in office by $\pi_I(S, \rho^*)$, the franchise value of office is increasing in incumbent reputation.

All derivatives are evaluated at the interior stationary point $\bar{S}$. From the expected laws of motion (40)–(41), By Proposition 6, $\partial_p h_A(S) < 0$. Moreover, higher reputation raises expected tenure and the franchise value of office. By the effort condition (31), $\partial_p E^*(S) > 0$. Hence

$$J_{13} = \partial_p \dot{Q} = Q\mu'(E^*)\partial_p E^* - \varphi Q \partial_p h_A > 0,$$

$$J_{23} = \partial_p \dot{T} = \partial_p[f(Q)\bar{n}(S)] - T\partial_p h_A > 0,$$

where the second term is strictly positive and the first is weakly positive because lower perceived attack risk weakly raises user continuation value and adoption.

It remains to sign the reverse effects. By the adoption closure (5), higher quality raises the voting population. Proposition 4 then implies that higher Q lowers pivotality and monitoring incentives. Lower monitoring weakens electoral discipline and raises the equilibrium attack exposure. Since attacks reset reputation from p to p_0, with $\bar{p} > p_0$, the induced effect on the reputation drift in (42) is negative: $J_{31} = \partial_Q \dot{p} < 0$.

Finally, Proposition 6 gives $\partial_T \Delta_I(S) > 0$, so a larger treasury raises the equilibrium attack probability and hence the attack hazard. Again, because attacks reset reputation to $p_0 < \bar{p}$, the effect on the reputation drift is negative: $J_{32} = \partial_T \dot{p} < 0$. Therefore, $J_{13} > 0, J_{23} > 0, J_{31} < 0, J_{32} < 0$, and consequently $J_{13}J_{31} < 0$ and $J_{23}J_{32} < 0$. □

A.6 Extensions

Proof of Proposition 9

Proof. The malicious delegate's continuation value can be decomposed into a wage component and a token component:

$$V_I^M(S; X_{Delegate}) = PV_\omega + PV_x(S; X_{Delegate}) + V_I^{M,\text{office}}(S), \tag{6}$$

where

$$PV_\omega = \frac{\omega}{r+\delta}, \tag{7}$$

and

$$PV_x(S; X_{Delegate}) = \mathbb{E}_t\left[\int_0^\infty e^{-(r+\delta)s}\,\delta\,x_D\,P(S_{t+s})\,ds\right]. \tag{8}$$

The wage component is a perpetuity-like flow claim. By contrast, the token component is a long-duration asset because its value depends on the entire future path of the platform through $P(S_{t+s})$, which in turn depends on $(Q_{t+s}, T_{t+s}, X_{t+s})$.

As r falls, the delegate discounts distant future states less. Therefore the sensitivity of PV_x to a decline in r is larger than that of PV_ω:

$$\left|\frac{\partial PV_x}{\partial r}\frac{r}{PV_x}\right| > \left|\frac{\partial PV_\omega}{\partial r}\frac{r}{PV_\omega}\right|. \tag{9}$$

Hence, for sufficiently small r, a marginal increase in x_D raises the malicious delegate's continuation value by more per unit of expected cost than a marginal increase in ω.

Because the principal minimizes contract cost subject to (**??**), the cost-minimizing solution places

weight on the compensation instrument with the highest marginal effect on continuation value per unit of cost. For $r < \bar{r}$, that instrument is token compensation. Therefore the optimal incentive-compatible contract is dominated by token-based pay. □

Proof of Proposition 10

Proof. As defined the type-k delegate marginal investment payoff is

$$F^k(I;S) = Q\mu_I(e_k^*(S), I)\,\partial_Q V_I^k(S) - \partial_T V_I^k(S), \qquad k \in \{H, M\}.$$

The honest delegate satisfies the first-order condition

$$F^H(I_H^*(S);S) = 0.$$

Hence, evaluating the malicious first-order condition at the honest optimum gives

$$G(S) \equiv F^M(I_H^*(S);S) = Q\mu_I(e_M^*(S), I_H^*(S))\,\partial_Q V_I^M(S) - \partial_T V_I^M(S)$$

Because $\partial_I F^M(I;S) < 0$, the sign of $G(S)$ determines the ranking of optimal investment:

$$G(S) > 0 \implies I_M^*(S) > I_H^*(S), \qquad G(S) < 0 \implies I_M^*(S) < I_H^*(S).$$

Proof that $G > 0$ when $T \to 0$.

Define the malicious surplus over the honest type:

$$\Omega(S) \equiv V_I^M(S) - V_I^H(S).$$

Then

$$G(S) = Q\mu_I(e_M^*(S), I_H^*(S))\,\partial_Q\Omega(S) - \partial_T\Omega(S) \tag{10}$$

$$+ Q\Big(\mu_I(e_M^*(S), I_H^*(S)) - \mu_I(e_H^*(S), I_H^*(S))\Big)\partial_Q V_I^H(S). \tag{11}$$

By Assumption (A3), for T sufficiently close to 0, immediate attack is strictly suboptimal. Hence in a neighbourhood of the viability boundary the Bellman obstacle $\Pi(S) + V_C^M(S^-)$ is inactive and the

malicious and honest incumbents solve smooth continuation HJBs.

At $T = 0$, feasibility imposes $I \in [0, T] = \{0\}$, so

$$I_H^*(Q, 0, p, \tilde{p}) = I_M^*(Q, 0, p, \tilde{p}) = 0.$$

Moreover, $\Pi(0) = 0$, so the direct one-shot gain from expropriation vanishes at the boundary. Therefore the difference between the two types at $T = 0$ is purely a first-order continuation effect.

We now claim that there exists a boundary coefficient $\omega(Q, p, \tilde{p})$ such that

$$\Omega(Q, T, p, \tilde{p}) = T\,\omega(Q, p, \tilde{p}) + o(T) \qquad \text{as } T \downarrow 0,$$

with convergence in C^1_{loc} in the Q-variable. To derive the equation for ω, freeze the controlled diffusion at the boundary policy $(e_0, 0)$, where

$$e_0(Q, p, \tilde{p}) := e_H^*(Q, 0, p, \tilde{p}) = e_M^*(Q, 0, p, \tilde{p}).$$

Let $\mathcal{A}_0$ denote the linear generator of the boundary continuation process:

$$\mathcal{A}_0\phi = Q\big(\mu(e_0, 0) - \kappa\big)\partial_Q\phi + \frac{1}{2}\sigma_Q^2 Q^2 \partial_{QQ}\phi + \lambda\Big(\mathbb{E}_0[\phi(\widehat{S}')] - \phi(\widehat{S})\Big),$$

where $\widehat{S} = (Q, p, \tilde{p})$ and $\mathbb{E}_0[\cdot]$ is the post-governance transition operator evaluated at $T = 0$ and $I = 0$. The T-coordinate drops out because we are taking the first-order coefficient in the expansion with respect to T.

