Economics Letters 268 (2026) 113176

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Do prediction markets price events or adjudication? Evidence from the disputed Polymarket markets

Guangzhe Wang

School of Accounting and Finance, The Hong Kong Polytechnic University, Hong Kong SAR, China

ARTICLE INFO

JEL classification:
G12
G14
D47
D82

Keywords:
Prediction markets
Decentralized settlement
Oracle
Blockchain governance
Price discovery

ABSTRACT

A prediction-market share is read as a probability about an event. In an oracle-settled market the cash flow is fixed not by the event but by the outcome that the resolution mechanism certifies, so the price can come apart from the event. We test whether prices forecast the event or the adjudication by reading them inside the on-chain settlement window of disputed Polymarket markets, the interval between a bonded proposer's assertion of the outcome and the UMA token vote that fixes the payoff. Inside that window the real-world event is already frozen while the payoff is still pending, so a discount on the asserted-correct claim cannot be ordinary forecasting of an unrealized event. Using 233 disputed markets with a clear audited rule-correct winner, coded by a search-grounded language-model audit formed independently of how the oracle ruled and matched to each market's UMA Optimistic Oracle lifecycle, we find the token vote overturns the audited winner in 12.4 percent of markets, an upper bound inherited from the coding model. Six hours after the first dispute the audited-correct claim already trades at a discount that is 0.83 higher in markets the oracle will overturn, and the discount climbs toward one as the vote approaches while the upheld claim falls toward zero. The pricing pattern survives the choice of coding model, the removal of deadline questions, restriction to the markets that escalate to the vote, and a treatment limited to the hand-verified overturns. The complete UMA voting record shows the deciding votes to be broad and near unanimous, and the dispute-time price predicts the vote share itself. Prices in contested prediction markets are therefore best read not as beliefs about the world but as forecasts of how a decentralized rule will be adjudicated.

1. Introduction

A prediction-market share that pays one dollar if an event occurs is conventionally read as the market's probability that the event occurs (Wolfers and Zitzewitz, 2004; Arrow et al., 2008). That reading presumes that, once the event is known, settlement is frictionless and faithful. In a decentralized market it is neither. Polymarket's payoffs are not certified by a regulated exchange but by the optimistic oracle of UMA,[1] under which anyone may propose an outcome and post a bond, anyone may dispute it, and a contested outcome is decided by a token-weighted vote of UMA holders whose voting power fixes the cash flow. The tradable claim is therefore a claim on the adjudication that this architecture produces, and that adjudication can differ from the outcome that the real-world events imply under the written rule.

When a question closes, a proposer posts a cash bond and asserts the outcome the written rule implies, and the assertion settles unchallenged unless a disputer posts an equal bond within a short window, in which case the payoff is decided by a token-weighted vote of UMA holders. A dispute therefore opens an interval in which an agent with capital at stake has already asserted what happened while the vote that fixes the payoff is still days away, so the event is frozen and only the adjudication is open. That interval is where this paper reads the price.

The distinction was on display on June 1, 2026, when a Polymarket contract with tens of millions of dollars in volume asking whether MicroStrategy would sell any Bitcoin by May 31 sat unresolved even though the firm's own regulatory filing that morning had disclosed a sale within the window.[2] The physical event was settled while the adjudication was not, because the payoff turned on a token vote over the wording of the rule.[3] Polymarket logged more than 1,150 disputed markets in the first half of 2026, and a *Wall Street Journal* analysis

E-mail address: guangzhelele.wang@connect.polyu.hk.

[1] For the mechanism and its dispute and voting procedure, see the UMA documentation at https://docs.uma.xyz.

[2] For contemporaneous reporting on this contract, see https://www.htx.com/news/polymarket-users-cry-foul-after-85-million-bet-on-microstrat-a0hmaJQ2/ and https://bitcoinist.com/polymarket-users-cry-foul-microstrategy-bitcoin/.

[3] This contract is a motivating illustration that postdates and sits outside the estimation sample, yet the pattern it shows is general.

https://doi.org/10.1016/j.econlet.2026.113176
Received 8 June 2026; Received in revised form 19 July 2026; Accepted 28 July 2026
Available online 1 August 2026

found that votes are dominated by a small set of large and sometimes conflicted wallets.[4]

This paper asks whether, in disputed markets, the price forecasts the event or the adjudication. Answering the question requires observing the rule-implied outcome separately from how the oracle ruled, which is the central measurement problem, because disputes arise precisely where the two might diverge. It also requires observing the price at a moment when the event is already settled but the adjudication is not, because otherwise a low price can simply mean the market thought the event unlikely. We address the first problem with a search-grounded large-language-model audit that codes each market's rule-correct winner from the written rule and the real-world events and that is instructed to ignore how any oracle or vote resolved. We address the second problem by reading the price inside the on-chain settlement window. We match every disputed market to its UMA Optimistic Oracle request log on Polygon and recover the exact times at which the outcome was proposed, disputed, and settled. The interval between the dispute and the settlement is a window in which a bonded agent has already asserted the outcome, so the event is frozen, while the token vote has not yet fixed the payoff. A discount on the audited-correct claim inside that window is a price on the adjudication and not a forecast of an unrealized event.

