

Optimistic Resolution Mechanisms for Prediction Markets: Incentives, Equilibria, and Design Limits

Mengzhong Ma¹, Te Bao¹, and Yonggang Wen¹

Nanyang Technological University, Singapore mengzhon001@e.ntu.edu.sg,
baote@ntu.edu.sg, ygwen@ntu.edu.sg

Abstract. This paper studies the incentive fragility of the resolution layer adopted by major crypto prediction markets. We model settlement as a strategic game between a proposer and potential disputers who face costly verification. Disputers may free-ride, producing a no-challenge equilibrium in which incorrect reports finalize. Adjudication noise further weakens deterrence by lowering the expected return to disputing. We microfound adjudication as token-weighted voting and derive a concentration-sensitive Hoeffding bound that motivates an empirical Adjudication Risk Index (ARI). Using UMA voting and staking data from 2020–2026, we estimate daily voting correctness and construct event-level and daily ARIs, documenting pronounced time variation and spikes that coincide with periods of increased voting-power concentration.

Keywords: Prediction markets · Optimistic Oracle resolution · token-weighted adjudication · stake concentration risk.

1 Introduction

In recent years, crypto-native prediction markets, led by Polymarket, have drawn outsized attention as a real-time alternative to polls and pundit forecasts. Polymarket drew attention for tracking (and outperforming) polls during the 2024 U.S. election, and has also been highlighted by Ethereum co-founder Vitalik Buterin as a leading example of prediction markets for “information finance” [3]. However, soundness of prediction markets depends on whether the final answer can be correctly settled (resolution). When a market settles, the protocol must map messy real-world events into an on-chain outcome, and any weakness in the pipeline can compromise incentives and trust. This paper studies that resolution layer, modeling the propose–dispute mechanism and the token-weighted adjudication that is invoked upon challenge, and develops an empirical risk index that quantifies when adjudication is most vulnerable.

Although prediction markets are supported by many platforms (e.g., Polymarket, Bookies, FORE Protocol, and PolyBet), almost all of them are resolved by Optimistic Oracle by UMA [11]. The oracle works under a three-stage design: report–dispute–adjudication. A reporter submits a outcome when a prediction

market finalize. The report is accepted by default unless challenged by a disputer during a fixed challenge window, in which case an on-chain adjudication (finalized by token-weighted voting) procedure determines the winner [11]. We build a model to formalize such procedure and show that the effectiveness of optimistic resolution hinges on two friction points.

First, because verification is a public good, multiple potential disputers can rationally free-ride on one another, generating a coordination failure in which no one verifies and dispute, so incorrect reports finalize (Theorem 1). Second, deterrence depends critically on adjudication quality. If the token-weighted voting is sufficiently noisy, disputing becomes privately unprofitable and enforcement collapses even absent verification costs. We then microfound adjudication through a token-weighted voting model and derive a concentration-sensitive Hoeffding bound that yields an empirical Adjudication Risk Index (ARI). Using UMA voting and staking data from 2020–2026, we estimate daily voting correctness and document that adjudication risk exhibits pronounced time variation, with spikes concentrated in late 2022, late 2024, and late 2025. The system-level ARI constructed from staking distributions aligns these spikes with temporary increases in voting-power concentration. Finally, we show how the protocol’s dispute-sink parameter can be used as a policy lever. By subsidizing disputes during high-risk periods, the mechanism can tolerate noisier adjudication while preserving incentives to challenge incorrect reports.

Related work Our paper contributes to three strands of literature. First, it builds on the economics of prediction markets, which study how market prices aggregate dispersed information [13,5]. Second, it relates to the growing work on blockchain oracles and decentralized adjudication [1], including token-incentivized dispute resolution and decentralized courts [7]. Third, recent empirical and systems-oriented studies of crypto prediction markets document that Polymarket relies on UMA’s Optimistic Oracle for market resolution and that complex resolutions and concentrated voting power can generate discrepancies between oracle outcomes and underlying events [9].

2 The Model

2.1 Environment and Players

We consider a single prediction market that settles on the realization of an exogenous event with an objectively correct outcome. Let the underlying event resolve to a state $\theta \in \Theta$, where Θ is a finite outcome space (e.g., $\Theta = \{0, 1\}$ for No/Yes). All agents share a common prior π over Θ . The protocol must output a finalized outcome $\tilde{\theta} \in \Theta$ that is used to settle contingent claims.

There is one proposer P who submits an outcome report, and a set of potential disputers $D = \{1, \dots, n\}$ who may challenge an incorrect report. We treat the adjudication device (UMA’s DVM) as a separate stage that is invoked only upon dispute. Agents are risk-neutral with quasi-linear preferences. The

proposer may have an incentive to misreport: if an incorrect outcome finalizes, the proposer obtains an incremental private benefit $\Delta \geq 0$ relative to correct finalization (which we normalize to zero). Potential disputers may incur verification costs $c_i \geq 0$ to learn θ with certainty; these costs capture off-chain effort required to check ground truth. We denote the minimum verification cost by $c_{\min} = \min_{i \in D} c_i$.