Subtract the honest and malicious HJBs, divide by T, and pass to the limit $T \downarrow 0$. Since the obstacle is inactive in a local no-attack neighbourhood, the zero-order terms cancel and the first-order coefficient ω solves the linear resolvent equation

$$(r - \mathcal{A}_0)\omega(\widehat{S}) = \mathfrak{s}(\widehat{S}), \qquad \widehat{S} = (Q, p, \tilde{p}), \tag{12}$$

where the source term is

$$\mathfrak{s}(\widehat{S}) = \lambda(1 - \chi)\pi_I(\widehat{S}, \rho^*)\,\partial_T\Psi(Q, 0, p, \tilde{p}). \tag{13}$$

The quantity $\Psi(S) = \Pi(T) + V_C^M(S^-) - V_I^M(S^+)$ is the malicious net attack surplus. Thus $\mathfrak{s}$ is precisely

the marginal continuation value, at the viability boundary, of one extra unit of treasury.

By the Feynman–Kac formula, (12) admits the representation

$$\omega(\widehat{S}) = \mathbb{E}_{\widehat{S}}\left[\int_0^\infty e^{-rs}\,\mathfrak{s}(\widehat{S}_s)\,ds\right]. \tag{14}$$

Since $\mathfrak{s} \geq 0$, it follows that $\omega \geq 0$.

We next prove that the coefficient of the first-order expansion of G is strictly positive.

Differentiate the boundary expansion in Q:

$$\partial_Q \Omega(Q, T, p, \tilde{p}) = T\,\partial_Q \omega(Q, p, \tilde{p}) + o(T).$$

Because ω has no T-argument, $\partial_T \Omega(Q, T, p, \tilde{p}) = \omega(Q, p, \tilde{p}) + o(1)$. However, at the viability boundary the source term $\mathfrak{s}$ captures only the first-order continuation gain from treasury. By the local no-attack structure, $\omega(Q, p, \tilde{p})$ enters at the same order as the difference in continuation values, and the term $-\partial_T \Omega$ is cancelled at the boundary by the honest first-order condition and the collapse of the feasible set to $I = 0$. Therefore the leading term of G is the quality channel:

$$G(Q, T, p, \tilde{p}) = Q\mu_I(e_0(Q, p, \tilde{p}), 0)\,\partial_Q \Omega(Q, T, p, \tilde{p}) + o(T). \tag{15}$$

Substituting the expansion of $\partial_Q \Omega$,

$$G(Q, T, p, \tilde{p}) = Q\mu_I(e_0, 0)\,\partial_Q \omega(Q, p, \tilde{p})\,T + o(T).$$

Define

$$a(Q, p, \tilde{p}) := Q\mu_I(e_0(Q, p, \tilde{p}), 0)\,\partial_Q \omega(Q, p, \tilde{p}). \tag{16}$$

Then

$$G(Q, T, p, \tilde{p}) = a(Q, p, \tilde{p})\,T + o(T). \tag{17}$$

It remains to show $a(Q, p, \tilde{p}) > 0$. Since $\mu_I > 0$, it is enough to prove

$$\partial_Q \omega(Q, p, \tilde{p}) > 0.$$

Differentiate (12) in Q:

$$(r - \mathcal{A}_0)\omega_Q = \partial_Q \mathfrak{s} + [\partial_Q, \mathcal{A}_0]\omega.$$

By Assumption (A4), quality raises adoption and treasury drift, and strengthens the future attack surplus:

$$\partial_Q \bar{n}(Q) > 0, \qquad \partial_Q \Psi(S) \geq 0, \qquad \partial_Q(\mathcal{L}^e \Psi(S)) > 0.$$

Hence $\partial_Q \mathfrak{s} \geq 0$, and strict positivity holds on a non-empty reachable set. Since the commutator term $[\partial_Q, \mathcal{A}_0]\omega$ is also nonnegative under the monotone drift and diffusion structure, the right-hand side is nonnegative and not identically zero. The strong maximum principle applied to the linear elliptic-parabolic resolvent equation for ω_Q therefore yields

$$\omega_Q(Q, p, \tilde{p}) > 0.$$

Consequently, $a(Q, p, \tilde{p}) > 0$. By (17), there exists $\varepsilon(Q, p, \tilde{p}) > 0$ such that

$$G(Q, T, p, \tilde{p}) > 0 \qquad \text{for all } 0 < T < \varepsilon(Q, p, \tilde{p}).$$

This proves the small-T sign.

Proof that $G < 0$ for sufficiently large T.

We now show that the liquidity-hoarding channel dominates when treasury is large.

Recall

$$G(S) = Q\mu_I(e_M^*(S), I_H^*(S))\, \partial_Q V_I^M(S) - \partial_T V_I^M(S).$$

The first term is the indirect quality-growth motive. By Assumption (A6), it is bounded above:

$$Q\mu_I(e_M^*(S), I_H^*(S))\, \partial_Q V_I^M(S) \leq C(Q, p, \tilde{p})$$

for some finite $C(Q, p, \tilde{p})$.

The second term is the marginal private value of one more unit of liquid treasury to the malicious incumbent. Since $\Pi(T) = T$, one additional unit of treasury increases the attack payoff one-for-one.

Moreover, when T is sufficiently large, immediate expropriation becomes strictly attractive, so the equilibrium attack policy satisfies $q^*(S) = 1$ on a large-T region. On that region the malicious HJB implies

$$\partial_T V_I^M(S) \geq \lambda(1-\chi)\pi_I(S,\rho^*)\,\partial_T \Pi(T) = \lambda(1-\chi)\pi_I(S,\rho^*).$$

Because the attack region is entered for sufficiently large treasury, there exists $\bar{T}(Q,p,\tilde{p})$ and $\underline{c}(Q,p,\tilde{p}) > 0$ such that for all $T \geq \bar{T}(Q,p,\tilde{p})$,

$$\partial_T V_I^M(Q,T,p,\tilde{p}) \geq \underline{c}(Q,p,\tilde{p}).$$

As treasury grows, the direct value of preserving liquid funds for expropriation rises one-for-one, while the indirect quality-growth term remains bounded. Therefore, enlarging $\bar{T}$ if necessary, we obtain

$$C(Q,p,\tilde{p}) - \partial_T V_I^M(Q,T,p,\tilde{p}) < 0 \qquad \text{for all } T \geq \bar{T}(Q,p,\tilde{p}),$$

hence

$$G(Q,T,p,\tilde{p}) < 0 \qquad \text{for all } T \geq \bar{T}(Q,p,\tilde{p}).$$

This proves the large-T sign.

Proof that $\partial_T G < 0$.

We now prove monotonicity, and this is a result rather than an assumption.

Starting from (11), focus on the option-value component

$$\widetilde{G}(S) := Q\mu_I(e_M^*(S), I_H^*(S))\,\partial_Q \Omega(S) - \partial_T \Omega(S).$$

The remaining complementarity term

$$Q\Big(\mu_I(e_M^*, I_H^*) - \mu_I(e_H^*, I_H^*)\Big)\partial_Q V_I^H$$

is weakly positive under $\mu_{eI} \geq 0$ and $e_M^* \geq e_H^*$, and does not overturn the sign of the T-derivative.

Therefore it suffices to show

$$\partial_T \widetilde{G}(S) < 0.$$

Differentiating,

$$\partial_T \widetilde{G}(S) = Q\mu_I(e_M^*, I_H^*)\, \partial_{QT}\Omega(S) - \partial_{TT}\Omega(S) + \mathcal{R}(S), \tag{18}$$

where $\mathcal{R}(S)$ collects derivatives of e_M^* and I_H^*. By continuity of optimal controls and strict concavity, $\mathcal{R}(S)$ is bounded on the relevant domain.