The evidence yields three results. First, the settlement genuinely diverges from the audited outcome. On the 233 disputed markets with an objectively clear audited winner, the settlement departs from the audited label in 12.4 percent of markets under the primary coding model, and a hand audit of every flagged market against primary sources places the genuine rate at 3.4 percent. Second, the market prices the divergence inside the settlement window and not merely before the event. Six hours after the first dispute the audited-correct claim trades at a discount that is 0.83 higher in markets the oracle will overturn, the gap widens exactly at the dispute, and the discount climbs toward one as the vote approaches while the upheld claim falls toward zero. The pricing pattern is robust to a second coding model, to dropping deadline questions, and to controlling for the time remaining to settlement, and it holds without controls, in the markets that escalate to the token vote, and when the treatment is restricted to the hand-verified overturns. Third, the voting layer favors reading the price as aggregation rather than as self dealing. The deciding votes are cast by hundreds of independent stakers and are near unanimous even when they overturn the audited outcome, the market's own proposers and disputers hold a negligible share of the deciding stake, and the dispute-time price predicts the vote share itself.

The contribution is to turn the institutional observation that decentralized settlement can be captured into a statement about secondary-market prices, identified inside the window where the event is frozen and only the adjudication is pending. Where the existing work asks whether the mechanism can fail or whether ambiguous wording raises dispute risk, we use an independent benchmark for the rule-implied outcome and the on-chain settlement clock to show that prices forecast the adjudication rather than the event, we measure how the market prices the divergence as the vote approaches, and we trace the price to a voting layer whose breadth makes aggregation rather than self dealing the natural reading.

2. Related literature

A long literature reads prediction-market prices as probabilities that efficiently aggregate dispersed information (Wolfers and Zitzewitz, 2004; Arrow et al., 2008). Recent work sharpens this view in financial settings. Gómez-Cram et al. (2025) build a high-frequency measure of earnings expectations from prediction-market prices and find it more accurate than analyst forecasts and leading in price discovery, while Gómez-Cram et al. (2026) trace that accuracy to a small informed minority rather than to the crowd. Abid (2026) documents that the same platform can be highly calibrated in aggregate yet host large single-trader price impact and resolution disputes settled by concentrated token holders. This strand treats the price as a signal about the world and, implicitly, treats settlement as faithful. We depart from it by showing that, once a market is disputed, the price is better read as a forecast of the adjudication than as a belief about the event.

A second strand prices prediction-market claims as something other than pure probabilities. Yang (2026) develops a pricing framework in which incomplete markets, selection rules, and risk premia drive a wedge between prices and beliefs. Gebele and Matthes (2026) show that, because collateral in oracle-settled markets stays locked until resolution, a near-certain dollar is a delayed dollar, and they recover a settlement discount that reflects lock-up, liquidity, and residual oracle risk. Our object is related but distinct, because we isolate the component of the price that anticipates the oracle ruling against the audited outcome rather than the component that compensates for delay, and we isolate it in the on-chain window between the dispute and the settlement.

A third strand studies the resolution layer itself. Rezaei Sanjabi (2026) constructs a semantic risk score from contract language, validates it against UMA dispute variance, and shows that linguistically ambiguous markets shed liquidity and converge faster to extreme prices, measuring prices in a fixed window before resolution rather than around the dispute. Ma et al. (2026) model the proposer and disputer game, microfound adjudication as token-weighted voting, and build an empirical adjudication-risk index from UMA voting and staking data, without pricing the risk in the secondary market. These papers ask whether the mechanism can be corrupted and what makes a market risky to resolve. We instead ask whether traders price the possibility that the mechanism rules against the audited outcome, we measure that pricing inside the settlement window, and we observe the resolution electorate directly, vote by vote, rather than through an index.

Closest in setting are two contemporaneous papers on UMA-resolved Polymarket markets. Wen et al. (2026) ask whether language models can reproduce UMA rulings, reporting agreement near 90 percent once a dispute has been raised. Because they treat the UMA ruling as the ground truth and measure how often the model agrees with it, their design cannot detect cases in which the oracle settles against the rule-implied outcome, which is exactly the object we study. Jia et al. (2026) assemble a full-lifecycle dataset spanning market, trade, and oracle layers and document that disputed markets keep trading, so their contribution is data infrastructure and descriptive behavior. Neither uses an independent benchmark for the rule-implied outcome together with the on-chain settlement clock to ask whether prices forecast the adjudication rather than the event. That gap is the space this paper occupies.

3. Data and sample construction

The data are assembled from public Polymarket sources, the UMA Optimistic Oracle subgraph on Polygon, the revealed-vote record of the UMA voting contract on Ethereum, and a language-model audit that is validated by hand. Market metadata, written rules, outcomes, token identifiers, and the dispute trajectory come from the Gamma application programming interface, whose *umaResolutionStatuses* field records each market's resolution path as a sequence of states such as proposed, disputed, and resolved. Hourly prices for both outcome tokens come from the CLOB *prices-history* endpoint. Value at stake measured as open interest and the concentration of token holders come from the Data application programming interface.

A sweep of the Gamma listing returns complete coverage of all closed markets with cumulative volume above roughly fifty thousand

[4] See https://thedefiant.io/news/markets/usd85m-polymarket-dispute-over-strategy-s-may-bitcoin-sale-puts-uma-s-token-voting-oracle-on.

dollars, a total of 131,288 markets.[5] A market is classified as disputed when its resolution sequence contains a disputed state, which holds for 783 markets. We retain the 629 disputed markets that are binary with Yes and No outcomes, resolved, and not settled deterministically by a Chainlink data feed, since the question of event against adjudication is meaningful only where a token vote rather than an automated feed certifies the outcome.

