Two Venues, One Policy Event: Measuring Repricing in Prediction Markets

Orhan Torul

Department of Economics, Boğaziçi University

September 2026

Abstract

How much does venue choice change measured announcement repricing? I compare Polymarket and Kalshi contracts on the same Federal Reserve decision around 31 releases spanning 13 meetings. Thirty-minute changes correlate at 0.94 (meeting-block 95% interval: 0.82–0.99), yet the venues share only four of their five largest events. Classification disagreement varies with the movement threshold: seven events differ at one percentage point and two at five. Kalshi's quote-implied ranges permit both classifications for 22 and 17 events, respectively, limiting what these point-price differences reveal. Release-type mean rankings agree, with imprecise venue effects. The retained releases have larger Polymarket movements than 35 excluded releases. The contribution is a reproducible comparison of event selection, raw movement summaries, and quote-implied classification uncertainty within this selected 2024–2026 archive.

Keywords: Prediction markets; policy expectations; announcements; market microstructure.
JEL classification: G14, E44, E58, C81.

1 Introduction

Researchers using prediction-market prices must choose both a contract and a historical price series. Even when two contracts concern the same policy decision, replacing the venue can change which announcements enter a large-response sample. Across 31 matched releases, Polymarket and Kalshi share four of their five largest thirty-minute movements and seven of their eight largest. Their observed release-type means have the same ordering, but their event classifications differ. This paper measures these consequences while holding the release and policy payoff fixed.

The distinction matters for using price changes as measured monetary-policy news. The cross-venue correlation is 0.94, with a meeting-block 95% interval of [0.82, 0.99], but correlation does not determine which observations a threshold rule selects. I therefore report the full classification-disagreement curve, event rankings, and paired release-type contrasts. A one-point rule gives seven different classifications and a five-point rule gives two. Historical Kalshi quote ranges span both sides of those rules for 22 and 17 events, respectively. Every midpoint classification disagreement at

these thresholds is compatible with the corresponding Kalshi range. The data establish differences between recorded measurements, without attributing them to different beliefs.

Two construction choices precede that comparison. The complement of Polymarket's no-change price covers both cuts and hikes. Kalshi's rate thresholds represent the same change event only after the preceding meeting establishes the base rate. I retain both legs of the Kalshi construction and preserve the archives' timestamps. The estimand is agreement conditional on valid matched records around the same releases. The applied comparisons concern event selection and raw price movements; they do not estimate response coefficients per unit of economic surprise.

Coverage materially restricts that estimand. The calendar contains 66 releases, of which 31 yield valid matched prices with Kalshi endpoint records no more than thirty minutes old. Their average absolute Polymarket change is 5.12 points, compared with 1.70 among the 35 excluded releases. A five-minute age bound leaves eight events. Results for the retained sample therefore cannot establish interchangeability for the full calendar or a different liquidity environment.

Quote ranges describe another limit on precision. Kalshi's synthetic change ranges have a median width of seven points and contain the Polymarket change in 29 of 31 cases. For comparison, pairing each range with every other event's Polymarket change yields 46.3% containment. The actual event pairing contains more information than this descriptive mismatching benchmark, but the widths remain too large to validate one-point precision. These ranges are historical quote constructions, not confidence intervals or executable arbitrage bounds.

The analysis complements recent uses of prediction markets for macroeconomic expectations (Diercks et al., 2026; Swanson et al., 2025; Eichengreen et al., 2025) and work separating price discovery from measurement noise (Hasbrouck, 1995; Putniņš, 2013). Section 2 develops that connection. Sections 3 and 4 define the records and comparison. Section 5 reports agreement and construction sensitivity; Section 6 measures the consequences of venue replacement. The separate supplement contains the full event audit and auxiliary timing, forecast-score, and historical-regime checks.

2 Related literature

2.1 What prediction-market prices reveal

The interpretation of a binary contract price is an economic question before it is a data-cleaning question. Wolfers and Zitzewitz (2004) emphasize that contract design determines which feature of an uncertain outcome a market can reveal. Manski (2006) shows that, even with risk-neutral price-taking traders, a price does not generally identify the full distribution or mean of heterogeneous beliefs. Wolfers and Zitzewitz (2006) provide sufficient conditions connecting prices to mean beliefs and examine departures from those conditions. Together, these contributions motivate the distinction maintained here between a price-based expectation measure and a directly observed population probability. Payoff equivalence makes the cross-venue comparison meaningful; it does not establish identical preferences, information, or equilibrium pricing across venues.

Market-based macroeconomic forecasts also predate the current prediction platforms. Gürkaynak and Wolfers (2007) study economic derivatives to extract forecasts, uncertainty, and risk, comparing their information with survey forecasts. Snowberg et al. (2013) survey the use of prediction markets for economic forecasting. This literature establishes the potential value of explicitly contingent claims relative to conventional assets whose payoffs depend on many economic outcomes. The present paper takes that potential as its starting point. It asks whether two available implementations of the same contingent claim supply sufficiently similar changes for an empirical application, rather than whether prediction markets are useful in principle.

