Prediction Markets Are Efficient at Rest and Dislocated in Motion

Benjamin M. Ampel
Georgia State University
bampel@gsu.edu (ORCID 0000-0003-0603-0270)

Joseph R. Buckman
Georgia State University
jbuckman@gsu.edu (ORCID 0000-0001-6883-6072)

Abstract

Prediction markets are often described as real-time probability sensors, but two venues like Kalshi and Polymarket may price the same event differently because of distinct user bases and trading volumes, raising questions about fee-adjusted cross-venue price gaps. We analyzed 5,038 moneyline contracts on 2,519 sports games across NBA, MLB, NFL, and NHL from September 2025 to May 2026 to study calibration, price discovery, and cross-venue dislocation. Formal equivalence tests show the two venues are equivalent, not merely indistinguishable, in calibration and discrimination both before and during play. Measured at the same instant before tip-off from executed trades, their closing prices are nearly identical, and neither price improves predictions once the other is known. During live play the venues reprice together with no detectable leader, yet their trade-inferred entry prices are transiently dislocated amid volatility. These dislocations are upper bounds on what a trader could capture. Applying the same trade-print inference within a single venue produces impossible bid-above-ask crossings in 4% of live observations. The dislocations also shrink sharply under the venues' actual fee schedules, cross-venue frictions, and execution latency. Kalshi's within-event overround, modestly above Polymarket's, is statistically robust but economically negligible. Cross-venue parity in forecast quality is state-invariant. Execution-level agreement is not.

Keywords: prediction markets, market efficiency, calibration, price dislocation, price discovery

1. Introduction

Two prediction-market venues, Polymarket (launched 2020) and Kalshi (the first U.S. CFTC-regulated event-contract exchange, 2021), rose to prominence in 2024 by offering real-time trading on outcomes such as sports, politics, and cryptocurrency prices. Prediction markets aggregate dispersed information into a single tradeable probability (Manski, 2006; Wolfers & Zitzewitz, 2004), and belong to the broader class of crowd-aggregation mechanisms studied in information systems (IS) (Du et al., 2017). Researchers, journalists, and decision-makers increasingly treat those probabilities as ground truth (Arrow et al., 2008; Ng et al., 2026).

Shortly before tip-off in an NBA playoff game we captured live, Kalshi and Polymarket priced the same outcome at 58% and 60%. Identical claims should converge across venues once fees, spreads, and execution frictions are accounted for. Each game can also be traded live, which may open further gaps as events shift betting behavior.

Kalshi (USD) alongside Polymarket (USDC) creates a natural laboratory for a basic but unanswered question. When two venues price identical events, do they agree, and can a disciplined trader profit when they do not? Answering this requires separating whether prices are accurate in aggregate from whether cross-venue disagreements are executable in real time. The intuition that miscalibration creates arbitrage conflates these claims. A venue can be poorly calibrated without offering any exploitable trade. Two venues can also momentarily disagree while both remain well calibrated. Identical events across venues can violate the law of one price while arbitrage stays capital-intensive or unenforceable (Gebele & Matthes, 2026).

Recent research has begun to examine prediction markets across venues. Ng et al. (2026) analyze common election contracts on Polymarket, Kalshi, PredictIt, and Robinhood, finding price disparities and price-discovery leadership by Polymarket. Gebele and Matthes (2026) document persistent execution-aware deviations from parity across semantically aligned contracts, and Le (2026) shows calibration varies by domain, horizon, and trade size. These studies establish that prediction-market prices need not be interchangeable, but do not determine whether cross-platform relationships change within a single rapidly resolving event as it moves from rest into live play.

To address this gap, we assembled a matched panel that pairs the same game across both platforms at the contract level. We use this dataset to evaluate three questions. First, are the two prediction markets equally well calibrated, and does one price carry information the other lacks (RQ1)? Second, is there executable cross-venue arbitrage in the pre-game line (accounting for trade exe-

cution fees) (RQ2)? Third, does this relationship change during live play (RQ3)?

Our findings suggest these markets are efficient at rest but dislocated in motion. By efficient at rest we mean joint pre-game convergence after fees, covering equivalent calibration, redundant predictive content, and negligible trade-based locks. Before tip-off the two venues' closing prices are nearly identical once measured at the same instant from executed trades, neither price adds detectable signal beyond the other, and their small disagreements carry no predictive information. Consistent with this resting-state efficiency, net-of-fee (after both venues' trading costs) cross-venue locks are rare and economically trivial before tip-off, appearing in only 1.9% of tested games and at a median observed transaction size of six contracts.

During the game, the two venues continue to reprice simultaneously, and we detect no price-discovery leader. However, brief net-of-fee dislocations in trade-inferred entry prices appear in the vast majority of games amid volatility. We treat these as an upper bound on what a trader could capture. Section 4 quantifies why. A dislocation is net-of-fee when:

$$e_{\text{net}} = \max(b_K - a_P,\, b_P - a_K) - \phi_K - g_P > 0 \quad (1)$$

where a and b denote inferred ask- and bid-side entry prices from observed trades, K and P index Kalshi and Polymarket, $\phi_K = \lceil 0.07\,\bar{p}_K(1-\bar{p}_K) \rceil_{0.01}$ is the Kalshi fee per contract evaluated at the Kalshi midpoint $\bar{p}_K = (a_K + b_K)/2$ and rounded up to the cent, largest near $\bar{p}_K = 0.5$ and vanishing at the extremes, and g_P is the Polymarket cost under the fee schedule that market carried on that date (Section 3.2). Kalshi carries a small positive within-event overround, about one percentage point overall, modestly above Polymarket's, and this difference survives multiple-comparison control.