UMA's Optimistic Oracle records, for each price request, the exact block times at which the outcome was proposed, disputed, and settled, together with the proposed and settled prices and the requesting Polymarket adapter. We query the public OptimisticOracleV2 subgraph on Polygon and match each request to its market by the question title recovered from the request's ancillary data, with the decoding mechanics in Appendix. This matches 387 of the 783 disputed markets, of which 374 carry an on-chain dispute time. The markets that do not match are overwhelmingly those that resolved through the managed proposer oracle introduced in late 2025, which is not indexed by this subgraph, so the matched sample is the non-managed Optimistic Oracle era. The matched markets are smaller and less frequently disputed than the unmatched markets, which Table A.1 in the Appendix records and which we return to when bounding the external validity of the estimates. A Polymarket market can generate several oracle requests, because the adapter re-requests after the first dispute, so we keep all requests per market and reconstruct the first proposal time, the first dispute time, and the final settlement time. A second on-chain dispute identifies the markets that escalate to the token vote.

The final layer is the vote itself. Every on-chain dispute is relayed to UMA's data verification mechanism on Ethereum, where stakers commit and then reveal token-weighted votes, and we collect the complete revealed-vote record of the VotingV2 contract over the sample period, 1,619,774 votes, from the public Ethereum event log. The relay preserves each request's ancillary data, hashing it when it is long, so every vote maps back to its Polymarket market exactly, and Appendix details the matching. All 233 clear markets defined below match to their votes. Not every vote fixes a payoff, because the vote on a first dispute frequently certifies only that the question was asked too early, after which the adapter re-requests the price and an unchallenged proposal settles the market. We therefore distinguish the payoff-decisive votes, those whose token-weighted majority is a substantive Yes or No, which exist for 103 of the 233 clear markets, and the majority side of the payoff-decisive vote coincides with the market's settlement in all 103.

The audited rule-correct winner is coded by querying *gemini-2.5-flash* under Google-Search grounding. For each market the model is asked to read the written rule, search for the real-world events around the market's end date, and report the outcome that the rule should yield given those events. The prompt instructs the model not to consider how any oracle, committee, or token vote ruled, so that the resulting label measures the rule-implied outcome and not the adjudication. Each market is queried three times. A market is treated as having an objectively clear audited winner when the three queries agree and the mean confidence is high, which holds for 233 of the on-chain matched binary markets with a valid Yes or No label. The clear sample of 233 markets is the basis for the analysis, because a divergence between the audited outcome and the adjudication is interpretable only where the audited outcome is itself well defined. The label is a measurement instrument and not an oracle of truth, so we refer throughout to the audited rule-correct winner rather than to the facts, we repeat the entire audit with *gemini-3.5-flash* and carry its label as a robustness check, and we read the divergence as the settlement departing from the audited label rather than from the truth.

[5] The endpoint caps offset pagination, so the sweep descends through bands of cumulative volume with a volume cursor. Appendix records the mechanics.

4. Empirical design and results

4.1. Variable construction and design

For each market we compare three binary objects. The settlement r_i is the UMA payoff winner, the token that pays one dollar at resolution, read from the resolved Gamma outcome and cross-checked against the on-chain settled price, which agree for 168 of the 170 clear markets that carry a numeric on-chain price. The market favorite at horizon h is the higher-priced of the two outcome tokens h hours before an event time, which we read as the market's adjudication forecast. The audited winner w_i is the rule-correct winner from the language-model audit. The realized divergence between the audited outcome and the settlement is captured by the indicator $capture_i = \mathbf{1}\{w_i \neq r_i\}$, equal to one when the oracle overturns the audited winner. The discount on the audited-correct claim is $1 - P^{w}_{i,t}$, where $P^{w}_{i,t}$ is the price of the audited-correct token. Controls comprise the ambiguity score equal to one minus mean audit confidence, log open interest, log cumulative volume, the share of holdings held by the five largest holders, and category fixed effects assigned from the question text by a transparent keyword map. The voter-layer variables, comprising the number of revealed voters, the largest voter's stake share, the concentration of the deciding vote, and its winning margin, are constructed in Section 4.4 and defined in Table A.2. All standard errors are clustered by Polymarket event, because sibling markets in a series are not independent. The 233 clear markets fall into 188 event clusters.

We anchor on the on-chain lifecycle rather than on the finalization stamp. The proposal asserts the outcome with a bond, the dispute opens the contest over the payoff, and the settlement fixes it. Between the dispute and the settlement the real-world event is already frozen, because a bonded agent has asserted it, while the token vote has not yet decided the payoff. The tradeable settlement window runs from the dispute to the finalization stamp, at which trading on the outcome tokens ceases, and lasts about 2.4 days at the median, with the full lifecycle timing recorded in Appendix. We measure the discount on the audited-correct claim at the dispute, six hours after the dispute, and as the mean over the tradeable settlement window, and we trace it in event time on both sides of the dispute and the settlement. The figure in the appendix sketches this timeline.