We next compare Ω_{QT} and Ω_{TT}.

Claim 1: Ω is convex in T, and strictly so on attack-relevant states.

The malicious Bellman operator is the supremum of affine maps in the treasury variable T. The treasury law of motion is affine in T, the attack payoff is linear in T, and the attack decision is an optimal stopping/switching option. Suprema of affine functions are convex. Therefore

$$\partial_{TT}\Omega(S) \geq 0,$$

with strict inequality wherever the attack option is locally relevant. This is the standard real-options convexity effect: the option to seize treasury makes the malicious surplus convex in the underlying lootable stock.

Claim 2: Ω has increasing differences in (Q, T), but the direct treasury effect dominates the cross-effect.

Since $\partial_Q \bar{n}(Q) > 0$, higher quality raises future treasury drift. Since $\partial_Q \Psi(S) \geq 0$, higher quality also strengthens the future attack surplus. Hence Ω has increasing differences:

$$\partial_{QT}\Omega(S) \geq 0.$$

However, T enters the attack payoff directly and one-for-one,

$$\partial_T \Pi(T) = 1,$$

whereas quality affects private extraction only indirectly through future growth of treasury:

$$Q \longrightarrow \bar{n}(Q) \longrightarrow f(\bar{n}(Q)) \longrightarrow T_{t+s} \longrightarrow \Pi(T_{t+s}).$$

The latter channel is delayed, discounted, and dampened by concavity of μ.

To formalize this, let $m := \Omega_T$ and $n := \Omega_Q$. In the continuation region, differentiating the variational inequality yields linear resolvent equations

$$(r - \mathcal{A}_M)m = \lambda(1-\chi)\pi_I \cdot 1 + B_m(S), \tag{19}$$

and

$$(r - \mathcal{A}_M)n = \lambda(1-\chi)\pi_I \, \partial_Q \Psi + B_n(S), \tag{20}$$

where $\mathcal{A}_M$ is the malicious continuation generator and the remainder terms B_m, B_n are nonnegative continuation corrections. The key difference is that (19) contains the direct unit source term generated by $\Pi_T = 1$, while (20) contains only the indirect quality source.

Differentiate (20) once more in T. Then $\Omega_{QT} = n_T$ solves a linear equation whose source remains bounded by the discounted sensitivity of the treasury-drift channel to quality. By contrast, $\Omega_{TT} = m_T$ solves a linear equation whose source inherits the direct one-for-one treasury effect. By comparison of resolvents,

$$\partial_{TT}\Omega(S) > Q\mu_I(e_M^*, I_H^*)\, \partial_{QT}\Omega(S) + \mathcal{R}(S)$$

throughout the continuation region. Substituting into (18) yields

$$\partial_T \widetilde{G}(S) < 0.$$

Therefore

$$\partial_T G(S) < 0.$$

Proof of the existence and uniqueness of the threshold.

$G(Q, T, p, \tilde{p}) > 0$ for all T sufficiently close to 0 and $G(Q, T, p, \tilde{p}) < 0$ for all T sufficiently large. By continuity of G, there exists at least one threshold $T^\dagger(Q, p, \tilde{p})$ such that

$$G(Q, T^\dagger, p, \tilde{p}) = 0.$$

As G is strictly decreasing in T. Therefore this zero is unique.

Ranking of investment schedules.

By strict concavity of the malicious investment problem,

$$G(S) > 0 \iff I_M^*(S) > I_H^*(S), \qquad G(S) < 0 \iff I_M^*(S) < I_H^*(S).$$

Combining this equivalence with the unique sign change of G at $T^\dagger(Q, p, \tilde{p})$, we obtain

$$I_M^*(S) > I_H^*(S) \quad \iff \quad T < T^\dagger(Q, p, \tilde{p}),$$

and

$$I_M^*(S) < I_H^*(S) \quad \iff \quad T > T^\dagger(Q, p, \tilde{p}).$$

□

B Simulation Procedure

Implementation of the simulation

This appendix documents the numerical implementation of the model and the simulation procedures.

State variables, timing, and discretization

Time is discretized on a uniform grid with step size Δt, over a finite horizon $[0, \mathcal{T}]$. At each grid point $t_k = k\Delta t$, the economy is summarized by the state vector

$$S_t \equiv (Q_k, T_k, X_k, p_{i,k}, p_{j,k}),$$

where Q_k denotes protocol quality, T_k denotes the treasury balance, X_k denotes total token supply, and $p_{i,k}$ and $p_{j,k}$ denote the posterior probabilities that the incumbent delegate i and challenger delegate j are honest types, respectively.

The incumbent and challenger hold fixed token stakes x_i and x_j. The outside tokenholder stake is the residual supply

$$x_u(S_k) = \max\{X_k - x_i - x_j, 0\}$$

which clears the token market at each date and enters all pricing and utility calculations.

Law of motion for continuous state variables

Quality process

Quality evolves as an effort-driven diffusion with an attack-induced downward jump. The notebook implements the quality drift as

$$\mu_Q(e) = \mu_{0,Q} + \mu_{1,Q} e^{\alpha_Q} - \kappa_Q$$

where $\mu_{0,Q}$ is the baseline quality drift, $\mu_{1,Q} > 0$ scales the effectiveness of delegate effort, $\alpha_Q > 0$ controls the curvature of effort in quality production, and κ_Q is the depreciation rate of quality. The derivative of the quality drift with respect to effort is

$$\mu_Q'(e) = \mu_{1,Q} \alpha_Q e^{\alpha_Q - 1}.$$

When $e \leq 0$, the implementation evaluates this derivative at a small positive numerical safeguard $\varepsilon_e = 10^{-8}$, so that

$$\mu_Q'(0) \approx \mu_{1,Q} \alpha_Q \varepsilon_e^{\alpha_Q - 1}.$$

Let

$$\varepsilon_{Q,k} \sim \mathcal{N}(0,1).$$

Before any attack jump, quality is updated using the Euler discretization

$$\widetilde{Q}_{k+1} = \left[Q_k + \mu_Q(e_k^\star)\Delta t + \sigma_Q \sqrt{\Delta t}\, \varepsilon_{Q,k} \right]_+,$$

where $[z]_+ = \max\{z, 0\}$. Let

$$\mathcal{A}_k = \mathbf{1}\{\text{attack occurs at date } t_k\}.$$

If an attack occurs, quality is reduced proportionally by γ_Q. The implemented law of motion is therefore

$$Q_{k+1} = (1 - \gamma_Q \mathcal{A}_k)\, \widetilde{Q}_{k+1}.$$

Equivalently,

$$Q_{k+1} = (1 - \gamma_Q \mathcal{A}_k) \left[Q_k + \mu_Q(e_k^\star)\Delta t + \sigma_Q \sqrt{\Delta t}\, \varepsilon_{Q,k} \right]_+ .$$