This paper makes three contributions. First, it extends research on cross-platform prediction markets to sports by creating a matched panel of identical moneyline events traded simultaneously on Kalshi and Polymarket, comparing resting and live markets. Second, it shows that signal quality is state-dependent, separating forecast quality from executable-liquidity quality. While both platforms agree on prices and reprice together, in-game volatility briefly dislocates their trade-inferred entry conditions (consistent with limits-to-arbitrage logic) (Shleifer & Vishny, 1997). Finally, the study reconstructs entry conditions from trade data, controlling for timing, state, venue- and date-specific fees, and multiple comparisons. It then bounds how much of the measured dislocation could be executable, benchmarking it against the same statistic computed within a single venue, where arbitrage is impossible by construction.

2. Theoretical background

Two decades of literature establish that prediction-market prices can be read as probabilities (Arrow et al., 2008; Tziralis & Tatsiopoulos, 2007; Wolfers & Zitzewitz, 2004): across election, sports, and corporate-event markets, contracts trading at price p resolve favorably close to a fraction p, which lets organizations treat a market price as a calibrated forecast.

Calibration is distinct from informativeness: a market can be calibrated yet uninformative, or sharp yet biased (Page & Clemen, 2013). Calibration is often measured with the expected calibration error (ECE) in Equation 2:

$$\text{ECE} = \sum_{m=1}^{M} \frac{n_m}{n} \left| \bar{y}_m - \bar{p}_m \right|, \quad (2)$$

where the price range is partitioned into $M = 10$ equal-width bins, n_m is the number of contracts whose price falls in bin m (with $n = \sum_m n_m$), $\bar{p}_m$ is the mean predicted probability in the bin, and $\bar{y}_m$ is the realized win frequency (Guo et al., 2017). The literature also recommends assessing discrimination, a price's ability to rank winners over losers and an aspect of its refinement beyond calibration (DeGroot & Fienberg, 1983), and information content, whether one price subsumes another (Fair & Shiller, 1990). Holding these three properties is vital. Two venues can match on aggregate calibration while differing on bias or on which price is the sufficient statistic for the outcome.

Betting and prediction markets also often exhibit a favorite-longshot bias, longshots being overbet relative to true odds (Snowberg & Wolfers, 2010). Expressed as a calibration slope, regressing outcomes on the logit of price yields a slope below 1 when prices are systematically too extreme, and a cross-venue difference in that slope would reveal differing trader populations.

Whether a trader can profit from such biases requires separate analysis. The efficient markets hypothesis predicts that prices for the same asset cannot diverge by more than transaction costs (Fama, 1970), but real arbitrage is bounded by fees, capital lockup, execution and timing risk, and segmented access (Shleifer & Vishny, 1997). Prediction markets add their own frictions: Kalshi and Polymarket use different currencies, settlement rails, and legal access, which may leave genuine gaps unexploitable. Earlier work documents cross-venue arbitrage on electronic markets (Franck et al., 2013; Oliven & Rietz, 2004), and recent studies measure execution-adjusted divergence directly (Gebele & Matthes, 2026; Saguillo

et al., 2025). Neither determines whether the pre-game-to-live transition dislocates cross-venue entry conditions.

When the same contract trades on multiple venues, one venue can lead price discovery, quantified by Hasbrouck's information share and by Granger causality (Granger, 1969; Hasbrouck, 1995). In fast-moving sports markets, public information is incorporated within seconds, so transient cross-venue gaps can open before both venues finish adjusting (Croxson & Reade, 2014), and apparent mispricings shrink once realistic execution is imposed (Franck et al., 2013). This motivates our distinction between the pre-game resting state and the in-game moving state.

Beyond what a trader can capture, a market price is a signal others read and act on. In IS research, prediction markets are information-aggregation mechanisms and decision-support systems whose output is consumed by downstream users and algorithms (Berg & Rietz, 2003; Hanson, 2003). Treating the price as a data product makes its cross-platform reliability, latency, and exploitability first-order design concerns.

These strands converge on one idea. The quality of a market-derived probability separates into two layers. The first is *forecast quality*, the accuracy of the consensus price (calibration, discrimination, information content). The second is *executable-liquidity quality*, which governs whether a disagreement can be traded (spread, depth, executability). The layers respond differently to state. Cross-venue parity in forecast quality is a property of the aggregated belief and is robust to state, so venues that converge at rest keep tracking each other in motion, even as both forecasts sharpen with the information the game reveals. Executable-liquidity quality is a property of order flow and is state-dependent. Under rapid information arrival, two separately designed order books are supplied by unsynchronized flow. Their executable quotes can then momentarily cross even as the consensus price stays aligned. Venue design enters mainly through this second layer. The model predicts cross-venue parity in forecast quality in both states, and parity in executable-liquidity quality at rest that breaks in motion by an amount growing with volatility and with venue-design heterogeneity. It is agnostic about which venue leads price discovery under fast public shocks. Our design, as we show below, cannot settle that question either way. Efficiency at rest and fragility in motion together form the observable signature of forecast quality holding while executable-liquidity quality decays.

2.1. Research Gaps

Our study differs from this work on three dimensions. First, prior execution-aware work measures arbitrage but not calibration or price discovery (Gebele & Matthes, 2026; Saguillo et al., 2025). Second, matched-event work tests price discovery but treats calibration as peripheral (Ng et al., 2026). Third, calibration work does not match identical events (Le, 2026). We combine matched-event calibration, price discovery, and fee-adjusted execution bounds in one design, and report execution claims as upper bounds benchmarked against a within-venue null. What remains unclear is whether cross-venue relationships are stable within a single rapidly resolving event. Prior analyses examine slow-moving or aggregated markets, whereas live sports impose identical real-time shocks both venues must price within seconds, separating aggregate calibration from incremental information and price disagreement from executable opportunity.