2.2 Timing

Time is discrete with three dates $t \in \{0, 1, 2\}$. At $t = 0$, the proposer observes the realized outcome and chooses to report; at $t = 1$, potential disputers observe $\hat{\theta}$ and decide whether to dispute; and at $t = 2$, adjudication is implemented regarding the dispute.

At $t = 0$ (report stage) After observing θ , the proposer chooses $\hat{\theta} \in \Theta$ and posts bond $B > 0$. If the report is ultimately finalized as $\tilde{\theta} = \hat{\theta}$ without a dispute, the proposer's bond is returned and the proposer obtains utility $\Delta \cdot \mathbf{1}\{\hat{\theta} \neq \theta\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function. If a dispute occurs, bond redistribution depends on adjudication: if the proposer wins, the proposer receives $1.5B$ from the pool of $2B$ bonded funds; if the proposer loses, the proposer forfeits its bond B . Since the proposer has already posted B , the proposer's net bond payoff conditional on winning a dispute equals $+0.5B$, and conditional on losing equals $-B$. In addition, the proposer receives the misreport benefit Δ only when an incorrect report finalizes, which under dispute can occur only if the proposer wins despite $\hat{\theta} \neq \theta$.

At $t = 1$ (challenge stage) Each potential disputer $i \in D$ observes the report $\hat{\theta}$ and chooses whether to verify and whether to dispute. Verification is modeled as an information acquisition action. By paying cost c_i , disputer i learns θ perfectly, while otherwise she relies on a private signal s_i yielding posterior $q_i(\hat{\theta}) = \Pr(\hat{\theta} \neq \theta \mid s_i, \hat{\theta})$. To dispute, disputer i posts the dispute bond B before the challenge window ends. If no dispute occurs, disputers receive payoff 0. If a dispute occurs and disputer i is the disputing party, then upon winning the dispute she receives $1.5B$ and upon losing she forfeits B , implying net bond payoffs of $+0.5B$ (win) and $-B$ (lose), respectively, in addition to any verification cost incurred. The remaining $0.5B$ from the combined bonds is transferred to a sink (the UMA Store), which makes disputes negative-sum and therefore affects disputers' incentives. During the challenge window, at most one dispute is active.

At $t = 2$ (adjudication stage) Adjudication is implemented via a token-weighted voting process (the DVM). Rather than modeling the full voter game in the baseline specification, we summarize adjudication quality with a soundness parameter $\varepsilon \in [0, 1/2]$: conditional on a dispute, the adjudicator O selects the truthful side with probability at least $1 - \varepsilon$. Thus, if $\hat{\theta} = \theta$, the proposer wins

with probability at least $1 - \varepsilon$. If $\hat{\theta} \neq \theta$, the disputer wins with probability at least $1 - \varepsilon$. This reduced-form treatment isolates the optimistic-resolution incentives in a way that can endogenize ε from token distribution and voting incentives separately. The finalized outcome $\tilde{\theta}$ equals $\hat{\theta}$ if the proposer wins and equals θ if the disputer wins. Bond redistribution follows as: the winner obtains $1.5B$ and $0.5B$ is transferred to the sink.

2.3 Equilibrium

Fix a report $\hat{\theta} \in \Theta$ submitted at $t = 0$. At $t = 1$, a set of $m \geq 1$ potential disputers (“active disputers”) simultaneously decide whether to incur a verification cost in order to learn the ground truth and, if warranted, trigger a dispute. We impose a reduced-form tie-breaking rule consistent with the protocol constraint that at most one dispute is active.¹ This assumption captures the public-good nature of verification: multiple agents may redundantly verify, but only one dispute can be realized.

Each disputer $i \in \{1, \dots, m\}$ observes $(\hat{\theta}, s_i)$ and forms posterior belief $q_i(\hat{\theta}, s_i) = \Pr(\hat{\theta} \neq \theta \mid \hat{\theta}, s_i)$. For tractability we analyze a symmetric information case in which, conditional on $(\hat{\theta}, s_i)$, all active disputers share a common posterior $q \in [0, 1]$ and have identical verification cost $c > 0$.

Each disputer chooses an action from $a_i \in \{N, V\}$, where N denotes *no action*, and V denotes *verify*: pay c to learn θ perfectly and submit a dispute *iff* the report is incorrect (i.e., *iff* $\hat{\theta} \neq \theta$). If $\hat{\theta} = \theta$, a verifier does not dispute.

Conditional on an incorrect report and a dispute, the disputing party wins with probability at least $1 - \varepsilon$. Given the bond mechanics, the selected disputing party receives net payoff $+0.5B$ if she wins and $-B$ if she loses. Define the expected net bond payoff from disputing a *known-wrong* report as

$$A = (1 - \varepsilon)(0.5B) + \varepsilon(-B) = (0.5 - 1.5\varepsilon)B. \quad (1)$$

Observe that $A > 0$ *iff* $\varepsilon < 1/3$. When $A \leq 0$, even a certainly incorrect report is not privately profitable to dispute, reflecting the negative-sum dispute sink.

Expected Payoff from Verifying Fix a candidate symmetric mixed strategy in which each other disputer verifies with probability $p \in [0, 1]$. Consider disputer i ’s deviation payoff from choosing V .