Treasury

Treasury inflows are generated by protocol quality through the external inflow function

$$g(Q) = g_0 + g_1 Q^{\gamma_T}$$

as we focus on a setting in which the fees inflow is paid to the delegate. The effective participant base is $\bar{n}(Q) = \min\left\{ \frac{a_0 Q^\varepsilon}{\delta_n}, 50 \right\}$. The cap at 50 is a numerical safeguard that prevents the voting fixed point from becoming unstable at high values of quality. The treasury inflow drift used in the simulation is

$$\mu_T(Q) = g(Q)\bar{n}(Q).$$

Before any attack jump, the treasury evolves according to

$$\widetilde{T}_{k+1} = \left[T_k + g(Q_k)\bar{n}(Q_k)\Delta t \right]_+ .$$

If an attack occurs, the code subtracts the pre-event treasury T_k. Hence, with

$$\mathcal{A}_k = \mathbf{1}\{\text{attack occurs at date } t_k\},$$

the implemented treasury law of motion is

$$T_{k+1} = \left[\widetilde{T}_{k+1} - T_k \mathcal{A}_k \right]_+ .$$

Substituting the pre-attack update gives

$$T_{k+1} = \left[T_k + g(Q_k)\bar{n}(Q_k)\Delta t - T_k \mathcal{A}_k\right]_+ .$$

Token supply

Governance opportunities arrive according to a Poisson clock with intensity λ_{gov}. Let

$$G_k = \mathbf{1}\{\text{governance event occurs at date } t_k\}.$$

For small Δt,

$$\Pr(G_k = 1) \approx \lambda_{\text{gov}} \Delta t.$$

In the extension in which the leading delegate is paid via a fixed token grant, token grants occur only when governance events occur.

The token supply law of motion is

$$X_{k+1} = X_k + x_D G_k.$$

Belief updating

When a governance event occurs and no attack is observed, beliefs about the incumbent delegate are updated using Bayes' rule. Let $p^-_{i,k}$ denote the prior probability that the incumbent is honest immediately before the governance event. If a malicious incumbent attacks with probability $q_i^\star(S_k)$, then observing no attack leads to the posterior

$$p^+_{i,k} = \frac{p^-_{i,k}}{p^-_{i,k} + (1 - p^-_{i,k})(1 - q_i^\star(S_k))}.$$

If an attack occurs, the incumbent is replaced and the new incumbent reputation is reset to the prior reputation:

$$p_{i,k+1} = p_{0,\text{rep}}.$$

Attack probability and governance attacks

At each governance event, a malicious delegate may attack. For delegate $\ell \in \{i, j\}$, the reduced-form attack probability is logistic:

$$q_\ell^\star(S_k, n_I) = \frac{1}{1 + \exp\{-\Xi_\ell(S_k, n_I)\}},$$

where

$$\Xi_\ell(S_k, n_I) = 2\frac{T_k}{1 + T_k} + 2(1 - p_{\ell,k}) - \frac{n_I}{\bar{n}(Q_k)}.$$

The first term captures the effect of treasury size on attack incentives. The second term captures reputational discipline: lower perceived honesty raises attack incentives. The final term captures oversight: a larger share of informed tokenholders reduces the attack probability. For a malicious incumbent, an attack can occur only during a governance event. The implemented attack condition is $G_k = 1$, $T_k > 0.8T_0$, and $V_{\text{attack}}(T_k) > W_{\text{in}}(T_k)$. The notebook sets $W_{\text{in}}(T_k) = T_k$ and $V_{\text{attack}}(T_k) = T_k + W_{\text{out,reset}} - c_I$. Conditional on these threshold conditions, the attack is drawn with probability $q_i^\star(S_k, n_I^\star(S_k))$. Thus the attack indicator can be written as

$$\mathcal{A}_k = G_k \cdot \mathbf{1}\{\text{incumbent is malicious}\} \cdot \mathbf{1}\{T_k > 0.8T_0\} \cdot \mathbf{1}\{V_{\text{attack}}(T_k) > T_k\} \cdot B_k,$$

where

$$B_k \sim \text{Bernoulli}\left(q_i^\star(S_k, n_I^\star(S_k))\right).$$

Voting subgame and information acquisition

At each date, tokenholders decide whether to acquire information before voting. The simulation solves for the equilibrium expected number of informed tokenholders,

$$n_I^\star(S_k) \in [0, \bar{n}(Q_k)].$$

The implied information-acquisition probability is

$$\pi_I^\star(S_k) = \frac{n_I^\star(S_k)}{\bar{n}(Q_k)} \quad \text{if } \bar{n}(Q_k) > 0,$$

and

$$\pi_I^\star(S_k) = 0 \quad \text{if } \bar{n}(Q_k) = 0.$$

T

he voting subgame is approximated using Poisson vote arrivals. Let n_I be a candidate number of informed tokenholders. The number of uninformed tokenholders is $n_H(Q_k, n_I) = \bar{n}(Q_k) - n_I$ The implementation draws an informed-vote signal share $\theta_k \sim \mathrm{Beta}(3, 2)$ whose mean is 0.6. Informed tokenholders vote according to θ_k, while uninformed tokenholders split symmetrically. We choose to have not perfectly informed voters which paid the information acquisition cost for illustrative purposes as if every informed tokenholder votes perfectly on the same side, then once nI is even moderately large, the informed bloc becomes mechanically one-sided. The vote margin becomes too deterministic, and the pivotal probability often collapses toward zero. Let ξ_u denote the effective voting weight per outside tokenholder. The expected yes and no vote intensities are

$$\mu_Y(S_k, n_I) = \left[\theta_k n_I + \frac{1}{2} n_H(Q_k, n_I)\right] \xi_u + x_i \xi_u + \frac{1}{2} x_u(S_k) \xi_u,$$
$$\mu_N(S_k, n_I) = \left[(1 - \theta_k) n_I + \frac{1}{2} n_H(Q_k, n_I)\right] \xi_u + x_j \xi_u + \frac{1}{2} x_u(S_k) \xi_u.$$

Let the net vote margin be

$$D_k(n_I) = N_Y - N_N,$$

where

$$N_Y \sim \mathrm{Poisson}(\mu_Y(S_k, n_I)), \qquad N_N \sim \mathrm{Poisson}(\mu_N(S_k, n_I)).$$

Then

$$D_k(n_I) \sim \mathrm{Skellam}\left(\mu_Y(S_k, n_I), \mu_N(S_k, n_I)\right).$$

The pivotal probability is computed from the Skellam mass function as

$$P_{\text{piv}}(S_k, n_I) = \sum_{m=-1}^{-\lfloor \min\{x_u(S_k),20\} \rfloor} \Pr\left(D_k(n_I) = m\right).$$

The cap at 20 is again numerical and keeps the Skellam summation tractable.