4.2. The settlement diverges from the audited outcome

Table 1 summarizes the sample and the discount on the audited-correct claim across the lifecycle. The headline divergence is a pair of bounds. The token vote settles against the audited label in 12.4 percent of the clear disputed markets, a noisy upper bound because the audited label may carry the classification error.[6] The hand audit of every flagged market places the genuine divergence at 3.4 percent, the lower bound. Settlement therefore departs from the audited outcome in at most one disputed market in eight and in about one in thirty on the hand-verified count. The lower panel shows the discount on the audited-correct claim by lifecycle stage. One day before the proposal the discount is 0.29 in markets the oracle will uphold and 0.64 in markets it will overturn, a gap that already exists before the event is formally asserted and that a finalization-anchored design would also see. At the first dispute the gap widens to 0.71, six hours after the

[6] We hand-checked all 29 flagged markets and a stratified sample of 40 upheld markets, researching each against primary sources such as regulatory filings, official transcripts, and agency releases. The check confirms a genuine divergence in 8 of the 29 flagged markets, attributes 20 to label error, and leaves 1 unresolved as genuinely ambiguous, while reversing none of the 40 upheld markets. The hand-verified divergence rate is therefore 3.4 percent, with a 95 percent Wilson interval from 1.8 to 6.6 percent, the precision of the model's overturn flag is 28 percent, and no false negatives were found among the upheld controls.

Table 1
Audited outcome, price, and settlement in clear markets.

Panel A. Sample and divergence			
Clear audited disputed markets, on-chain matched	233		
Event clusters	188		
Oracle overturns the audited winner, upper bound	0.124		
Hand-verified overturns, lower bound	0.034		
Reaches the token vote, two or more on-chain disputes	0.335		
Panel B. Discount on the audited-correct claim by lifecycle stage			
Stage	*N*	Upheld	Overturned
One day before the proposal	189	0.287	0.637
At the first dispute	205	0.102	0.809
Six hours after the dispute	205	0.052	0.894
Tradeable settlement window, mean	232	0.039	0.937
At the close	233	0.001	0.998
Seven days before finalization, prior design	187	0.274	0.757

The settlement is the UMA payoff winner and the audited outcome is the *gemini-2.5-flash* rule-correct winner. Capture equals one when the two differ. The hand-verified rate counts the overturns confirmed by the hand audit described in the text. The discount is one minus the price of the audited-correct token. The price at the first dispute is taken at or just before the dispute, about half an hour before at the median, and the six-hour measure is the cleanest strictly post-dispute reading.

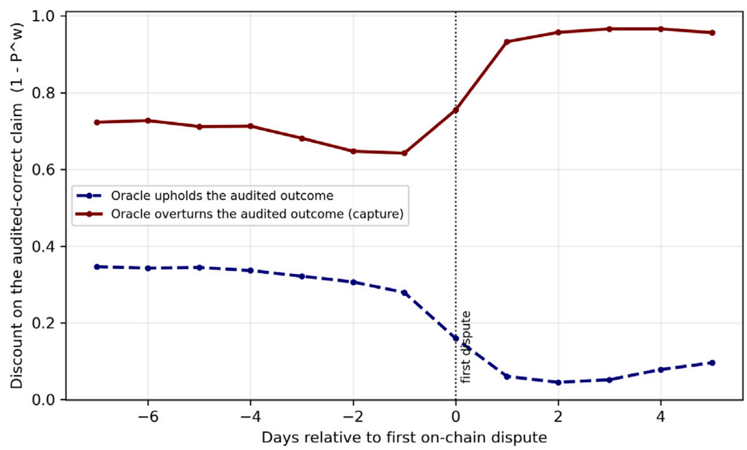

Fig. 1. Discount on the audited-correct claim, $1 - P^w$, by days relative to the first on-chain dispute, separately for markets in which the oracle overturns the audited outcome and markets in which it upholds it, on the clear sample. The gap widens at the dispute, where the event is already asserted and only the token vote is pending.

dispute it reaches 0.84, and across the tradeable settlement window it reaches 0.90, while at the close it is essentially one. The widening from the pre-proposal stage into the settlement window is the part of the pattern that a forecast of an unrealized event cannot generate, because by the dispute the event has been asserted and only the vote remains.

4.3. The market prices the divergence inside the settlement window

The central result is that the market prices the divergence between the audited outcome and the settlement inside the window where the event is frozen and only the vote is pending. Fig. 1 plots the discount on the audited-correct claim in event time relative to the first dispute, separately for markets the oracle ultimately overturns and markets it upholds. The two paths are separated before the dispute and then move sharply apart at the dispute, the overturned claim climbing toward one and the upheld claim falling toward zero, which is the contest over the payoff being priced. Fig. 2 plots the same discount in event time relative to the settlement and shows the paths converging to one and zero as the token vote arrives.

Panel A of Table 2 makes the magnitude precise. Regressed on realized capture with controls and category fixed effects and standard errors clustered by event, the discount on the audited-correct claim measured at the dispute is 0.724 higher in markets the oracle will overturn, with a t value of 12.6, and the strictly post-dispute discount six hours later is 0.832 higher, with a t value of 19.8. The coefficient grows monotonically from 0.309 one day before the proposal to 0.997 at the close, and the finalization-anchored discount measured seven days before resolution, which the prior design used, is 0.443 on the same markets. The dispute-window estimate is therefore both larger and better identified than the finalization-anchored estimate. Read the other way, in Panel B the dispute-window discount predicts capture with a coefficient of 0.715 and a t value of 8.4. Panel C shows that on-chain escalation to the token vote does not by itself predict capture or the discount once the price is observed, so the contest over the price, and not the count of disputes, carries the information.