If the report is incorrect (an event of probability q), then disputer i submits a dispute, and she is selected as the active disputing party with probability equal to her expected selection share among all challengers. Let $K \sim \text{Binomial}(m - 1, p)$ denote the number of other disputers who verify (hence submit a dispute) when the report is incorrect. Disputer i is selected with probability $1/(1 + K)$, so her

¹ If multiple disputers attempt to dispute within the challenge window, exactly one disputer is selected uniformly at random to be the active disputing party. Only the selected party posts the dispute bond and participates in adjudication; all verification costs are borne by those who chose to verify regardless of selection.

expected selection probability conditional on choosing V is $\mathbb{E}\left[\frac{1}{1+K}\right] = \frac{1-(1-p)^m}{mp}$. Therefore, the expected payoff from choosing V is

$$U_V(p; q) = q \cdot A \cdot \frac{1 - (1-p)^m}{mp} - c, \quad (2)$$

where A is given by Equation (1). The payoff from choosing N is normalized to $U_N = 0$.

Proposition 1 (Symmetric mixed-strategy verification equilibrium). *Assume $\varepsilon < 1/3$ so that $A > 0$. Consider the $t = 1$ verification subgame with $m \geq 1$ identical disputers with verification cost $c > 0$ and common posterior $q \in [0, 1]$ that the report is incorrect.*

1. **No-verification equilibrium.** *If $qA \leq c$, then $p^* = 0$ is a (symmetric) equilibrium. Moreover, $p^* = 0$ is the unique symmetric equilibrium in this region.*
2. **Full-verification equilibrium.** *If $\frac{qA}{m} \geq c$, then $p^* = 1$ is a (symmetric) equilibrium.*
3. **Interior mixed equilibrium.** *If $\frac{qA}{m} < c < qA$, then there exists a unique symmetric mixed equilibrium with $p^* \in (0, 1)$ characterized by indifference:*

$$qA \cdot \frac{1 - (1-p^*)^m}{mp^*} = c. \quad (3)$$

The resulting probability that an incorrect report is challenged is

$$\alpha^* = 1 - (1-p^*)^m \in (0, 1). \quad (4)$$

Proof. See Appendix A

Coordination Failure and Endogenous Deterrence Proposition 1 highlights a public-good problem: as the number of potential disputers m grows, each verifier's expected selection share shrinks, weakening incentives to verify and potentially driving the system toward non-enforcement.

Let $\alpha \in [0, 1]$ denote the probability that an incorrect report is challenged at $t = 1$. Under the symmetric strategy p , this probability is

$$\alpha(p) = 1 - (1-p)^m.$$

The proposer's expected payoff from misreporting (conditional on $\hat{\theta} \neq \theta$) is

$$U_P^{\text{mis}}(\alpha) = (1-\alpha)\Delta + \alpha\left((-1 + 1.5\varepsilon)B + \varepsilon\Delta\right), \quad (5)$$

since with probability $1-\alpha$ no dispute occurs and the proposer obtains Δ , while with probability α a dispute occurs and the proposer wins (incorrectly) with probability at most ε .

Truthful reporting is sequentially rational for the proposer whenever $U_P^{\text{mis}}(\alpha) \leq 0$, i.e., whenever

$$\alpha \geq \alpha_{\min} = \frac{\Delta}{(1-1.5\varepsilon)B + (1-\varepsilon)\Delta}. \quad (6)$$

Theorem 1 (Coordination failure breaks deterrence). *Assume $\varepsilon < 1/3$ and consider the challenge subgame described above. Let (p^*, α^*) be the symmetric verification equilibrium from Proposition 1 and define α_{\min} as in (6).*

1. **Enforcement collapse.** *If $qA \leq c$, then $\alpha^* = 0$ and the mechanism provides no endogenous deterrence: a proposer with $\Delta > 0$ strictly benefits from misreporting whenever it is payoff-relevant.*
2. **Partial enforcement.** *If $\frac{qA}{m} < c < qA$, then $\alpha^* \in (0, 1)$. A truthful PBE exists only if $\alpha^* \geq \alpha_{\min}$. If $\alpha^* < \alpha_{\min}$, then there exists a PBE in which misreports occur with positive probability.*
3. **Full enforcement.** *If $\frac{qA}{m} \geq c$, then $p^* = 1$ (hence $\alpha^* = 1$) is an equilibrium of the challenge subgame, and the proposer is truthful whenever*

$$B \geq \frac{\varepsilon \Delta}{1 - 1.5\varepsilon},$$

which is the special case of (6) with $\alpha = 1$.

Theorem 1 formalizes the core fragility of optimistic resolution. Even when bonds are large enough to make disputing privately profitable conditional on being the disputing party, the presence of many potential disputers (larger m and then greater free-riding) can reduce each agent's expected selection share and induce a low-challenge equilibrium, undermining deterrence. Moreover, high verification cost breaks disputing, characterized by $qA \leq c$ (enforcement collapse). In practice, ambiguity of the questions raises the effective verification cost, thereby increasing the likelihood of no dispute and incorrect finalization.²

Theorem 1 mainly characterizes enforcement failures that arise from non-cooperative free-riding or miscoordination. Enforcement failure can also arise from agents collusion either by bribing or centralization of voting power.