3. Research design

Our three research questions translate into four testable hypotheses that structure the analysis. Corresponding to RQ1, **H1** holds that the Kalshi and Polymarket are equally well calibrated, with no difference in ECE. **H2** further holds that neither price encompasses the other, so both carry independent predictive information. Corresponding to RQ2, **H3** holds that any pre-game cross-venue price gap is exploitable, meaning a net-of-cost locked round-trip exists at the closing line. Corresponding to RQ3, **H4** holds that the pre-game price-discovery and price-agreement relationship is unchanged during live play. We frame H1–H4 as naive efficiency nulls a believer in interchangeable venues would hold, and determine which survive on executed-trade data.

3.1. Data collection

We study moneyline (game-winner) contracts, which both platforms offer for the same event. Game outcomes resolve objectively and quickly, and predicting them is an active analytics problem in IS (Eryarsoy & Delen, 2019). We cover the four major U.S. leagues listed on both venues (NBA, MLB, NFL, NHL), September 2025 through May 2026.

Kalshi's API provides trade records with price, size, taker-side, and timestamps. Polymarket metadata comes from its Gamma API and executed trades from its public data API. Every analysis is measured from executed trades, so the two price paths share a common trade-time basis. For the resting-state panel we use each venue's last executed trade before tip-off as its closing line.

Resolved-market trade histories are immutable and cached locally for reproducibility. We matched contracts across venues on a canonical key of league, ordered team pair, calendar date, and subject team, with team codes normalized through a shared alias table. The join is sym-

metric. Table 1 traces every filter from the listed universe to each analysis sample. The resting-state panel requires a settled outcome and an executed pre-game trade on *both* complementary team contracts on *both* venues, leaving 5,038 contracts on 2,519 games. Requiring an executed pre-game trade excludes thin sides and is stricter than a resampled quote. The in-game samples are slightly larger (2,698 subject contracts) because they require only in-game trades, and a small number of games trade on Polymarket once play begins but not before it. The NBA dominates the panel because its season spans the full window and both venues list every game.

Table 1. Sample flow from the listed universe to each analysis sample.

	Filter	Contr.	Games
S1	Kalshi moneyline contracts listed	7,003	3,504
S2	Polymarket subject markets	3,521	3,521
S3	Canonical cross-venue match	2,732	2,732
S4	$\geq$1 executed trade, both venues	2,732	2,732
S5	**Resting panel**: both team contracts, settled, pre-game trade both venues	5,038	2,519
S6	**Price discovery**: $\geq$10 in-game trades both venues	2,698	2,698
S7	**In-game lock test**: all four entry legs fresh at a common grid point	2,671	2,671
S8	**Pre-game lock test**: same rule, pre-game window	2,270	2,270
S9	**Two-phase paired** comparison (S7 $\cap$ S8)	2,260	2,260

S5 counts both complementary contracts per game; S3–S4 and S6–S9 count one subject contract per game. Drops at S5: no pre-game Kalshi trade 8, no pre-game Polymarket trade 50, unsettled 366. S5 by league (contracts/games): NBA 2,136/1,068, MLB 1,424/712, NHL 1,014/507, NFL 464/232. The per-observation panel (Table 5) uses a stricter 15-second grid over a fixed three-hour pre-game window and is computed on 1,235 pre-game games.

The two venues differ institutionally in ways that bear directly on arbitrage. Kalshi is a U.S. CFTC-regulated exchange settling in dollars, with fractional-dollar tick sizes and a per-contract trading fee. Polymarket is an on-chain venue settling in USDC, with a bid–ask spread and cryptocurrency transaction costs (Schär, 2021). Polymarket charged no explicit trading fee on these leagues for most of our window, but introduced a taker fee on sports markets partway through it. We therefore use the schedule each market actually carried on each date (Section 3.2). Polymarket has also faced different legal-access and settlement frictions, so a genuine price gap may be unexploitable for any single trader.

3.2. Method

Because settlement and late-game prints can pin to 0 or 1, we treat a trade as terminal when its price is at most 0.03 or at least 0.97. The live window ends at the last non-terminal trade and begins one sport-specific game duration earlier. The pre-game closing line is the last executed trade before that window opens. Polymarket's market metadata records an official scheduled start time for 2,723 games, which lets us validate this inferred window and re-run the live analysis anchored to the official tip-off instead (Section 4.4). We compute ECE with ten equal-width bins per venue, overall and by league, treating the game outcome as ground truth. To test which price is informative we estimate a forecast-encompassing logit (Fair & Shiller, 1990), $\Pr(\text{win}) = \sigma(\beta_0 + \beta_k \operatorname{logit}(p_k) + \beta_p \operatorname{logit}(p_p))$, with standard errors clustered by game. A vanishing coefficient for one venue and a strong one for the other indicates the latter subsumes the former. We measure the favorite–longshot bias as the calibration slope, testing cross-venue differences with a platform $\times$ logit-price interaction, and capture within-venue efficiency by the overround.

Cross-venue mid-price gaps overstate exploitability because they ignore the spread. Lacking the historical order book, we use real trades to approximate the prices a trader could have paid at each venue, and ask whether buying complementary outcomes on both venues nearly simultaneously would have locked in a positive return after fees. Completed trades do not guarantee that both positions could have been filled simultaneously at sufficient size. We therefore call such instances trade-print dislocations and treat their frequency as an upper bound.