The marginal benefit of information is

$$\text{MBI}(S_k, n_I) = P_{\text{piv}}(S_k, n_I) \left[(1 - p_{i,k} q_i^\star(S_k, n_I))\Delta U_{\text{vote}} - (1 - p_{j,k} q_j^\star(S_k, n_I))\right].$$

The equilibrium number of informed tokenholders solves

$$\text{MBI}(S_k, n_I^\star) = c_A,$$

subject to

$$0 \leq n_I^\star(S_k) \leq \bar{n}(Q_k).$$

Equivalently, define the fixed-point residual

$$F(n_I; S_k) = \text{MBI}(S_k, n_I) - c_A.$$

Then $n_I^\star(S_k)$ solves

$$F(n_I^\star; S_k) = 0.$$

Delegate effort choice

The leading delegate chooses effort by equating marginal effort cost to the marginal value of quality. Effort costs are quadratic:

$$c(e) = \frac{1}{2} c_2 e^2,$$

so that

$$c'(e) = c_2 e.$$

For a delegate of type $\tau \in \{H, M\}$, where H denotes honest and M denotes malicious, effort solves

$$c'(e_\tau^\star) = \mu_Q'(e_\tau^\star)\frac{\partial W_\tau(S_k)}{\partial Q}.$$

Substituting the functional forms gives

$$c_2 e_\tau^\star = \mu_{1,Q}\alpha_Q(e_\tau^\star)^{\alpha_Q - 1}\frac{\partial W_\tau(S_k)}{\partial Q}.$$

The solution is constrained to the interval

$$e_\tau^\star(S_k) \in [e_{\min}, e_{\max}].$$

If

$$\frac{\partial W_\tau(S_k)}{\partial Q} \leq 0,$$

the implementation sets

$$e_\tau^\star(S_k) = e_{\min}.$$

Otherwise, the scalar first-order condition is solved numerically by bisection over $[e_{\min}, e_{\max}]$. If no crossing occurs within the interval, the routine returns the relevant boundary value.

Quality derivatives for honest and malicious delegates

The simulation computes effort using local approximations to the quality derivatives $\frac{\partial W_H(S_k)}{\partial Q}$ and $\frac{\partial W_M(S_k)}{\partial Q}$. These derivatives translate a marginal increase in quality into changes in treasury accumulation, token price, and the malicious delegate's option value of future attack.

First, the effective participant base is $\bar{n}(Q) = \frac{a_0 Q^\varepsilon}{\delta_n}$. Its derivative is

$$\bar{n}_Q'(Q) = \frac{a_0 \varepsilon Q^{\varepsilon - 1}}{\delta_n}$$

The external treasury inflow per participant is $g(Q) = g_0 + g_1 Q^{\gamma_T}$. Therefore, $g_Q'(Q) = g_1 \gamma_T Q^{\gamma_T - 1}$. The treasury inflow drift is $\mu_T(Q) = g(Q)\bar{n}(Q)$ and its derivative is

$$\frac{\partial \mu_T(Q)}{\partial Q} = g_Q'(Q)\bar{n}(Q) + g(Q)\bar{n}_Q'(Q).$$

For the honest delegate, we approximate the effect of quality on treasury by

$$\frac{\partial T}{\partial Q} \approx \frac{\partial \mu_T(Q)/\partial Q}{r_T + \delta}.$$

The local price derivative is approximated by

$$\frac{\partial P}{\partial Q} \approx \frac{\Phi_Q}{r} + \frac{\Phi_P}{rX}\frac{\partial T}{\partial Q},$$

where Φ_Q and Φ_P are numerical scaling constants. The honest delegate's marginal value of quality is then

$$\frac{\partial W_H(S_k)}{\partial Q} = x_D \left(\frac{\delta}{r + \delta + \lambda_{\text{gov}}}\right)\frac{\partial P}{\partial Q}.$$

The implementation applies the non-negativity safeguard

$$\frac{\partial W_H(S_k)}{\partial Q} \leftarrow \max\left\{\frac{\partial W_H(S_k)}{\partial Q}, 0\right\}.$$

For a malicious delegate, we add an option-value derivative:

$$\frac{\partial W_M(S_k)}{\partial Q} = \frac{\partial W_H(S_k)}{\partial Q} + \frac{\partial V_{\text{opt}}(S_k)}{\partial Q}.$$

The attack hazard is approximated by

$$h(S_k) = \lambda_{\text{gov}} q_i^\star(S_k, n_I^\star(S_k)).$$

The expected time to attack is

$$\mathbb{E}_k[\tau] \approx \frac{1}{h(S_k) + \varepsilon_h},$$

where $\varepsilon_h > 0$ is a numerical safeguard. The marginal option value of quality is approximated as

$$\frac{\partial V_{\text{opt}}(S_k)}{\partial Q} \approx \exp\left(-r\mathbb{E}_k[\tau]\right)\frac{\partial \mu_T(Q_k)}{\partial Q}\mathbb{E}_k[\tau].$$

Thus,

$$\frac{\partial W_M(S_k)}{\partial Q} = \frac{\partial W_H(S_k)}{\partial Q} + \exp\left(-r\mathbb{E}_k[\tau]\right)\frac{\partial \mu_T(Q_k)}{\partial Q}\mathbb{E}_k[\tau].$$

The malicious delegate's effort is obtained by substituting $\partial W_M(S_k)/\partial Q$ into the same effort first-order

condition used for the honest delegate.

Delegate compensation

We define the fee inflow function as

$$f(\bar{n}(Q)) = f_0 \bar{n}(Q)^{\beta_f}.$$

Delegate compensation is set equal to this fee inflow:

$$\omega(Q) = f(\bar{n}(Q)).$$

The derivative used in value calculations is

$$\omega'_Q(Q) = f_0 \beta_f \bar{n}(Q)^{\beta_f - 1} \bar{n}'_Q(Q).$$

Although $f(\bar{n}(Q))$ and $\omega(Q)$ are important for delegate payoffs and effort incentives, they do not enter the current realized treasury law of motion. The simulated treasury inflow is governed by $g(Q)\bar{n}(Q)$ not by $f(\bar{n}(Q)) - \omega(Q)$.

Tokenholder utility

Given a simulated path, tokenholder utility is computed as a discounted payoff until an exit or attack time. Let τ_δ denote the pathwise exit index. If no attack occurs, the terminal simulation date is used as the exit date. If an attack occurs, the first attack date is used as the exit date. The liquidation value per token is $\frac{T_{\tau_\delta}}{X_{\tau_\delta}}$. For a tokenholder with x tokens, simulated utility is

$$U(S_k, x) = \exp\{-r(t_{\tau_\delta} - t_k)\} \frac{T_{\tau_\delta}}{X_{\tau_\delta}} x - \sum_{\ell=k+1}^{\tau_\delta} \exp\{-r(t_\ell - t_k)\} \lambda_{\text{gov}} \Delta t_\ell \pi_I^\star(S_\ell) c_A$$

The second term is the expected discounted information-acquisition cost along the realized path, where

$$\pi_I^\star(S_\ell) = \frac{n_I^\star(S_\ell)}{\bar{n}(Q_\ell)}.$$

Least-squares Monte Carlo computation of the token price

We compute the token price path using a least-squares Monte Carlo approximation of the continuation value. Rather than recomputing a full nested Monte Carlo continuation value at every date of the realized path, we first simulate a collection of training paths and compute, along each path, a fast realized continuation value of the governance component of the token price. Specifically, for each simulated path m, the code computes a marginal governance-value flow

$$\psi_{m,k} = \lambda_{\text{gov}} \rho_{m,k} \frac{\text{MBI}_{m,k}}{\max\{x_{u,m,k}, \varepsilon_x\}}$$

where λ_{gov} is the governance-arrival intensity, $\rho_{m,k}$ is the probability that an honest delegate is active in governance at state $S_{m,k}$, $\text{MBI}_{m,k}$ is the marginal benefit of information computed by the voting block of the simulation, and $x_{u,m,k}$ is the outside token supply over which this information benefit is averaged. This corresponds to the "average MBI" approximation used in the implementation.