Proposition 2 (Two cases of collusion). *Suppose side-payments (bribes) between agents can be implemented off-chain or on-chain and are enforceable. Then each of the following conditions is sufficient for deterrence to fail.*

1. **Adjudication failure (increasing ε destroys private enforcement).** *If collusion centralization of voting power in the adjudication layer raises the voting error rate to*

$$\varepsilon \geq \frac{1}{3},$$

then $A \leq 0$ and disputing a known-wrong report is weakly unprofitable even absent verification costs. Consequently, there is no parameter choice of B that can restore private incentives to dispute through the bond mechanism alone, and the resolution layer becomes non-credible.

² There are tremendous episodes of resolution failure on Polymarket where the core problem is ambiguous in wording. For example, readers can refer to the following media outlets: <https://www.wired.com/story/volodymyr-zelensky-suit-polymarket-rebellion/>? or <https://www.ft.com/content/985ae542-1ab4-491e-8e6e-b30f6a3ab666?>.

2. **Proposer–disputer buyoff (silencing the marginal enforcer).** Suppose some disputer i would, absent collusion, verify and dispute a known-wrong report (e.g., $A > c_i$). If

$$\Delta \geq A - c_i,$$

then the proposer can offer an enforceable transfer $t \in [A - c_i, \Delta]$ that deters a dispute while preserving a profitable misreport. In particular, if

$$\Delta \geq A - c_{\min},$$

the proposer can buy off the lowest-cost enforcer, allowing an incorrect report to finalize without challenge in equilibrium.

It is worth noting that under free entry, a buyoff must deter every potential disputer, making side-payments costly and fragile. Hence proposer–disputer buyoffs are comparatively unlikely. In contrast, adjudication failure is more systemic. If voting power is sufficiently concentrated, the induced error rate ε can make disputes unprofitable for all challengers (e.g., $A = (0.5 - 1.5\varepsilon)B \leq 0$ when $\varepsilon \geq 1/3$). Crucially, a dishonest proposer need not spend to acquire voting power. It can exploit an existing staking distribution or turnout pattern that makes an incorrect report likely to prevail at adjudication, thereby collapsing private enforcement incentives.

3 Adjudication Voting and Empirical Risk Indices

While the truthful reporting may fail due to deterrence breaking in the reporting–disputing game, its presumption ($\varepsilon < \frac{1}{3}$) does not necessarily always hold (i.e., in cases of adjudication failure). Rather than abstracting the procedure into the adjudication quality ε , in this section we microfound the procedure through a token-weighted voting model, i.e., the procedure of DVM under the UMA protocol.³ With the microfound basis, we further build empirical risk indices of adjudication failure.

3.1 Token-Weighted Voting

Consider a binary adjudication between the truthful outcome T and an incorrect alternative F . Voters $j \in \{1, \dots, N\}$ hold staked voting weights $w_j \geq 0$, with total stake $W = \sum_{j=1}^N w_j$. Let $X_j \in \{0, 1\}$, $\Pr(X_j = 1) = \rho$, denote whether voter j votes for the truthful outcome, assumed independent across j . The truthful vote weight is $S = \sum_{j=1}^N w_j X_j$. Let $\tau \in (0, 1)$ denote the activation threshold (e.g., $\tau = 0.65$).⁴ We model adjudication as selecting F (an error) whenever the

³ Readers can find the documentation of the UMA voting procedure from [12].

⁴ If any single voting option reaches the consensus threshold (currently 65% of staked UMA), it is considered the correct outcome [10].

incorrect side reaches threshold τW , i.e., $W - S \geq \tau W$ and thus $S \leq (1 - \tau)W$. Accordingly, the adjudication error probability is

$$\varepsilon = \Pr(S \leq (1 - \tau)W).$$

Lemma 1 (Weighted Hoeffding bound for adjudication error). *Define the informational margin $\delta = \rho + \tau - 1$. Applying Hoeffding’s inequality [6], if $\delta > 0$, we can construct*

$$\varepsilon = \Pr(S \leq (1 - \tau)W) \leq \exp\left(-\frac{2\delta^2 W^2}{\sum_{j=1}^N w_j^2}\right), \quad (7)$$

which is the concentration-adjusted upper bound on the probability of adjudication error. If $\delta \leq 0$, the bound is vacuous.

Proof. See Appendix A

Motivated by Lemma 1, we define the Adjudication Risk Index (ARI) as

$$\text{Risk}(\rho; \tau, \{w_j\}) = \exp\left(-\frac{2(\rho + \tau - 1)^2 W^2}{\sum_{j=1}^N w_j^2}\right), \quad (8)$$

characterizing the upper-bound proxy for adjudication error when $\rho + \tau > 1$. Equation (8) implies how to decrease the the risk of bad adjudication. Protocols can either incentivize voters to vote for the correct option (i.e., higher ρ), increase the threshold of consensus τ , or increase the effective number of independent voters $N_{\text{eff},e} = \frac{W_e^2}{S_{2,e}}$ by decentralizing token staking.