The pricing pattern is robust to the coding model, the sample, and the mechanics, which Table 3 records. Recoding both the audited winner and the discount with *gemini-3.5-flash*, which places the overturn level near 7 percent rather than 12, leaves the dispute-window discount on capture at 0.703, and the coefficient is 0.627 when capture is restricted to the markets both models overturn, 0.776 when deadline questions whose event might never occur are dropped, 0.799 in the markets that escalate to the token vote, and 0.706 without controls. Restricting the treatment to the eight overturns the hand audit certifies as genuine and dropping the flagged markets it attributes to label error yields 0.636 at the dispute and 0.826 six hours after it, the latter indistinguishable from the baseline 0.832, so the pricing result does not live in the mislabeled markets. Nor is it an artifact of proximity to the settlement, because adding the hours remaining to the settlement leaves the capture coefficient at 0.721 while the time control itself is

Table 2
The settlement diverges from the audited outcome and prices anticipate it.

Stage	Coefficient on capture	t	N
Panel A. Discount on the audited-correct claim regressed on capture			
One day before the proposal	0.309***	4.15	189
At the first dispute	0.724***	12.64	204
Six hours after the dispute	0.832***	19.80	204
Tradeable settlement window mean	0.903***	52.10	231
At the close	0.997***	–	232
Seven days before finalization	0.443***	5.85	187
Panel B. Direction, capture regressed on the discount			
	Coefficient	t	N
Discount at the dispute	0.715***	8.39	204
Discount over the settlement window	0.975***	25.02	231
Panel C. On-chain escalation			
	Coefficient	t	N
Capture on escalation to the vote	0.023	0.44	232
Dispute discount on escalation to the vote	0.053	0.89	204

Ordinary least squares with standard errors clustered by Polymarket event. Controls comprise the ambiguity score, log open interest, log cumulative volume, the top-five holder share, and category fixed effects. The discount on the audited-correct claim is $1 - P^w$. The close coefficient is mechanical because the closing price equals the settlement. Significance is denoted * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

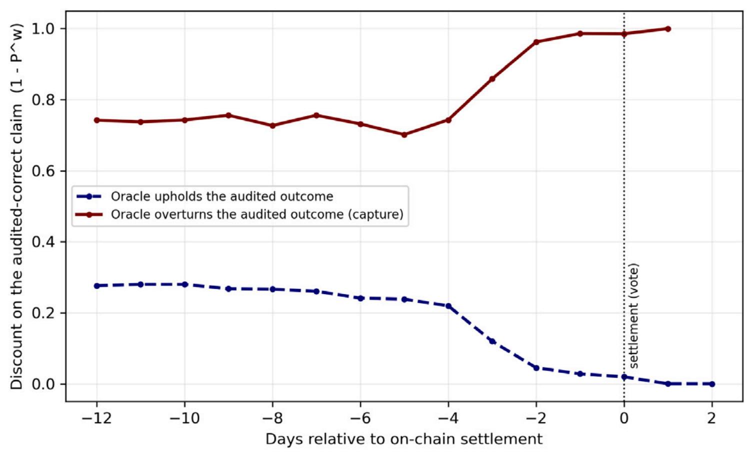

Fig. 2. Discount on the audited-correct claim, $1 - P^w$, by days relative to the on-chain settlement, separately for markets the oracle overturns and markets it upholds, on the clear sample. The paths converge to one and zero as the token vote settles the payoff.

Table 3
Robustness of the dispute-window discount on capture.

Specification	Coefficient	t	N
Cross-model, *gemini-3.5-flash* audit and capture	0.703***	9.94	234
Capture under both coding models	0.627***	6.37	192
Drop deadline questions	0.776***	9.46	144
Escalated to the token vote	0.799***	14.99	67
Bivariate, no controls	0.706***	13.08	205
Hand-verified overturns as treatment	0.636***	5.71	185
Controlling for hours to the settlement	0.721***	12.57	204

The dependent variable is the discount on the audited-correct claim at the dispute, except in the first row, where it is measured on the second model's audited-correct token. Standard errors are clustered by event. The cross-model overturn level is 6.7 percent against 12.4 percent under the primary model. The hand-verified row keeps as treated the eight overturns the hand audit confirms as genuine and drops the flagged markets it attributes to label error or leaves unresolved, the upheld markets serving as controls. Significance is denoted * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

insignificant. The exact frequency of overturning is a property of the coding instrument, but the finding that prices forecast the adjudication rather than the audited outcome is a property of the markets.

4.4. *The voter layer, aggregation against self dealing*

A price that anticipates the vote admits two readings. The market may aggregate dispersed information about how the electorate will read the rule, or the voters themselves may set the price, in which case the forecast is self dealing rather than aggregation. For each clear market we take its final substantive vote, aggregate each voter's revealed stake, and compute the number of revealed voters, the largest voter's stake share, the stake-weighted concentration measured as a Herfindahl index, the winning margin as the stake share of the settled side among substantive ballots, and the stake share cast by the market's own on-chain proposers and disputers.