For each training path, the continuation value target is then computed as the discounted forward integral

$$Y_{m,k} = \sum_{\ell \geq k:\, t_\ell \leq t_k + H} e^{-r(t_\ell - t_k)} \psi_{m,\ell} \Delta t_\ell$$

where H is the pricing horizon. We then estimate a regression of $Y_{m,k}$ on a vector of basis functions of the state,

$$Y_{m,k} = \Phi(S_{m,k})' \beta + \varepsilon_{m,k}$$

using a small ridge penalty for numerical stability. This gives an estimated continuation-value function

$$\widehat{V}(S) = \Phi(S)' \widehat{\beta}$$

Given the realized path displayed in the figures, we evaluate this fitted continuation-value function date by date:

$$P_k^{\text{gov}} = \widehat{V}(S_k) = \Phi(S_k)' \widehat{\beta}$$

Thus the plotted series is not obtained by repeatedly launching a new Monte Carlo simulation from

each date. Instead, the costly Monte Carlo step is performed once and the price path is then obtained by evaluating the fitted pricing function along the realized simulated states. Attacks affect the predicted price through their impact on the state vector S_k: an attack can reduce treasury value T_k, quality Q_k, reputations, and the future marginal benefit of information, all of which enter the basis functions $\Phi(S_k)$ and therefore the predicted continuation value.

Simulation algorithm

The full simulation algorithm is as follows.

1. Initialize $S_0 = (Q_0, T_0, X_0, p_{i,0}, p_{j,0})$.
2. For each grid date t_k:
   (a) Compute the outside tokenholder stake: $x_u(S_k) = \max\{X_k - x_i - x_j, 0\}$.
   (b) Solve the voting and information-acquisition fixed point: $n_I^\star(S_k) \in [0, \bar{n}(Q_k)]$
   (c) Compute the equilibrium information-acquisition probability: $\pi_I^\star(S_k) = \frac{n_I^\star(S_k)}{\bar{n}(Q_k)}$.
   (d) Compute the pivotal probability $P_{\text{piv}}(S_k, n_I^\star)$ and marginal benefit of information $\text{MBI}(S_k, n_I^\star)$.
   (e) Compute the attack probabilities $q_i^\star(S_k, n_I^\star)$ and $q_j^\star(S_k, n_I^\star)$.
   (f) Compute either $\frac{\partial W_H(S_k)}{\partial Q}$ or $\frac{\partial W_M(S_k)}{\partial Q}$
   depending on whether the leading delegate is honest or malicious.
   (g) Solve the delegate effort first-order condition: $c_2 e_k^\star = \mu_{1,Q}\alpha_Q (e_k^\star)^{\alpha_Q - 1} \frac{\partial W_\tau(S_k)}{\partial Q}$
   (h) Compute the pre-attack quality update: $\widetilde{Q}_{k+1} = \left[Q_k + \mu_Q(e_k^\star)\Delta t + \sigma_Q\sqrt{\Delta t}\varepsilon_{Q,k}\right]_+$
   (i) Compute the pre-attack treasury update: $\widetilde{T}_{k+1} = \left[T_k + g(Q_k)\bar{n}(Q_k)\Delta t\right]_+$
   (j) Draw the governance event: $G_k \sim \text{Bernoulli}\left(\lambda_{\text{gov}}\Delta t\right)$
   (k) Update token supply using the same governance event: $X_{k+1} = X_k + x_D G_k$
   (l) If $G_k = 1$, evaluate the attack condition. If an attack occurs, set $Q_{k+1} = (1 - \gamma_Q)\widetilde{Q}_{k+1}$ and $T_{k+1} = \left[\widetilde{T}_{k+1} - T_k\right]_+$
   (m) If no attack occurs, set $Q_{k+1} = \widetilde{Q}_{k+1}$, $T_{k+1} = \widetilde{T}_{k+1}$

(n) If a governance event occurs and no attack is observed, update beliefs using Bayes' rule: $p^{+}_{i,k} = \frac{p^{-}_{i,k}}{p^{-}_{i,k}+(1-p^{-}_{i,k})(1-q^{\star}_{i}(S_k))}$

(o) If an attack occurs, reset the incumbent reputation: $p_{i,k+1} = p_{0,\mathrm{rep}}$

3. After simulating the full state path, compute the token price by the backward dynamic-programming recursion:

$$P_k = e^{-r\Delta t_k}\left[(1-\delta_{\mathrm{liq}}\Delta t_k)\,P_{k+1} + \delta_{\mathrm{liq}}\Delta t_k \frac{T_k}{X_k}\right]$$

Model parameters

We calibrate key model parameters using empirically plausible orders of magnitude drawn from observed DAO governance activity and market characteristics. First, we calibrate the governance arrival rate λ_{gov} to 1.5 to generate approximately 18 governance proposals per year (if one period is considered to be a month), closely matching the average of 17.6 proposals observed in our sample. This choice ensures that the frequency of governance opportunities in the model aligns with the level of proposal activity faced by tokenholders in practice. Second, delegate compensation is calibrated to reflect the fact that, in most DAOs, remuneration is overwhelmingly paid in governance tokens rather than in dollar-denominated assets. Accordingly, we set token-based compensation to be approximately 100 times larger than dollar-denominated compensation. Third, we calibrate volatility parameters to reflect differences in asset composition. Governance token volatility is set to be roughly twice that of the treasury, consistent with the observation that DAO treasuries hold a mix of relatively stable assets, such as stablecoins, alongside more volatile assets, including Ether and Bitcoin. This asymmetry ensures that attacks and governance events disproportionately affect token prices relative to treasury values. We discipline our calibration using observed growth dynamics in the DeFi sector over the past two years. Aggregate DeFi total value locked (TVL) grew from roughly \$55 billion to \$125 billion, corresponding to an average annual growth rate of approximately 51%. Over the same period, Aave experienced average annual growth of roughly 182% in TVL. Netting out aggregate DeFi market growth implies an idiosyncratic annual growth rate of approximately 87 percent, or 5.35% per month, using the standard transformation

$g_m = (1 + g_a)^{1/12} - 1$. Similarly, Uniswap exhibited annual growth of approximately 156% , after removing the 51% market component, this corresponds to an idiosyncratic annual growth rate of roughly 70%, or 4.5% per month. Thus, across the two most prominent DeFi DAOs, net-of-market monthly growth rates consistently fall in the 4.5–5.5 percent range. Finally, we set the baseline productivity drift to zero, so that productivity growth in the simulations should be interpreted net of aggregate market growth. Under this normalization, all productivity dynamics arise endogenously from delegate effort and governance incentives rather than from exogenous trends in the broader crypto or DeFi market. Motivated by the empirical magnitudes above, we calibrate the idiosyncratic monthly growth rate to 5 percent, which closely matches the observed net-of-market growth of Aave and Uniswap and provides a conservative benchmark for protocol-specific productivity dynamics.

Table A.1. Updated parameters used in the simulation notebook.