A locked round-trip buys the cheaper YES and the complementary NO, and its net edge subtracts each venue's trading cost. Kalshi charges a ceiling-rounded taker fee ($0.07\,p(1-p)$ per contract, rounded up to the cent) that is stable across our window. Polymarket's cost is market- and date-specific: each market's metadata records whether fees were enabled and, if so, the schedule and rate in force. In our panel, 2,456 markets carried no explicit fee, while 1,065 carried a sports taker fee of $0.03\,p(1-p)$, first appearing on 30 March 2026. We apply each market's own recorded schedule and add \$0.01 for on-chain transaction cost. In the live analysis, we count a lock only when all four entry prices are within 60 seconds at a common grid point, and neither venue sits in a terminal ($\leq$0.03 or $\geq$0.97) state, and we record the available size.

A trade print is not a posted quote, so a measured lock is an upper bound. To quantify how much is an artifact of combining asynchronous prints, we compute the identical statistic *within* a single venue, where a bid above an ask is impossible. That within-venue crossing rate is our model-free benchmark.

For each matched game, we align both venues' price paths on a one-minute grid and estimate the cross-correlation peak lag during live play. We build both paths from executed trades such that the two series share

a common trade-time basis. Because Polymarket trades less frequently than Kalshi, a bidirectional Granger test is confounded by the resulting forward-fill lag, so we rely on the symmetric peak-lag statistic and report the Granger asymmetry only as a robustness check.

To identify what drives the live-play gaps we build a per-bar panel on a 30-second grid over each game's matched subject contract. The unit of observation is one contract-bar. At each grid point we take each venue's last executed trade, keeping the bar only when both venues traded within 60 seconds and neither price is terminal. For Kalshi price k and Polymarket price p, the signed gap is $k - p$ and the midpoint $(k + p)/2$. Realized volatility is the absolute change in the midpoint from the previous bar, and trade intensity the combined trade count. We regress the absolute gap on volatility and trade intensity and measure reversion of the signed gap on its lagged level (Table 4).

3.3. Evaluation

Interval estimates for calibration, overround, and discrimination use a cluster bootstrap resampling entire matchups (team pairs) (Cameron et al., 2008), accounting for dependence between a game's two contracts. Because the analysis comprises a family of related tests, we control the false discovery rate with the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) and report which results survive correction. Because a failure to reject a difference is not evidence of equivalence, every cross-venue parity claim is additionally tested with two one-sided tests (TOST) against margins fixed before estimation: ± 0.02 for ECE, Brier score, and AUC, ± 0.10 for the calibration slope, and ± 0.01 for the overround. These margins are economically conservative. Each sits at or below the venues' own absolute calibration error and well inside the 2.5–3 cent round-trip breakeven implied by spreads and fees. A difference we declare equivalent is therefore too small to change what a consumer of the price would do. We summarize lead–lag direction with the peak-lag distribution and treat binomial and Granger diagnostics as caveated robustness checks (McNemar, 1947). The seeded pipeline is reproducible from a single matched panel.

4. Results

4.1. Parity and joint efficiency at rest (prior to the match start)

The two venues are equivalent in calibration, not merely indistinguishable. Overall ECE is 0.017 for Kalshi and 0.018 for Polymarket. The matchup-clustered 95% confidence interval for the difference is $[-0.010, +0.005]$, no league shows a significant gap (Table 2), and the equivalence test rejects a difference as large as ± 0.02 (90% CI $[-0.008, +0.004]$). Figure 1 shows the two reliability curves tracking the diagonal and each other. The favorite–longshot slope is likewise equivalent. Both calibration slopes sit near one (overall 0.91 for Kalshi and 0.91 for Polymarket, difference -0.003, $p = 0.46$, equivalent at ± 0.10), and no league shows a significant cross-venue slope gap, so neither venue prices more sharply than the other once both prices are read at the same instant.

Measured contemporaneously from executed trades, the two closing prices are nearly identical, with a matchup correlation of 0.998 and a mean absolute cross-venue gap of 0.006. Neither price carries predictive information that the other lacks, which falsifies H2. In a deviation-form logit, the Polymarket level remains strongly predictive while the disagreement term $\mathrm{logit}(p_k) - \mathrm{logit}(p_p)$ is indistinguishable from zero. These results suggest that the small cross-venue disagreements carry no incremental predictive information. Discrimination tells the same story. The pooled AUC is 0.6795 for Kalshi and 0.6793 for Polymarket, with a clustered ΔAUC interval $[-0.0010, +0.0014]$ that is equivalent at ± 0.02, and no league shows a significant gap.

The one resting difference is structural. Kalshi's overround sums to 101.21% overall, significantly above Polymarket's 100.80%. This gap of 0.41 percentage points, 95% CI $[+0.37, +0.46]$, is positive in every league and survives false-discovery control. It is not a mechanical artifact of Kalshi's fee. That fee is charged on top of the executed price and so cannot enter a sum of executed prices. At 1.55% per contract it is also roughly four times the gap. On the subset quotable on both sides of both venues ($n = 2{,}638$), where the gap is $+0.40$ points, decomposing at the spread-neutral midpoint leaves a structural component of $+0.26$, CI $[+0.22, +0.30]$, the remaining $+0.14$ reflecting where trades sit within the spread. The difference is statistically robust but economically negligible. It is equivalent to zero against a one-point margin and amounts to 17% of the round-trip breakeven.

Table 2 summarizes the structural picture. The overround is positive and tightly estimated on both venues and modestly larger on Kalshi in every league. The calibration slope is near 1 in both venues and does not differ between them. Of the thirteen calibration and overround tests in the family, five survive Benjamini–Hochberg correction, namely all five overround-difference tests. Aggregate parity holds and, unlike in non-contemporaneous comparisons, it is not masked by tail disagreement.