3.2 Empirical Method of ARI Estimation

To operationalize the voting-quality parameter ρ , we estimate a day-specific correctness rate ρ_d from event-level vote outcomes using a binomial generalized linear mixed model (GLMM) with day random intercepts [8,2,4]. For voting event e on day $d(e)$, let $y_e \in [0, 1]$ denote the observed correct-side weight share. Using the realized weight distribution, we form an effective trial size $n_e = N_{\text{eff},e} = \frac{W_e^2}{\sum_{j \in V_e} w_{j,e}^2}$, $k_e = n_e y_e$, and model

$$k_e \mid \rho_{d(e)} \sim \text{Binomial}(n_e, \rho_{d(e)}), \quad \text{logit}(\rho_d) = \alpha + u_d, \quad u_d \sim \mathcal{N}(0, \sigma^2). \quad (9)$$

We report $\hat{\rho}_d = \text{logit}^{-1}(\hat{\alpha} + \hat{u}_d)$ and construct day-level uncertainty bands from the fitted model. In practice, we adopt the posterior distribution over ρ_d implied by the mixed model (via Bayesian GLMM computation).

Given a day-level estimate $\hat{\rho}_d$ and the protocol activation threshold τ , we compute an event-level ARI from Equation (8). For each event e on day $d(e)$,

define total cast weight W_e and the second moment $S_{2,e} = \sum_{j \in V_e} w_{j,e}^2$. The event-level risk index is

$$\text{ARI}_e(\hat{\rho}_d) = \exp\left(-\frac{2(\hat{\rho}_d + \tau - 1)^2 W_e^2}{S_{2,e}}\right), \quad (10)$$

which can be aggregated to form the day-level ARI (i.e., the $N_{\text{eff},e}$ -weighted mean).

3.3 Data and Estimation Results of Historical ARIs

Data We adopt the voting and staking data from the UMA protocol to build ARIs. The dataset starts from Jan 2020 and ends on Feb 2026. UMA’s dispute-resolution voting infrastructure updates in design around the DVM 2.0 upgrade during Q1 2023. Accordingly, our sample includes a pre-DVM 2.0 period that relies on the legacy DVM voting contract (“Voting”, i.e., Dune label `Voting_v1`) and a post-DVM 2.0 period that relies on the upgraded voting contract (“VotingV2”, i.e., Dune label `Voting_v2`). We extract voting-related events (e.g., commit-reveal and tally outcomes) from the corresponding contract addresses and align all resolution activity to this contract transition for subsequent analyses.⁵

Estimation Fig. 1 illustrates the evolution of the UMA protocol since its launch in 2020. Although the \$UMA token price has declined gradually since 2021, the number of events requiring voting has risen steadily, consistent with the proliferation of prediction markets in 2025 and 2026.

According to Equation (9), we estimate voters’ correctness rate ρ , with the resulting daily series $\hat{\rho}_d$ reported in Table 1. The estimates range from 0.6353 to 0.9995. Using $\hat{\rho}_d$, we construct the event-level adjudication risk index (ARI) via Equation (10) and summary statistics are reported in Table 2. While the point estimates ARI_e remain below $\frac{1}{3}$ (with a maximum of 0.2648), the conservative bound based on the lower endpoint of the ρ_d credible interval exceeds $\frac{1}{3}$ for a small set of events (up to 0.76), highlighting the tail risk of adjudication error. The same pattern can be observed in the daily ARI, which we construct by aggregating the event-level indices, as shown in Fig 2. Spikes in adjudication error risk cluster around late 2022, late 2024, and late 2025.

⁵ Legacy DVM voting contract (“Voting”, `Voting_v1` on Dune): https://dune.com/data?category=decoded_project&namespace=umaproject&contract=Voting_v1&blockchains=ethereum. Upgraded DVM 2.0 voting contract (“VotingV2”): <https://etherscan.io/address/0x004395edb43efca9885cedad51ec9faf93bd34ac>. UMA documentation notes that the rebuilt DVM 2.0 iteration was released in Q1 2023: <https://docs.uma.xyz/protocol-overview/dvm-2.0>. Polymarket’s resolution documentation (including the 2-hour challenge window and escalation to UMA voting upon dispute): <https://docs.polymarket.com/polymarket-learn/markets/how-are-markets-resolved>.

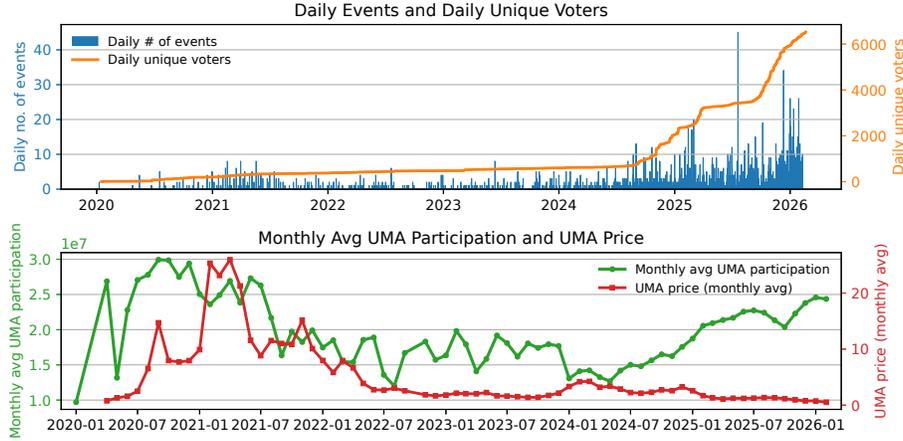


Fig. 1. Voting events, voter participants, and \$UMA staking participation each event.