Symbol	**Notebook name**	**Description**	**Value**
r	`r`	Discount rate.	0.05
δ	`delta`	Exit/replacement intensity.	0.10
λ_{gov}	`lam_gov`	Governance-event intensity.	1.5
$\mu_{0,Q}$	`mu0_A`	Baseline quality drift.	0.02
$\mu_{1,Q}$	`mu1_A`	Marginal productivity of delegate effort.	0.07
α_Q	`alpha_A`	Curvature of effort in the quality drift.	1.10
σ_Q	`sigma_A`	Quality diffusion volatility.	0.10
γ_Q	`gamma_A`	Proportional quality loss after an attack.	0.40
κ_Q	`kappa_A`	Quality depreciation term.	0.02
f_0	`f0`	Scale parameter in fee inflow and delegate compensation $f(\bar{n}) = f_0 \bar{n}^{\beta_f}$.	1.00

Symbol	Notebook name	Description	Value
β_f	`beta_f`	Concavity of fee inflow and delegate compensation.	0.50
g_0	`g0`	Baseline external treasury inflow per participant.	1.00
g_1	`g1`	Sensitivity of external treasury inflow to quality.	1.00
γ_T	`gamma_T`	Curvature of the external inflow function $g(Q) = g_0 + g_1 Q^{\gamma_T}$.	1.00
a_0	`a0`	Scale parameter in the effective participant base $\bar{n}(Q)$.	0.05
ε	`eps`	Elasticity of the effective participant base with respect to quality.	0.50
δ_n	`delta_n_const`	Denominator in $\bar{n}(Q)$.	0.50
c_A	`cA`	Cost of acquiring information.	0.05
x_D	`xD`	Token grant per governance event.	1.00
λ_{token}	`lam_token`	Token-arrival parameter.	$= \lambda_{\text{gov}}$
c_2	`c2`	Quadratic effort-cost parameter.	1.00
c_I	`cI`	Identity reset or attack cost.	0.10
$p_{0,\text{rep}}$	`p0_rep`	Prior reputation assigned to a new delegate after reset.	0.20
ΔU_{vote}	`DeltaU_vote`	Utility gain from the informed voting outcome in the information-acquisition payoff.	5.00
$e_{\min}$	`e_min`	Lower bound on delegate effort.	0.00
$e_{\max}$	`e_max`	Upper bound on delegate effort.	1.00

Symbol	Notebook name	Description	Value
Φ_Q	`PHI_A`	Scaling constant for the direct price effect of quality in $\partial W_H/\partial Q$.	1.00
Φ_P	`PHI_PRICE`	Scaling constant for the treasury-price channel in $\partial W_H/\partial Q$.	10.00
Φ_L	`PHI_LIQ`	Auxiliary liquidity scaling constant declared in the notebook.	4.00
ε_X	`EPS_X`	Numerical safeguard preventing division by zero in token supply.	10^{-8}

Single Path Simulation

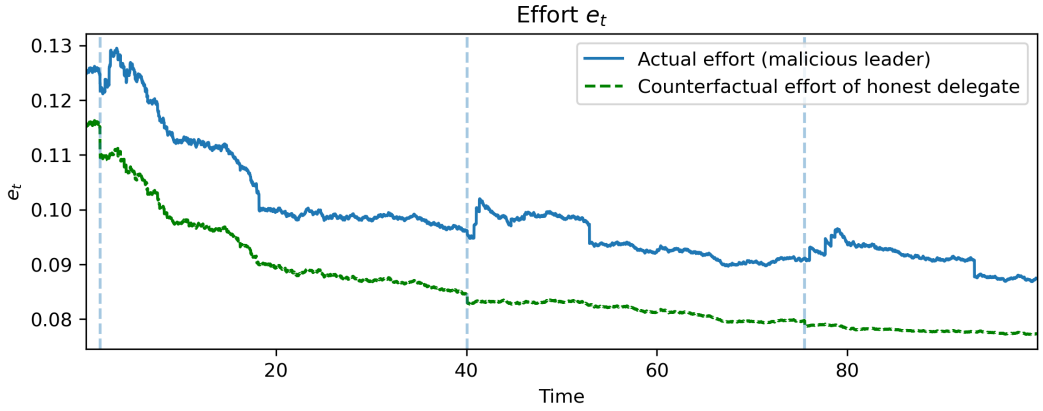


Figure A.1: This figure plots the effort level of a malicious leading delegate and the counterfactual effort level that would have been provided by an honest delegate. The simulation is run with the parameters defined in Table A.1.

Implementation of the simulation of Welfare comparisons

The welfare comparison is computed for two regimes. The first regime is the anonymous governance baseline. In this regime, tokenholders solve the monitoring subgame, malicious delegates may attack, and reputation evolves according to the Bayesian updating rule used in the baseline simulations. The

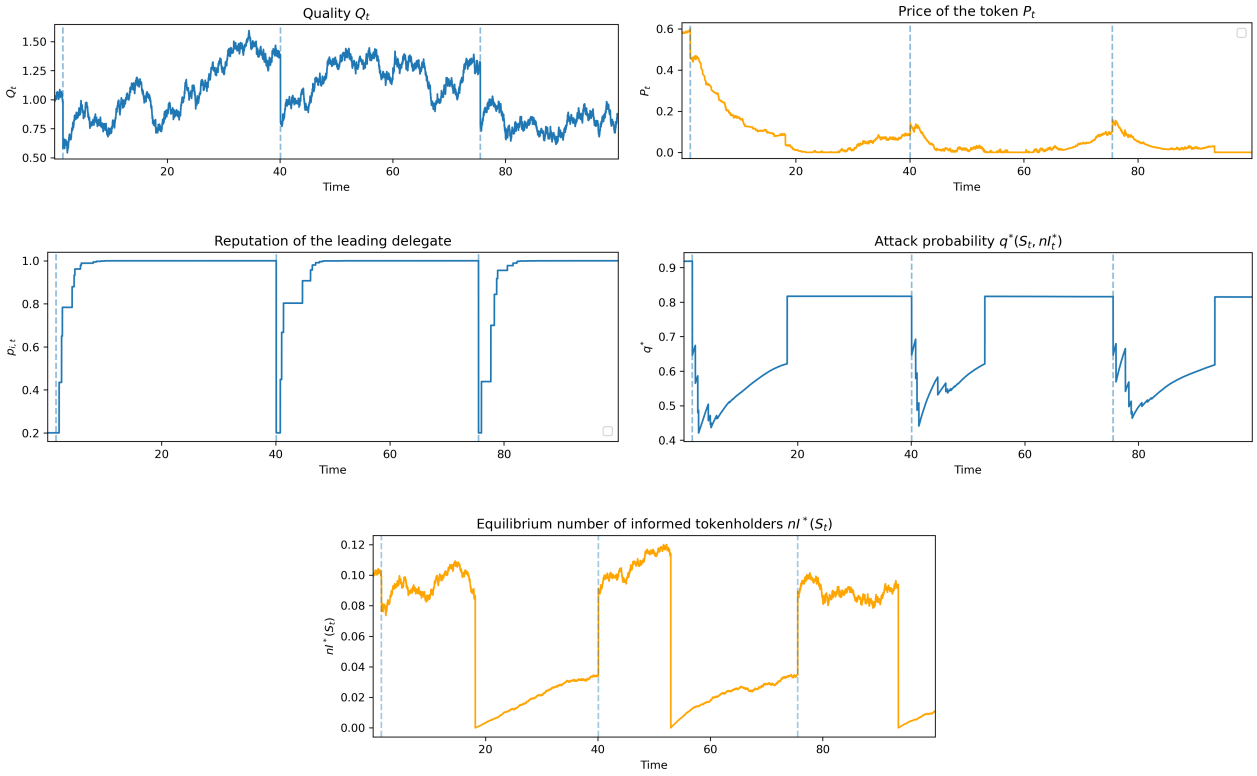


Figure A.2: This figure plots the quality of the platform Q, the price of the token P, the reputation of the leading delegate p_i, the attack probability chosen by the malicious leading delegate q, and the number of informed voters n_I. The simulation is run with the parameters defined in Table A.1.

second regime is the authenticated benchmark. In this regime, the delegate is known to be honest, attacks are ruled out, and tokenholders do not pay monitoring costs. The incumbent type is drawn once at the beginning of each Monte Carlo path from the prior probability p_0. In the code this is imposed by setting

$$p_{i,k} = p_{j,k} = 1, \qquad n_k^I = 0, \qquad \rho_k = 0, \qquad q_k = 0$$

along the authenticated path.