Kalshi is no worse calibrated than Polymarket on longshots ($p < 0.15$, $\Delta\mathrm{ECE} = +0.001$, $q = 0.97$), in

the 0.4–0.6 midrange (+0.002, $q = 0.80$), or on heavy favorites (+0.003, $q = 0.97$). Every aggregate, per-league, and tail ECE comparison is insignificant after FDR correction, so calibration parity holds across the full price range.

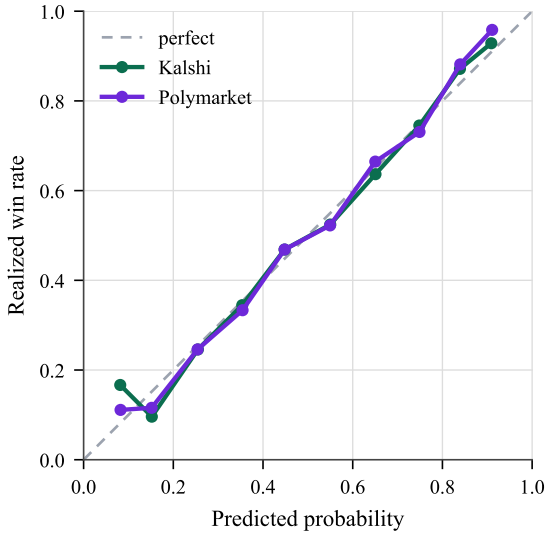

Figure 1. Reliability curves for Kalshi and Polymarket on the matched panel.

Table 2. Calibration, overround, and calibration slope by league (matchup-clustered).

Lg	n	$\text{ECE}_{K/P}$	ΔECE CI	$\text{Ovr}_{K/P}$	Slp
All	5038	.017/.018	$[-.009, .005]$	$1.21^{*}/0.80$	.91
NBA	2136	.022/.024	$[-.011, .009]$	$1.28^{*}/0.78$	.97
MLB	1424	.017/.013	$[-.009, .011]$	$1.12^{*}/0.90$	.56
NFL	464	.048/.048	$[-.015, .019]$	$1.18^{*}/0.78$	.93
NHL	1014	.028/.040	$[-.024, .008]$	$1.22^{*}/0.71$	.85

Overround in percentage points. * Kalshi−Polymarket overround difference significant at $q < 0.05$ (FDR). Slope shown for Kalshi; cross-venue slope differences are not significant and are equivalent at ±0.10.

4.2. Economically negligible pre-game locks

Before tip-off, cross-venue locks are effectively absent, which rejects H3. Using executed trades to infer entry prices, effective spreads are about 1 cent per venue. Although 42 tested games contained at least one net-of-fee locked moment, these instances represented only 1.9% of games and cleared at a median observed transaction size of six contracts. The typical game's net edge is only −0.03. Non-contemporaneous closing-line gaps are not exploitable, and the simultaneous gap typically falls below the ≈2.5–3% breakeven implied by spreads and fees.

The closing lines also actively converge as information accumulates. Tracking the absolute cross-venue gap at fixed horizons before tip-off, the mean gap falls from 0.0123 a full day out ($n = 2{,}160$) to 0.0085 within twelve hours, 0.0076 at three hours, and 0.0074 in the final hour ($n = 2{,}691$). A paired test on the same games (the gap at $T-24$h minus the gap at $T-3$h, clustered by matchup) gives a mean reduction of +0.0048 with interval $[+0.0041, +0.0056]$ that excludes zero. The two prediction markets converge at a common price as the game approaches, the signature of a jointly efficient resting state.

4.3. Simultaneous repricing and transient trade-print dislocation in motion

During live play, the two prediction markets track each other closely and reprice simultaneously. Measuring both venues from their executed trades on a common minute grid, the cross-correlation peak lag has a median of zero, and roughly 96% of games show no lag at all, while in-game coupling is strong (mean peak $|\rho| = 0.72$, Table 3). The minority of games that do show a one-bar lag tilt toward Kalshi (about 3% versus 1% for Polymarket), and a directional Granger signal toward Kalshi survives.

We do not interpret either as price-discovery leadership, and we cannot rule leadership out. Kalshi trades a median of 28.8 times per minute during play against Polymarket's 8.3, and Polymarket has no trade in 7% of one-minute bars at the median (12% mean), so the sparser series must be forward-filled. Forward-filling is not neutral. Resampling any series onto Polymarket's observed trade times injects a systematic one-bar lag, even when the two paths are identical by construction. That is the direction the Granger asymmetry points (significant Kalshi→Polymarket in 2,359 of 2,698 games against 1,335 in reverse). Absence of a detected leader in this design is therefore not evidence of absence, and we report the result as unidentified.

Table 3. In-game price discovery (1-min grid, 2,698 games).

Measure	In-game
Median peak lag	0 min
Simultaneous (zero-lag)	96%
1-bar lag, Kalshi / Poly†	3% / 1%
Mean peak $\|\rho\|$	0.72
Kalshi / Poly trades per min	28.8 / 8.3
Polymarket empty bars (median)	7%

† Confounded by Polymarket's lower trade frequency: forward-filling the sparser series injects a lag toward Kalshi. Read as unidentified, not as leadership.

Brief trade-level discrepancies still emerge during live play, separating the live from the resting state and rejecting H4. Requiring all four trade-based entry prices within 60 seconds and excluding terminal states, 2,436 of 2,671 tested games (91.2%, clustered 95% CI $[0.900, 0.924]$)

contain at least one positive net-of-fee locked-payoff window (Table 5). Among the 2,260 games quotable in both phases on the closing-line grid, locks occur in 92.2% of games during play but 1.9% before, a clustered paired difference of $+0.90$, CI $[+0.89, +0.92]$ (McNemar, 1947). The finer 15-second bar panel, which fixes a three-hour pre-game window, gives 91.2% and 3.6% (Table 5). The two pre-game figures differ because that panel exposes each game to more observations, not because the venues differ. The windows are fleeting and shallow, with a median of six per game, a best edge near 5%, a size of six contracts, and durations under one minute (20,593 windows).