Table 1. Daily estimates of ρ_d and GLMM fit statistics.

Panel A: Distribution of daily estimates								
Statistic	N	Mean	SD	P25	P50	P75	Min	Max
$\hat{\rho}_d$	747	0.9757	0.0508	0.9879	0.9929	0.9955	0.6353	0.9995
CI width (95%)	747	0.0589	0.0589	0.0258	0.0429	0.0747	0.0022	0.4125
Panel B: GLMM parameters and diagnostics								
Parameter	Estimate	95% interval (HDI)		\hat{R}	ESS			
α (intercept)	5.6152	[5.3095, 5.9101]		1.004	563			
σ_u (day RE s.d.)	2.2284	[2.0015, 2.4768]		1.004	700			
$\text{logit}^{-1}(\alpha)$ (pooled mean)	0.9964							

Notes: ρ_d is estimated via a binomial GLMM with day random intercepts: $k_e \sim \text{Binomial}(n_e, \rho_{d(e)})$, $\text{logit}(\rho_d) = \alpha + u_d$, $u_d \sim \mathcal{N}(0, \sigma_u^2)$. Here $n_e = W_e^2 / \sum_{j \in V_e} w_{j,e}^2$ and $k_e = n_e y_e$. $\hat{\rho}_d$ reports posterior means; “CI width” is $\rho_{d,97.5\%} - \rho_{d,2.5\%}$. Panel B reports posterior means and 95% HDIs, with MCMC diagnostics \hat{R} and ESS.

Besides constructing event-level ARI from realized voting events, we also construct system-level ARI from the \$UMA token distribution, using the same daily estimates $\hat{\rho}_d$. Because UMA enabled on-chain staking only after the DVM 2.0 (V2) upgrade (with earlier rounds relying on round-specific snapshots), we can observe staking distributions only from the DVM 2.0 launch in Q1 2023 onward. Using data from 2023 onward, Fig. 3 plots the system-level ARI together with the daily effective number of independent voters, $N_{\text{eff},d}$. While $N_{\text{eff},d}$ exhibits a gradual upward trend overall, it experiences pronounced declines followed by recoveries in late 2023, late 2024, and late 2025. These three “V-shaped” episodes coincide with spikes in the system-level ARI, indicating that increased concentration in voting power is associated with sharp increases in adjudication risk.

Table 2. Event-level adjudication risk index.

Statistic	N	Mean	SD	P25	P50	P75	Min	Max
$\text{Risk}_e(\hat{\rho}_d)$	2738	0.0015	0.0086	1.36e-8	5.16e-8	1.07e-6	2.08e-9	0.2652
Lower bound	2738	0.0011	0.0075	1.13e-8	3.43e-8	3.26e-7	6.28e-10	0.2563
Upper bound	2738	0.0053	0.0323	2.50e-8	1.44e-7	8.45e-6	3.95e-9	0.7426

Notes: The event-level risk index is computed as $\text{Risk}_e(\rho) = \exp(-2(\rho + \tau - 1)^2 W_e^2 / S_{2,e})$ with $\tau = 0.65$. The “Lower bound” and “Upper bound” evaluate $\text{Risk}_e(\rho)$ at the upper and lower endpoints of the daily 95% credible interval for ρ_d , respectively.

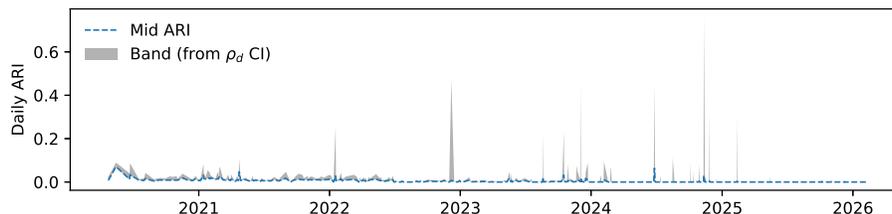


Fig. 2. Daily Adjudication Risk Index (ARI) constructed from event-level indices.

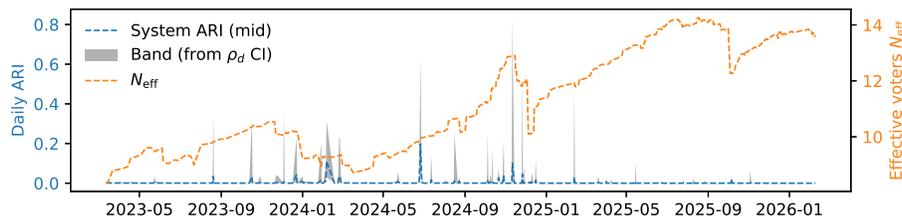


Fig. 3. Daily Adjudication Risk Index (ARI) constructed from \$UMA staking distribution.