Table 4 reports the result, which is uniformly hostile to the self-dealing reading. The deciding votes are broad. The median matched market's final vote has about 350 distinct revealed voters, the largest voter carries 11 percent of the stake, and the concentration index is 0.047, the equivalent of about twenty equally sized voters. Overturned markets are not decided by more concentrated votes than upheld markets, the dispute-window discount on capture does not strengthen with vote concentration, and splitting the sample at the median concentration leaves the discount on capture at 0.746 in the more dispersed half against 0.664 in the more concentrated half. Direct self dealing is negligible. A market's own proposer or disputer appears among its voters in 27 of the 233 markets, and where present such voters hold less than one hundredth of one percent of the deciding stake. The deciding votes are also near unanimous. Every payoff-decisive vote carried at least 90 percent of the substantive stake, the median margin is one, and the overturned markets are settled as unanimously as the upheld

Table 4
The deciding votes are broad, near unanimous, and forecast by the price.

Panel A. The deciding votes			
Median revealed voters per market	355		
Median top-voter stake share	0.108		
Median vote concentration, Herfindahl	0.047		
Payoff-decisive votes, majority equals the settlement	103 of 103		
Winning margin at least 0.90, median margin	all, 1.00		
Markets whose own proposer or disputer votes, their stake share	27 of 233, <0.0001		
Panel B. Concentration, self dealing, and the price			
	Coefficient	t	N
Capture on vote concentration	5.76	0.96	232
Capture on top-voter share	2.43*	1.82	232
Dispute discount on capture × concentration	−11.15	−0.79	204
Dispute discount on capture, dispersed half	0.746***	9.20	107
Dispute discount on capture, concentrated half	0.664***	9.83	98
Vote share against the audited winner on the dispute discount	0.825***	7.43	99

Panel A describes the final substantive vote of each matched clear market from the complete VotingV2 revealed-vote record. Panel B reports ordinary least squares with standard errors clustered by event and the same controls as Table 2 in the first three rows. The concentration is the stake Herfindahl of the final vote and is demeaned in the interaction row. The top-voter coefficient is marginal and reverses sign without controls. The halves split the matched sample at the median concentration, with the discount regressed on capture alone. The last row regresses the stake share cast against the audited winner on the dispute-time discount in the payoff-decisive markets. Significance is denoted * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ones, so an overturn is a broad electorate certifying a reading of the rule that departs from the audited one and not a marginal coalition flipping a contested tally. Finally, the price at the dispute predicts the vote share itself, since the stake share ultimately cast against the audited winner rises 0.825 with the dispute-time discount, with a t value of 7.4, before the voting round has begun. Taken together, the voter layer says that the adjudication forecast in the price is a forecast of a dispersed electorate's reading of the rule, which is aggregation, and not the electorate quoting its own decision to itself. Two limits of this evidence are worth stating. Vote concentration varies little in this era of the mechanism, so the tests compare degrees of breadth rather than breadth against capture by a single wallet, and Polymarket traders hold outcome tokens through proxy wallets on Polygon while voters stake on Ethereum, so we cannot match holders to voters at the identity level.

4.5. Interpretation

Taken together the results say that in contested prediction markets the price is best read not as a belief about the world but as a forecast of how a decentralized rule will be adjudicated, and that traders form this forecast inside the settlement window, after a bonded agent has asserted the outcome and before the token vote fixes the payoff. The design answers the objection that a low price merely reflects an unrealized event, because the discount is read where the event is already frozen and only the adjudication is pending, and because it widens precisely at the dispute. Three features of the design bound the reading. The benchmark for the audited outcome is coded by a language model rather than by a court of facts, its overturn level moves with the model, and the hand audit shows the model's overturn flag to be right about one time in four, so we bound the divergence between the hand-verified rate and the model-defined rate, read it as the settlement departing from an audited label and not from the truth, and rest the pricing claims on treatments that survive the hand audit. The eight hand-verified overturns are markets in which primary sources contradict the settled outcome, which is the cleanest evidence that the divergence is adjudication and not mislabeling. The voter layer sharpens what such a divergence is, because the deciding votes are near unanimous even in the overturned markets, so a capture is a broad electorate consensually certifying a different reading of the rule and the risk the market prices is the risk of how the rule will be read. The matched sample is the non-managed Optimistic Oracle era and is smaller and less disputed than the unmatched markets, so the overturn frequency should not be extrapolated to the highest-stakes contests, even though the pricing pattern is what the theory predicts.

5. Conclusion

Decentralized settlement turns the question of who certifies the outcome into a tradable risk. Matching disputed Polymarket markets to their UMA Optimistic Oracle lifecycle and using an independent audit of the rule-implied outcome, we show that prices inside the settlement window forecast the oracle's adjudication rather than the audited outcome, that the two diverge in up to one disputed market in eight under the primary coding model and in about one in thirty once every flagged case is checked by hand, and that the market prices the divergence in the window between the dispute and the vote, where the event is frozen and only the adjudication is pending. The pattern is sharper and better identified than at a fixed horizon before finalization, it holds under a second coding model and on the hand-verified overturns, and it widens at the dispute. The complete voting record adds that the adjudication being priced is the consensus reading of a broad and near unanimous electorate rather than the private choice of a few conflicted wallets, and that the dispute-time price predicts the vote share itself. For the asset pricing of state-contingent claims the result identifies a measurable wedge between an event and its adjudication, and for market design it quantifies the price of a settlement architecture in which capital can influence the certification of the outcome.