The per-game rate overstates the contrast, because the states do not offer comparable observation exposure. A live game supplies a median of 168 eligible grid points against six before tip-off, so a per-game indicator is close to a test of whether *any* dislocation occurred. Per observation, dislocations occur in 6.2% of live bars, CI $[6.0\%, 6.3\%]$, against 0.8% of pre-game bars, $[0.5\%, 1.2\%]$. The state difference is roughly eight-fold per observation, against twenty-five-fold on the same panel's per-game rates.

How much of this is executable? Applying the identical statistic within a single venue answers the question, because inside one order book a bid can never exceed an ask (lower panel, Table 5). Kalshi's own trade prints cross in 3.98% of live bars and 91.2% of games, a per-game rate identical to the cross-venue lock rate. Polymarket's own prints cross in 11.83% of bars. The per-game statistic is thus close to a detector for whether any trade-print asynchrony occurred, which is almost always. The per-bar comparison is more informative but still bounded. A within-Kalshi rate of 3.98%, or 1.61% net of Kalshi's own fee, is the same order of magnitude as the 6.17% cross-venue rate. We therefore read the live results as an upper bound on dislocation, not as capturable arbitrage.

The dislocations are still specific to the contract and the instant. Pairing the Kalshi legs of one game with the Polymarket legs of an unrelated game yields a per-bar rate of 89.7%, and shifting the Polymarket legs five minutes later within the same game yields 56.2%. Both far exceed the contemporaneous 5.7% on the same net-of-fee basis. The measure therefore isolates same-contract, same-instant disagreement. However, it cannot establish that the disagreement was tradeable.

Forecast quality does not decay alongside execution quality, but neither is it literally invariant. Scored against realized outcomes on the same live bars ($n = 186{,}466$), the venues remain equivalent: ECE is 0.0083 for Kalshi and 0.0085 for Polymarket (difference -0.0002, 95% CI $[-0.004, +0.005]$), AUC 0.7925 against 0.7920, Brier 0.1850 against 0.1852, and slope 1.037 against 1.026. All are equivalent at the pre-specified margins, and parity holds in each quarter of the live window separately. What changes with state is the sharpness of both forecasts, not the gap between them. Kalshi's AUC rises from 0.680 at the closing line to 0.792 across live bars, 0.920 in the final quarter, and 0.997 at the last observation. Brier falls in step, from 0.224 to 0.185, 0.115, and 0.027. The precise claim is therefore that *cross-venue parity* in forecast quality is state-invariant, not that forecast quality itself is. The calibration slope also drifts from 0.75 in the first quarter to 1.36 in the fourth, so late in-game prices on both venues are systematically not extreme enough.

The mechanism behind the dislocations is the transient liquidity effect anticipated above. In the in-game panel (Table 4), the absolute cross-venue gap rises with 30-second midpoint volatility, and about 74% of a displacement closes in the next bar. The gaps are two-sided but not symmetric, with Polymarket the cheaper side in 61% of bars, Kalshi in 24%, and the two equal in 15%, and the mean signed gap is only 0.0053. Trade intensity adds no explanatory power. Its coefficient is $+0.000025$ with a 95% confidence interval of $[-0.00012, +0.00017]$, a tightly estimated zero. With a shallow median size of about six contracts and sub-minute duration, these dislocations arise mainly from brief volatility-driven movements in trade-inferred entry prices.

Table 4. In-game gap mechanism (791,971 contract-bars, 2,704 clusters). Dependent variable is the absolute gap $|k - p|$ on a 30-second grid; standard errors clustered by contract.

Term	Coef.	SE	95% CI
Volatility (per 1%)	$+.00254^{***}$	.00010	$[.0023, .0027]$
$\log(1+\text{tr.})$	$+.000025$	.000075	$[-.0001, .0002]$
Constant	$+.00886^{***}$	.00019	$[.0085, .0092]$
Reversion, gap_t	$-.738^{***}$	.0232	$[-.783, -.693]$
R^2 (levels / reversion)	0.108 / 0.358		

$^{***}p < 0.001$. Trade activity is a tightly estimated zero.

4.4. Robustness

Fees and frictions. Table 6 imposes the costs a cross-venue trader would face. Replacing the fee-free assumption with each market's recorded schedule lowers the per-bar rate from 6.17% to 5.73%. In the post-30-March subsample, where the sports taker fee was live, it falls from 5.01% to 3.68%. Adding one cent per contract for conversion, settlement, and capital lowers it to 3.46%, and requiring the edge to survive a fifteen-second round trip lowers it to 1.37%. Under pessimistic assumptions of three cents and sixty seconds, 0.14% of observations and one game in ten remain. Most dislocations, though not all, vanish under realistic costs.

Table 5. Lock windows per game and per observation on the 15-second bar panel (four entry prices fresh within 60 s; pre-game window fixed at three hours, $n = 1{,}235$), with the within-venue crossing benchmark. The closing-line pre-game rate in Section 4.2 uses a coarser grid and a different sample. Median in-game window length is under one minute.