With historical ARIs, policies on the dispute sink can be adjusted to tolerant periods of noisy adjudication. According to Equation (1), rather than sinking $0.5B$, we parameterize it to αB , so the winner gets $(2 - \alpha)B$. Then the presumption enforcing disputing becomes $\varepsilon < \varepsilon^*(\alpha) = \frac{1-\alpha}{2-\alpha}$, which can be larger than $1/2$ for negative α . Therefore, the protocols can subsidizes disputes ($\alpha < 0$) during periods of high adjudication error (indicated by high ARI), making the sink parameter dynamically adjusted to fit the token staking distribution change.

4 Conclusion

This paper analyzes the resolution layer of prediction markets that rely on UMA’s Optimistic Oracle. We model settlement as a report–dispute–adjudication game and show that optimistic resolution can fail due to (i) free-riding in verifi-

cation, which can generate a no-challenge equilibrium, and (ii) noisy or captured adjudication, which can make disputing privately unprofitable. We then micro-found adjudication via token-weighted voting, derive a concentration-sensitive Hoeffding bound, and use it to construct an empirical Adjudication Risk Index (ARI).

Using UMA voting and staking data from 2020–2026, we estimate daily voting correctness and document substantial time variation in adjudication risk: daily ARI exhibits clustered spikes, and a system-level ARI based on staking distributions links these episodes to temporary increases in voting-power concentration. Finally, parameterizing the dispute sink as αB highlights a design lever. Dispute subsidies (negative α) can expand error tolerance and help sustain incentives to challenge during high-risk periods.

References

1. Beniiche, A.: A study of blockchain oracles. arXiv preprint arXiv:2004.07140 (2020)
2. Breslow, N.E., Clayton, D.G.: Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**(421), 9–25 (1993)
3. Buterin, V.: From prediction markets to info finance (Nov 2024), <https://vitalik.eth.limo/general/2024/11/09/infofinance.html>
4. Gelman, A., Hill, J.: Data analysis using regression and multilevel/hierarchical models. *Analytical Methods for Social Research* (2006)
5. Hanson, R.: Logarithmic market scoring rules for modular combinatorial information aggregation. Tech. rep., George Mason University (2002), working paper
6. Hoeffding, W.: Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**(301), 13–30 (1963)
7. Lesaege, C., George, W., Ast, F.: Kleros: Decentralized justice protocol (yellow paper). White paper (2021)
8. McCullagh, P., Nelder, J.A.: *Generalized Linear Models*. Chapman and Hall/CRC, 2 edn. (1989)
9. Saguillo, O., Suarez-Tangil, G., et al.: Unravelling the probabilistic forest: Arbitrage in prediction markets (2025), manuscript
10. UMA Protocol: Faqs: How is the correct answer decided in uma’s voting system? (nd), <https://docs.uma.xyz/faqs#how-is-the-correct-answer-decided-in-umas-voting-system>
11. UMA Protocol: How does uma’s oracle work? (nd), <https://docs.uma.xyz/protocol-overview/how-does-umas-oracle-work>
12. UMA Protocol: Voting walkthrough (nd), <https://docs.uma.xyz/using-uma/voting-walkthrough>
13. Wolfers, J., Zitzewitz, E.: Prediction markets. *Journal of Economic Perspectives* **18**(2), 107–126 (2004). <https://doi.org/10.1257/0895330041371321>

Appendix A Derivations and Proofs

Appendix A.1 Proof of Proposition 1.

By (2), verifying is optimal iff $U_V(p; q) \geq 0$, i.e., $qA \cdot \frac{1-(1-p)^m}{mp} \geq c$. Define

$$\phi(p) = qA \cdot \frac{1 - (1 - p)^m}{mp} \quad \text{for } p \in (0, 1]. \quad (\text{A1})$$

Let $g(p) = \frac{1-(1-p)^m}{p}$. Using the binomial expansion,

$$(1-p)^m = 1 - mp + \binom{m}{2}p^2 - \binom{m}{3}p^3 + \dots,$$

hence

$$1 - (1-p)^m = mp - \binom{m}{2}p^2 + O(p^3),$$

and therefore

$$g(p) = \frac{mp - \binom{m}{2}p^2 + O(p^3)}{p} = m - \binom{m}{2}p + O(p^2).$$

It follows that $\lim_{p \downarrow 0} g(p) = m$, so

$$\lim_{p \downarrow 0} \phi(p) = qA \cdot \lim_{p \downarrow 0} \frac{1 - (1-p)^m}{mp} = qA \cdot \frac{m}{m} = qA.$$

We can therefore extend ϕ continuously to $p = 0$ by setting $\phi(0) = qA$.

Monotonicity of $\phi(p)$ Since $qA/m > 0$ is a constant multiplier, it suffices to show that

$$h(p) = \frac{1 - (1-p)^m}{p}$$

is strictly decreasing on $(0, 1]$. Differentiate h :

$$h'(p) = \frac{(m(1-p)^{m-1})p - (1 - (1-p)^m)}{p^2}.$$

Thus $\text{sign}(h'(p))$ equals the sign of the numerator

$$N(p) = mp(1-p)^{m-1} - (1 - (1-p)^m).$$

Let $x = 1 - p \in [0, 1)$. Then $p = 1 - x$ and

$$\begin{aligned} N(p) &= m(1-x)x^{m-1} - (1 - x^m) \\ &= mx^{m-1} - mx^m - 1 + x^m \\ &= mx^{m-1} - 1 - (m-1)x^m. \end{aligned}$$