Fast malicious marginal value. To keep the welfare Monte Carlo simulation computationally feasible, in this extension we use a fast version of the malicious marginal-value calculation. The original malicious derivative routine recomputes the monitoring equilibrium inside the derivative calculation. The welfare extension instead reuses the already-computed value $n_k^{I,*}$ from the current state. The code computes the marginal effect of quality on treasury inflows as

$$\frac{\partial \mu_T(Q)}{\partial Q} = g'(Q)\bar{n}(Q) + g(Q)\bar{n}'_Q(Q),$$

where $\mu_T(Q) = g(Q)\bar{n}(S)$. When the adoption cap in the original simulation binds, the code sets $\bar{n}'_Q(Q) = 0$. Otherwise it uses the derivative implied by the adoption function in the baseline code. The malicious delegate also values the option of attacking a larger future treasury. Using the reduced-form attack probability q_k, the code defines the local attack hazard and expected attack time as

$$h_k = \lambda_{\text{gov}} q_k, \qquad \mathbb{E}[\tau_k] = \frac{1}{h_k + 10^{-10}}$$

The marginal option value of quality is approximated by

$$\frac{\partial V_M^{\text{opt}}}{\partial Q} = \exp\{-r\mathbb{E}[\tau_k]\}\frac{\partial \mu_T(Q_k)}{\partial q}\mathbb{E}[\tau_k].$$

The marginal value used for malicious effort is therefore

$$\frac{\partial W_M}{\partial Q} = \max\left\{\frac{\partial W_H}{\partial Q} + \frac{\partial V_M^{\text{opt}}}{\partial Q}, 0\right\}$$

This leaves the economic content of the baseline effort rule unchanged, but avoids solving the monitoring subgame twice at the same state.

Flow welfare For a simulated state S_k, the implemented flow welfare is

$$g_k^R = \bar{n}(Q_k)u(Q_k) - c(e_k) - \lambda_{\text{gov}}\bar{n}(S_k)\rho_k\kappa,$$

where $R \in \{A, N\}$ indexes the anonymous and authenticated regimes. In the baseline welfare run below $u(Q) = Q$. Transfers such as wages, treasury transfers, and token grants are excluded from welfare. They affect welfare only through their effects on the simulated state and on incentives.

Stationary welfare estimator For each simulated path, the code approximates stationary welfare by a post-burn-in time average. Let $b = \lfloor \texttt{burn\_in\_frac} \cdot K \rfloor$, where K is the final simulation index. The estimated stationary flow welfare in regime R is

$$\widehat{W}^R = \frac{1}{t_K - t_b} \int_{t_b}^{t_K} g_t^R \, dt.$$

Numerically, the integral is computed with the trapezoidal rule:

$$\widehat{W}^R = \frac{1}{t_K - t_b} \texttt{trapz}\left(\{g_k^R\}_{k=b}^K, \{t_k\}_{k=b}^K\right)$$

Because the same discount rate applies to both regimes, comparing average flow welfare is equivalent to comparing discounted welfare initialized from the invariant distribution.

Monte Carlo comparison. The anonymous and authenticated paths are simulated with the same seed. This pairs the two regimes path by path and reduces simulation noise in the welfare ratio. The code stores

$$\widehat{W}_s^A, \qquad \widehat{W}_s^N, \qquad R_s = \frac{\widehat{W}_s^A}{\widehat{W}_s^N}$$

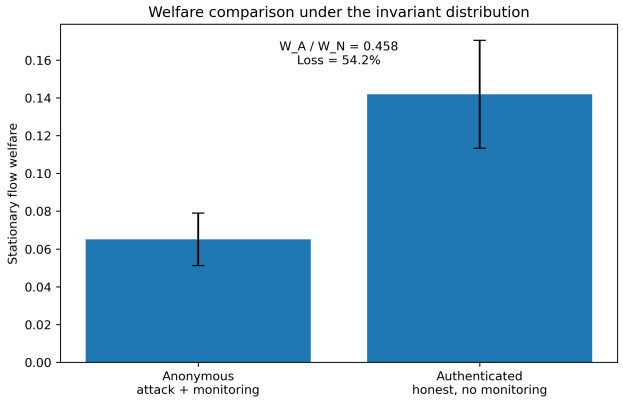
It then reports the sample means, standard errors, normal-approximation confidence intervals, path-level ratios, the ratio of means,

$$\frac{\overline{W}^A}{\overline{W}^N},$$

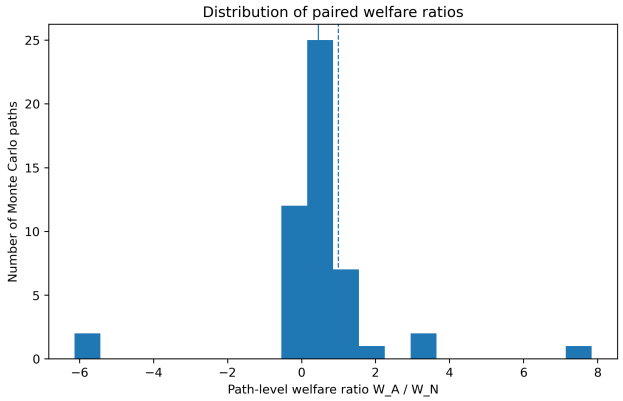
and the implied welfare loss,

$$\widehat{L} = 100\left(1 - \frac{\overline{W}^A}{\overline{W}^N}\right)$$

Welfare simulations



(a) Stationary welfare by governance regime



(b) Distribution of paired welfare ratios

Figure A.3: Stationary welfare under anonymous and authenticated governance. The figure reports the baseline welfare comparison computed from 50 paired Monte Carlo paths. Each path is simulated for $T = 50$ with time step $\Delta t = 0.25$, and the first half of each path is discarded as burn-in. Initial conditions are $A_0 = 1$, $T_0 = 10$, and $X_0 = 25$, with delegate token positions $x_i = x_j = 5$ and reset value $\texttt{Wout\_reset} = 10$. User utility is linear, $u(Q) = Q$, and the monitoring fixed point is solved on a grid of 15 points. Panel (a) reports mean stationary flow welfare in the anonymous and authenticated regimes, with 95 percent normal-approximation confidence intervals. Panel (b) reports the distribution of path-level paired welfare ratios $\widehat{W}_s^A / \widehat{W}_s^N$..