2.2 Recent evidence on macroeconomic prediction markets

The closest recent studies use prediction markets to address substantive questions about expectations. Diercks et al. (2026) compare Kalshi macroeconomic forecasts with survey and financial-market benchmarks and examine responses to economic news. Swanson et al. (2025) use Kalshi event contracts to study how macroeconomic expectations respond to monetary-policy announcements, bringing higher-frequency observations to the debate over the Federal Reserve information effect. Eichengreen et al. (2025) use Polymarket contracts and wallet-level positions to study political pressure on central-bank independence and policy expectations. Their work also compares Polymarket-implied policy expectations with fed-funds-futures benchmarks. Agreement with a conventional rate instrument is therefore not, by itself, a new contribution of the present paper.

These studies differ in outcomes, contracts, and empirical questions. The additional question here concerns portability across prediction venues: for a common next-meeting payoff, how much does a measured announcement response depend on the historical price field and its observation convention? I hold the release and terminal event fixed, retain both the cut and hike components, and report the coverage cost of restricting record age. Direct absolute errors and quote-implied ranges supplement regression slopes. This comparison assesses whether the recorded outcome measure changes with venue choice while holding the release and terminal policy event fixed. The substantive identification strategies in those studies answer different questions.

A further distinction concerns calibration and trading frictions. Bürgi et al. (2026) document informative Kalshi prices alongside favorite–longshot patterns and differences between makers and takers. Their evidence shows why a price can contain useful information without being an unbiased probability or yielding the same return to different trading sides. In the present setting, agreement inside recorded quote ranges does not show that both venues are correctly calibrated. Conversely, a difference between two midpoints need not be a difference that a trader could exploit. The supplementary fixed-horizon scoring exercise and the quote-range calculation address these separate issues, within the limits of the available outcomes and price records.

2.3 Macroeconomic announcements and the policy path

Rapid financial-market responses to public announcements are well established. Ederington and Lee (1993) document concentrated price adjustment and subsequent volatility around scheduled macroeconomic releases. Fleming and Remolona (1999) show that announcement-time price formation involves changes in spreads and trading activity as well as prices. Andersen et al. (2003) link high-frequency exchange-rate changes to announcement surprises and show that the response depends on the news content and its sign. These findings make Treasury adjustment a useful empirical benchmark here, but they also limit the novelty of documenting that a deep futures market records a response quickly.

The monetary-policy literature further separates an announcement from its unexpected component and a current policy decision from the future policy path. Kuttner (2001) uses fed-funds futures to distinguish anticipated from unanticipated policy actions. Gürkaynak et al. (2005) show that target-rate and policy-path factors have distinct asset-price effects. The term-structure synthesis of Gürkaynak and Wright (2012) explains why longer-maturity yields contain expectations and risk-premium components. A two-year Treasury future is consequently an informative timing comparator, but not an identical-payoff external truth for a binary next-meeting contract. Recovering a comparable expected rate from prediction contracts would require the full outcome distribution; a complement-of-hold price alone cannot provide it.

The multidimensional nature of releases also matters. Gürkaynak et al. (2020) show how non-headline news that is unobserved by the econometrician can affect event-window asset prices. The present paper compares the measurements of the same event rather than estimating response coefficients per unit of economic surprise. It therefore does not interpret larger absolute employment reactions than inflation reactions as evidence that employment news is intrinsically more informative. A change across cycles in the ranking of raw movements would not directly reconcile these results with surprise-normalized changes in the mean expected policy rate. That comparison would change both the news normalization and the outcome measure.

2.4 Price discovery, measurement noise, and agreement

The cross-market price-discovery literature asks which venue contributes to a common efficient price. Hasbrouck (1995) develops information shares for linked securities, and de Jong and Schotman (2010) model intraday price dynamics across fragmented markets while addressing structural identification. Putniņš (2013), in the *Journal of Empirical Finance*, shows that conventional discovery metrics can confound the speed of information incorporation with differences in price-series noise. This distinction is particularly relevant when comparing a sparsely sampled prediction history with a futures tape. A faster observed half-crossing is not an information share, and imposing an identical observation grid does not remove all microstructure differences.

The present analysis accordingly separates co-movement, absolute agreement, and observed

timing. Its difference-versus-mean display follows the measurement-comparison principle of Bland and Altman (1986): correlation does not determine whether one measurement can replace another at a specified tolerance. I adapt that principle to event changes and use meeting-level uncertainty rather than treating the releases as independent measurements. This is an application of an established distinction, not a proposed new agreement estimator. Quote ranges add an economically specific interpretation of the observed discrepancies.

Finally, evaluating a probability against a realized outcome differs from comparing two measurements. Gneiting and Raftery (2007) develop the theory and use of proper scoring rules for probabilistic forecasts. The supplementary Brier losses score the same terminal event at fixed calendar horizons, with at most one forecast per meeting at each horizon. Repeated daily forecasts can provide additional information about a forecast path, but they do not create additional policy realizations. That distinction governs the deliberately limited inference from the 16 linked meetings.