Quantity	Pre-game	In-game
Games with a lock	3.6%	91.2%
95% CI	[.027, .047]	[.900, .924]
Locked obs. (per bar)	**0.79%**	**6.17%**
95% CI	[.005, .012]	[.060, .063]
Median eligible bars/game	6	168
Median best edge	< 0	$+0.05$
Median size (contracts)	5	6
Within-venue benchmark (bid > ask)		
Kalshi, per bar / game	1.1% / 5.8%	4.0% / 91.2%
Polymarket, per bar	2.8%	11.8%
Kalshi net of own fee	0.3%	1.6%

Table 6. Live dislocation rate under stacked execution costs.

Specification	Per game	Per bar
Polymarket assumed fee-free	91.2%	6.17%
+ actual per-market fee schedule	89.8%	5.73%
+ 1¢ cross-venue friction	81.7%	3.46%
+ 15 s execution latency	60.8%	1.37%
2¢ friction, 30 s latency	32.0%	0.48%
3¢ friction, 60 s latency	10.6%	0.14%

Synchronization. Tightening the 60-second freshness requirement to 30, 15, 10, 5, and 2 seconds moves the per-game rate from 89.8% to 82.3%, 70.0%, 59.7%, 45.8%, and 33.9%, while the per-bar rate is flat or slightly rising (5.73% to 7.76%). The per-game figure is thus driven by how many eligible observations a game supplies, not by stale prices. The per-bar frequency is not an artifact of loose synchronization.

Terminal states and game timing. Settlement prints pin to the boundary, so we end each live window at the last non-terminal trade and trim the three-point boundary band. We validate the trade-inferred window against the official scheduled start recorded for 2,723 games. The inferred start runs a median of 27 minutes early, and only 37% of windows begin within 20 minutes of tip-off, so it is imprecise. The live result is nonetheless insensitive to it, with the official anchor giving 89.5% per game and 5.82% per bar. The pre-game analysis uses the official tip-off, which is better defined and yields a lower rate (0.31% of bars).

Selection. Requiring an executed pre-game trade on both venues could in principle retain only the contracts most likely to agree. It does not. The filter excludes 19 of 2,712 games (0.7%), and those are the highest-volume games (median 14,558 Kalshi pre-game trades against 1,239), not the thin one-sided tail. Their mean in-game gap is slightly *smaller* (0.0104 against 0.0120).

5. Discussion

5.1. Practical implications

For traders, the implication is restrictive, and more so than our headline rate suggests. Cross-venue locks are negligible before game time. During live play they are common in the trade record but brief, shallow, and largely not capturable. Trade-print asynchrony alone reproduces crossings of the same order of magnitude. A cent of friction then removes about 40% of what remains, a sixty-second execution delay removes about 85%, and the two together remove 98%. Contesting the residual would require automated sub-15-second execution on both venues and pre-funded balances in dollars and USDC. Kalshi's higher overround reflects a structural cost difference, not better pricing. At 0.4 percentage points it is too small to act on.

Because cross-venue parity in forecast quality holds in both states while executable-liquidity quality degrades in motion, the right way to treat a prediction-market price depends on how a consumer uses it. A forecaster or dashboard should read either venue's price and refresh it lazily, since the venues are equivalent in both states and pooling buys little. A live alerting system should require cross-venue agreement before firing a threshold crossing, and should widen displayed uncertainty under high volatility. Disagreement in motion signals unstable execution, not new information. An automated trader must account for quote freshness, shallow size, each venue's fee schedule on the relevant date, and latency. Any measured window should be benchmarked against the within-venue statistic before acting on it.

5.2. Theoretical contributions

Our results confirm the two-layer account of signal quality, with one refinement. Measured contemporaneously, Kalshi and Polymarket are equivalent at rest in calibration, discrimination, calibration shape, and incremental predictive information, and the equivalence tests show this holds during live play as well. What is state-invariant is the *parity* between venues, not the quality itself. Both forecasts sharpen substantially as a game progresses. Executable-liquidity quality behaves differently again, degrading in motion as volatility dislocates trade-inferred entry conditions. Signal quality is therefore multidimensional, and predictive equivalence does not imply execution-level stability.

Our findings are more cautious than the literature on price-discovery leadership. Prior work finds Polymarket leads Kalshi in election contracts (Ng et al., 2026). We detect no leader during live sports play, but cannot distinguish that from insufficient power. Polymarket trades

roughly a third as often as Kalshi here, and the resulting forward-fill biases the estimator toward a Kalshi-leads reading. The difference between their setting and ours may reflect trade frequency as much as any elections-versus-sports regime difference, and we do not claim the latter. Methodologically, cross-venue studies must use contemporaneous observations, apply the fee schedule in force on each date, and benchmark inferred crossings against a within-venue null before reading price gaps as opportunities. Accordingly, H1 survives, H2 is rejected because the two contemporaneous closing prices are informationally redundant, not because one venue dominates the other, H3 is rejected because pre-game locks are economically negligible, and H4 is rejected because live volatility generates frequent but shallow and largely non-executable dislocations.

5.3. Limitations

Our claims are subject to four limitations. First, we inferred cross-venue entry conditions from executed trades, not from posted quotes. A trade print confirms that a transaction occurred at a given price and size. It does not establish that either price remained available, or that both legs could have been filled simultaneously. The within-venue benchmark in Table 5 quantifies how much this matters. The same statistic applied inside one order book, where a crossing is impossible, fires on 3.98% of live observations. Our live estimates are therefore upper bounds on trade-print dislocation, not measurements of realized or scalable arbitrage (Saguillo et al., 2025). Establishing executability would require synchronized historical order-book depth, which neither venue publishes.

Second, game windows are inferred from trade timestamps, not official clocks. Against the official start available for 2,723 games, the inferred window opens a median of 27 minutes early, though re-running the analysis on the official tip-off leaves results essentially unchanged. Trade prints are also not synchronized quotes, so measurement uncertainty remains despite the terminal-state, freshness, and within-venue controls.