Define the auxiliary function

$$F(x) = mx^{m-1} - 1 - (m-1)x^m, \quad x \in [0, 1].$$

We claim $F(x) \leq 0$ for all $x \in [0, 1]$, with strict inequality for $x \in (0, 1)$. Compute

$$F'(x) = m(m-1)x^{m-2} - (m-1)mx^{m-1} = m(m-1)x^{m-2}(1-x) \geq 0,$$

so F is weakly increasing on $[0, 1]$. Moreover,

$$F(1) = m - 1 - (m - 1) = 0.$$

Since F is increasing and equals 0 at $x = 1$, it must satisfy $F(x) \leq 0$ for all $x \in [0, 1]$, with $F(x) < 0$ for $x \in (0, 1)$ whenever $m \geq 2$. Therefore $N(p) = F(1 - p) < 0$ for $p \in (0, 1)$ (for $m \geq 2$), implying

$$h'(p) = \frac{N(p)}{p^2} < 0 \quad \text{for } p \in (0, 1),$$

and hence h is strictly decreasing on $(0, 1]$. Consequently,

$$\phi(p) = \frac{qA}{m} h(p)$$

is strictly decreasing on $(0, 1]$ as well.⁶

Boundary values and uniqueness Together with $\phi(0) = qA$, we also have

$$\phi(1) = qA \cdot \frac{1 - (1 - 1)^m}{m \cdot 1} = \frac{qA}{m}.$$

By continuity and strict monotonicity on $(0, 1]$, ϕ maps $[0, 1]$ onto $[\frac{qA}{m}, qA]$ and is one-to-one. Hence:

- If $c \geq \phi(0) = qA$, then $\phi(p) \leq c$ for all $p \in [0, 1]$, so $p^* = 0$.
- If $c \leq \phi(1) = qA/m$, then $\phi(p) \geq c$ for all $p \in [0, 1]$, so $p^* = 1$.
- If $qA/m < c < qA$, then by the intermediate value theorem there exists a unique $p^* \in (0, 1)$ such that $\phi(p^*) = c$, which is exactly Equation (3).

Finally, Equation (4) follows from independence: an incorrect report is challenged iff at least one of the m disputers verifies, so $\alpha^* = 1 - (1 - p^*)^m$.

Appendix A.2 Proof of Lemma 1

Fix a realization of the true state and consider the random votes $\{X_j\}_{j=1}^N$. By assumption, $\{X_j\}$ are independent Bernoulli random variables with

$$X_j \in \{0, 1\}, \quad \Pr(X_j = 1) = \rho.$$

Let weights $\{w_j\}_{j=1}^N$ be nonnegative constants and define

$$W = \sum_{j=1}^N w_j, \quad S = \sum_{j=1}^N w_j X_j.$$

⁶ For $m = 1$, $h(p) \equiv 1$ and $\phi(p) \equiv qA$ is constant, so the equilibrium is trivial: verification is optimal iff $qA \geq c$.

Recall that adjudication is erroneous whenever the incorrect side reaches the activation threshold τW , i.e.,

$$W - S \geq \tau W \iff S \leq (1 - \tau)W,$$

so

$$\varepsilon = \Pr(S \leq (1 - \tau)W). \quad (\text{A2})$$

Define the bounded variables

$$Y_j = w_j X_j.$$

Then $Y_j \in [0, w_j]$ almost surely and $S = \sum_{j=1}^N Y_j$. Moreover,

$$\mathbb{E}[Y_j] = w_j \mathbb{E}[X_j] = w_j \rho \implies \mathbb{E}[S] = \sum_{j=1}^N \mathbb{E}[Y_j] = \rho \sum_{j=1}^N w_j = \rho W.$$

Using (A2),

$$\begin{aligned} \varepsilon &= \Pr(S \leq (1 - \tau)W) \\ &= \Pr(S - \mathbb{E}[S] \leq (1 - \tau)W - \rho W) \\ &= \Pr(S - \mathbb{E}[S] \leq -(\rho + \tau - 1)W). \end{aligned} \quad (\text{A3})$$

Let

$$\delta = \rho + \tau - 1.$$

When $\delta \leq 0$, the threshold $(1 - \tau)W$ is weakly above the mean ρW , and the bound in the statement is vacuous in the usual sense (it does not yield a small upper bound). Henceforth assume $\delta > 0$.

Since Y_1, \dots, Y_N are independent and satisfy $Y_j \in [a_j, b_j]$ with $(a_j, b_j) = (0, w_j)$, Hoeffding's inequality implies that for any $t > 0$,

$$\Pr(S - \mathbb{E}[S] \leq -t) \leq \exp\left(-\frac{2t^2}{\sum_{j=1}^N (b_j - a_j)^2}\right) = \exp\left(-\frac{2t^2}{\sum_{j=1}^N w_j^2}\right). \quad (\text{A4})$$

By (A3), the relevant deviation is $t = \delta W$, which can be plug into (A4), yielding

$$\varepsilon = \Pr(S - \mathbb{E}[S] \leq -\delta W) \leq \exp\left(-\frac{2\delta^2 W^2}{\sum_{j=1}^N w_j^2}\right).$$