Appendix. Design, lifecycle mechanics, and variables

Fig. A.1 sketches the identification. A bonded proposer asserts the outcome after the event resolves, a disputer challenges it, the question may escalate to a token vote, and the oracle settles the payoff. Between the proposal and the settlement the real-world event is frozen while the payoff is pending, and the discount on the audited-correct claim is read inside that window. Table A.1 records the selection into the on-chain matched sample and Table A.2 lists every variable used in the empirical section together with its source and construction.

The lifecycle timing is as follows. The proposal precedes the first dispute by about twenty minutes at the median. The first dispute precedes the on-chain settlement by about 3.7 days at the median. Trading on the outcome tokens ceases at the finalization stamp, which falls a median half day and a mean 2.3 days before the on-chain settlement, so the tradeable settlement window runs from the dispute to the finalization stamp and lasts about 2.4 days at the median. A Polymarket market can generate several oracle requests because the adapter re-requests the price after a first dispute, so the request chain is collapsed to the first proposal time, the first dispute time, and the

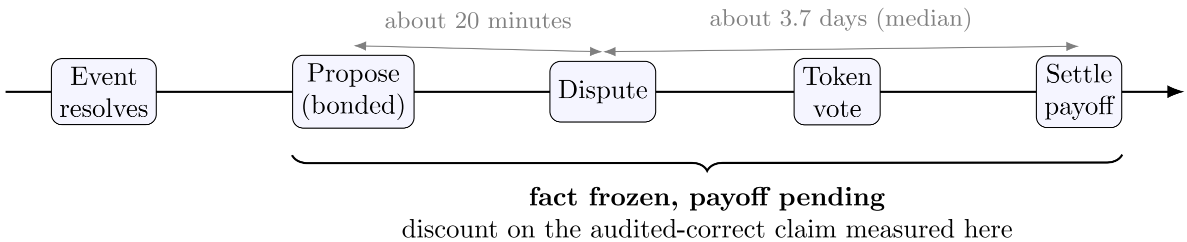

Fig. A.1. The on-chain settlement window. The discount on the audited-correct claim is measured between the dispute and the settlement, where a bonded proposal has frozen the event but the token vote has not yet fixed the payoff. Trading on the outcome tokens ceases at the finalization stamp, a median half day before the on-chain settlement, so the tradeable window runs from the dispute to the finalization stamp.

Table A.1
Selection, on-chain matched against unmatched clean disputed markets.

	Matched	Unmatched
Number of markets	327	302
Median volume, dollars	373,621	953,148
Two or more disputes, share	0.321	0.543
Resolved after September 5, 2025, share	0.566	0.967

The sample is the 629 clean disputed markets that are binary, resolved, and not Chainlink settled. Matched markets are those linked to an on-chain OptimisticOracleV2 request. The unmatched markets are larger and more frequently disputed and resolve overwhelmingly after the managed proposer oracle began on September 5, 2025, which the OptimisticOracleV2 subgraph does not index.

Table A.2
Variables in the empirical analysis.

Variable	Source	Construction
Audited winner w_i	Search-grounded *gemini-2.5-flash* audit	Majority of three queries for the rule-correct Yes or No outcome, formed independently of the oracle ruling and validated by the hand audit described in the text.
Settlement r_i	Gamma *outcomePrices*, cross-checked on chain	The resolved outcome that pays one dollar. Agrees with the on-chain settled price in 168 of 170 clear markets.
Capture	w_i and r_i	$\mathbf{1}\{w_i \neq r_i\}$, one when the token vote overturns the audited winner.
Capture, second model	*gemini-3.5-flash* audit and r_i	As above with the second model's audited winner.
Hand-verified capture	Hand audit	Capture restricted to the overturns the hand audit confirms as genuine divergences.
$P^{w}_{i,t}$	CLOB *prices-history*, hourly	Price of the audited-correct token at event time t, nearest hourly point at or before t.
Discount at the dispute	CLOB price, on-chain dispute time	$1 - P^{w}$ at or just before the first on-chain dispute.
Discount, dispute +6 h	CLOB price, on-chain dispute time	$1 - P^{w}$ six hours after the first dispute, strictly post-dispute.
Discount, window mean	CLOB price, on-chain times	Mean of $1 - P^{w}$ from the dispute to the last trade.
Discount, pre-proposal	CLOB price, on-chain proposal time	$1 - P^{w}$ one day before the first proposal.
Discount, prior horizon	CLOB price, finalization stamp	$1 - P^{w}$ seven days before the finalization stamp, the prior design's measure.
$T_{\text{propose}}, T_{\text{dispute}}, T_{\text{settle}}$	UMA OptimisticOracleV2 subgraph	Block times of the first proposal, first dispute, and final settlement, matched by decoded ancillary title.
Escalated to the vote	UMA subgraph	One when a market records two or more on-chain disputes.
Multi-dispute	Gamma *umaResolutionStatuses*	One when the resolution sequence contains two or more disputed states.
Revealed vote	VotingV2 *VoteRevealed* log, Ethereum	One record per voter, request, and round, carrying the voter, the vote, and the staked tokens, matched to markets by the relayed ancillary data or its hash.
Number of voters	VotingV2 log	Distinct revealed voters in the market's final substantive vote.
Top-voter share	VotingV2 log	Largest voter's share of revealed stake in the final substantive vote.
Vote concentration	VotingV2 log	Herfindahl index of revealed stake shares in the final substantive vote.
Payoff-decisive vote	VotingV2 log	One when the final vote's stake-weighted majority is a substantive Yes or No rather than the too-early value.
Winning margin	VotingV2 log	Stake share of the settled side among substantive ballots of the payoff-decisive vote.
Vote share against audited	VotingV2 log	Stake share cast against the audited winner among substantive ballots.