Third, cross-venue agreement is not itself evidence of efficiency, since both venues could share a common bias. Our calibration and discrimination tests partly address this by scoring prices against realized outcomes, not against each other. However, a bias common to both venues and invisible to a ten-bin reliability check cannot be ruled out here.

Fourth, external validity is bounded by the setting and sample. We examine Kalshi and Polymarket moneyline contracts for four U.S. leagues from September 2025 to May 2026. Other venues, seasons, sports, and instruments may exhibit different patterns of calibration, liquidity, or cross-venue dislocation. These limitations constrain claims about realized profitability and generalizability, but not the observed distinction between pre-game pricing and live-play trade-based dislocations.

6. Conclusion

Two prediction-market venues can produce nearly identical resting signals while their trade records briefly diverge amid live volatility. Before game time, Kalshi and Polymarket are equivalent in calibration, discrimination, and incremental predictive information, and meaningful trade-based locks are absent. During live play the venues reprice together, with no leader we can detect, yet their trade-inferred entry conditions are temporarily dislocated. Those crossings are an upper bound, not an opportunity. The same statistic within a single venue, where arbitrage is impossible, reproduces crossings of comparable magnitude. Realistic fees, frictions, and latency then remove most of what remains. Prediction markets can therefore be equivalent as resting forecasting mechanisms while their executable conditions come apart, at least as trade prints record them, during rapid information arrival.

These findings recast a prediction-market probability as a state-dependent data product. For information-systems researchers and designers, calibration and predictive equivalence do not guarantee live execution-level stability, and the relevance of a market feed depends on whether the consumer requires a stable forecast, a timely alert, or an actionable trading signal. They also carry a measurement lesson. Any study inferring executable conditions from trade prints should report the within-venue crossing rate of its own estimator, which bounds how much of a cross-venue result the method manufactures. Future research should test these relationships using synchronized order-book data, additional venues, and explicit models of execution, settlement, and access costs, clarifying whether live dislocations arise from game-state shocks, spread widening, liquidity withdrawal, or sub-minute latency. More broadly, the distinction motivates examination of other settings in which the same claim is priced across multiple venues, including cross-listed securities, cryptocurrency exchanges, and centralized versus decentralized prediction protocols.

References

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., Neumann, G. R., Ottaviani, M., Schelling, T. C., Shiller, R. J., Smith, V. L., Snowberg, E., Sunstein, C. R.,

Tetlock, P. C., Tetlock, P. E., … Zitzewitz, E. (2008). The promise of prediction markets. *Science*, *320*(5878), 877–878.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289–300.

Berg, J. E., & Rietz, T. A. (2003). Prediction markets as decision support systems. *Information Systems Frontiers*, *5*(1), 79–93.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427.

Croxson, K., & Reade, J. J. (2014). Information and efficiency: Goal arrival in soccer betting. *The Economic Journal*, *124*(575), 62–91.

DeGroot, M. H., & Fienberg, S. E. (1983). The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society Series D (The Statistician)*, *32*(1–2), 12–22.

Du, Q., Hong, H., Wang, G. A., Wang, P., & Fan, W. (2017). Crowdiq: A new opinion aggregation model. *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS)*. https://doi.org/10.24251/hicss.2017.211

Eryarsoy, E., & Delen, D. (2019). Predicting the outcome of a football game: A comparative analysis of single and ensemble analytics methods. *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS)*. https://doi.org/10.24251/hicss.2019.136

Fair, R. C., & Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *The American Economic Review*, *80*(3), 375–389.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Franck, E., Verbeek, E., & Nüesch, S. (2013). Inter-market arbitrage in betting. *Economica*, *80*(318), 300–325.

Gebele, J., & Matthes, F. (2026). Semantic non-fungibility and violations of the law of one price in prediction markets. https://doi.org/10.48550/arXiv.2601.01706

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 1321–1330.

Hanson, R. (2003). Combinatorial information market design. *Information Systems Frontiers*, *5*(1), 107–119.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, *50*(4), 1175–1199.

Le, N. A. (2026). Decomposing crowd wisdom: Domain-specific calibration dynamics in prediction markets. https://doi.org/10.48550/arXiv.2602.19520

Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters*, *91*(3), 425–429.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157.

Ng, H., Peng, L., Tao, Y., & Zhou, D. (2026). Price discovery and trading in modern prediction markets. https://doi.org/10.2139/ssrn.5331995

Oliven, K., & Rietz, T. A. (2004). Suckers are born but markets are made: Individual rationality, arbitrage, and market efficiency on an electronic futures market. *Management Science*, *50*(3), 336–351.

Page, L., & Clemen, R. T. (2013). Do prediction markets produce well-calibrated probability forecasts? *The Economic Journal*, *123*(568), 491–513.

Saguillo, O., Ghafouri, V., Kiffer, L., & Suarez-Tangil, G. (2025). Unravelling the probabilistic forest: Arbitrage in prediction markets. *7th Conference on Advances in Financial Technologies (AFT 2025)*, *354*, 27:1–27:24. https://doi.org/10.4230/LIPIcs.AFT.2025.27

Schär, F. (2021). Decentralized finance: On blockchain- and smart contract-based financial markets. *Federal Reserve Bank of St. Louis Review*, *103*(2), 153–174.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55.

Snowberg, E., & Wolfers, J. (2010). Explaining the favorite-long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy*, *118*(4), 723–746.

Tziralis, G., & Tatsiopoulos, I. (2007). Prediction markets: An extended literature review. *The Journal of Prediction Markets*, *1*(1), 75–91.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*(2), 107–126.