(*continued on next page*)

Table A.2 (*continued*).

Self-dealing share	VotingV2 log and OO subgraph	Stake share of the market's own proposers and disputers among its revealed voters.
Ambiguity	Audit confidence	One minus the mean confidence over the three audit queries.
Log open interest	Data API */oi*	$\log(1 + \text{open interest})$.
Top-five holder share	Data API/*holders*	Share of token holdings held by the five largest holders.
Log volume	Gamma *volumeNum*	$\log(1 + \text{cumulative volume})$.
Category	Question text	Seven groups assigned by a transparent keyword map, entered as fixed effects.
Event cluster	Gamma event identifier	The Polymarket event the market belongs to, used to cluster standard errors.
Deadline question	Question text	One when the question contains a by-date or before-date deadline.

final settlement time, and a second on-chain dispute marks escalation to the token vote.

The universe sweep, the oracle matching, and the vote matching work as follows. The Gamma listing caps offset pagination, so the sweep walks a volume cursor down through bands of cumulative volume until the listing is exhausted. Each OptimisticOracleV2 request carries the market's question inside its ancillary data, which we decode from hex and match to the disputed markets by normalized title. Each on-chain dispute is relayed from Polygon to the data verification mechanism on Ethereum, and the relay stamps the request with the identity of its requester and, when the payload is long, replaces the ancillary data with its hash, so a revealed vote is matched to its market either by the stamped full text or by recomputing the hash from the Polygon request. A vote on a re-requested question frequently resolves to the protocol's too-early value, which allocates the dispute bond without fixing the market payoff, and such votes are excluded from the payoff-decisive set used for the margin and vote-share variables.

Data availability

Data will be made available on request.

References

Abid, M.N., 2026. The Polymarket Paradox: Manipulation, Whale Concentration, and Predictive Accuracy in the World's Largest Prediction Market. Working paper, April 28, 2026. Available at SSRN: https://ssrn.com/abstract=6670638 or http://dx.doi.org/10.2139/ssrn.6670638.

Arrow, K.J., et al., 2008. The promise of prediction markets. Science 320 (5878), 877–878. http://dx.doi.org/10.1126/science.1157679.

Gebele, J., Matthes, F., 2026. When certainty is not worth it: Capital lock-up and settlement discounting in prediction markets. http://dx.doi.org/10.48550/arXiv.2605.31431, arXiv preprint arXiv:2605.31431 [cs.CE], May 29, 2026.

Gómez-Cram, R., Guo, Y., Jensen, T.I., Kung, H., 2025. Financial Prediction Markets: A New Measure of Earnings Expectations. Working paper, December 10, 2025. Available at SSRN: https://ssrn.com/abstract=5933475 or http://dx.doi.org/10.2139/ssrn.5933475.

Gómez-Cram, R., Guo, Y., Jensen, T.I., Kung, H., 2026. Prediction Market Accuracy: Crowd Wisdom or Informed Minority? Working paper, April 20, 2026. Available at SSRN: https://ssrn.com/abstract=6617059 or http://dx.doi.org/10.2139/ssrn.6617059.

Jia, H., Zhou, L., Zhang, W., Cong, L.W., Li, S., Sun, S., 2026. Unlocking the forecasting economy: A suite of datasets for the full lifecycle of prediction market: [Experiments & analysis]. http://dx.doi.org/10.48550/arXiv.2604.20421, arXiv preprint arXiv:2604.20421 [cs.LG], April 22, 2026.

Ma, M., Bao, T., Wen, Y., 2026. Optimistic Resolution Mechanisms for Prediction Markets: Incentives, Equilibria, and Design Limits. figshare, http://dx.doi.org/10.6084/m9.figshare.31429679.v1, Preprint, February 27, 2026.

Rezaei Sanjabi, R., 2026. The Semantic Risk Premium. Working paper, January 20, 2026. Available at SSRN: https://ssrn.com/abstract=6126806 or http://dx.doi.org/10.2139/ssrn.6126806.

Wen, J., Zhou, J., Huang, J., 2026. Can LLMs help decentralized dispute arbitration? A case study of UMA-resolved markets on polymarket. http://dx.doi.org/10.48550/arXiv.2604.15674, arXiv preprint arXiv:2604.15674 [cs.CY], April 17, 2026.

Wolfers, J., Zitzewitz, E., 2004. Prediction markets. J. Econ. Perspect. 18 (2), 107–126. http://dx.doi.org/10.1257/0895330041371321.

Yang, Y., 2026. Pricing Prediction Markets: Incomplete Markets, Selection Rules, and Risk Premia. Working paper, March 24, 2026. Available at SSRN: https://ssrn.com/abstract=6468338 or http://dx.doi.org/10.2139/ssrn.6468338.