3 Contracts, archives, and sample

3.1 The same policy event

The observation unit is an 8:30 a.m. Eastern Time macroeconomic release and the FOMC meeting to which its prediction contract refers. The calendar contains 22 Employment Situation releases, 22 Consumer Price Index releases, and 22 Producer Price Index releases from April 5, 2024, through February 27, 2026. It includes nine delayed releases, three of each type. The complete calendar and its contract assignment appear in Supplement H. This is the study's inherited release calendar, not a census of every macroeconomic announcement or every historical prediction market. In particular, the cross-market coverage filters further restrict the population to which the results apply.

Let r_0 denote the upper bound of the federal funds target range established before meeting m, and let r_m be the upper bound after that meeting. The target event is a change, $C_m = \mathbf{1}\{r_m \neq r_0\}$. Polymarket offers a binary no-change contract. I define its complementary price as

$$P_m^P(t) = 100[1 - H_m(t)], \tag{1}$$

where $H_m(t)$ is the archived Yes price of that contract. The superscript P denotes Polymarket and prices are in percentage points. This transformation defines a price-based measure of the change event. It is not a direct observation of the No token's executable bid or ask.

Kalshi's relevant contracts pay on the upper target bound exceeding a specified threshold. Write their Yes prices as $Q_m(k,t)$. On a quarter-percentage-point policy grid, a matched change price is

$$P_m^K(t) = 100\{1 - Q_m(r_0 - 0.25, t) + Q_m(r_0, t)\}. \tag{2}$$

The archived rules specify an upper target bound strictly greater than the strike; all 32 linked threshold metadata records—two per meeting for the 16 linked meetings—and their settlements are re-

tained and checked. Thus $Q_m(k,t)$ refers to $r_m > k$, not $r_m \geq k$. The first component represents a cut and the second a hike. Keeping the hike component avoids assuming its price is zero merely because no hike subsequently occurs. This mapping is conditional on the target-rate grid and the contracts' resolution definitions. An off-grid move or an intervening unscheduled decision could break the economic equivalence; the contract metadata are retained so that the assumption is inspectable.

The base rate must already be known at the observation date. Before the preceding meeting, a level contract and a change-from-the-prevailing-rate contract refer to different events. All reported release comparisons occur after the preceding scheduled meeting and before the linked decision. The supplementary fixed-horizon forecasts use the same restriction. Polymarket's archived outcome ordering is Yes, No for all 16 linked meetings; its terminal outcome and Kalshi's threshold settlements provide a check on the mapping. A machine-readable event crosswalk supplies market and token identifiers, both strikes, and archived Kalshi resolution rules for every matched release. Terminal settlement checks cover all 16 linked outcomes; repeated releases do not create additional independent settlement tests. Prices can nevertheless disagree even when terminal payoffs match, owing to trading frictions, risk preferences, or measurement differences.

3.2 What each price record measures

Polymarket histories come from the public CLOB price-history endpoint, requested at one-minute fidelity around each release.[1] The data contain timestamp–price pairs. I retain their actual timestamps rather than floor them to minute or hourly boundaries. These observations are sampled historical prices; the endpoint does not supply a contemporaneous bid, ask, trade identifier, or quote-update age for each observation. A recent sample is therefore not proof of a recent underlying trade or quote revision. The endpoint documentation defines minute fidelity but does not establish an OHLC interval-start or interval-end convention for these timestamp–price pairs. I retain the supplied times without a bar-end reassignment and treat the unresolved timing semantics as a limit on both the thirty-minute comparison and the auxiliary timing exercise. Section 5 tests hypothetical one-minute shifts for the main result.

Kalshi histories come from its public historical candlestick endpoint at one-minute intervals.[2] Each candle carries an end timestamp, separate bid and ask fields, trade-price fields, and volume. I require both bid and ask closes to be present and to satisfy $0 \leq b \leq a \leq 1$, and calculate their midpoint. Missing sides are never replaced with zero or one. Both fields belong to the same candle, although their last updates within that interval need not be simultaneous. This possible within-candle bid–ask asynchrony differs from the age of the candle itself; a record-age screen cannot remove it. I do not replace a missing transaction close with a mean or previous trade. The primary analysis uses quotes because requiring transactions in both legs of equation (2) sharply limits coverage, particularly for the hike leg. A candle's age is observable; the age of the quote carried in that candle is not necessarily

[1] The official research-data overview documents token-specific historical requests: https://institute.polymarket.com/data.

[2] See https://docs.kalshi.com/api-reference/historical/get-historical-market-candlesticks.

observable. Throughout the paper, "record age" has this narrower meaning.

The public responses were archived on September 17, 2026, with their request URLs and retrieval times. They are retrospective histories, not a contemporaneously captured, immutable market tape. This limits assertions about exactly what an investor could have observed in real time. The replication materials preserve the raw responses used in the estimates, so later API changes need not silently change the results.

3.3 Coverage and selection

Table 1 separates the release calendar and matched prediction-market sample. The supplement reports futures coverage for the auxiliary exercise. Both Kalshi legs have some records for 42 releases. Thirty-four releases yield valid pre- and post-release matched prices and Polymarket observations. Requiring Kalshi endpoint record ages of at most thirty minutes leaves 31 events; tightening that requirement to five minutes leaves eight. The main sample thus reflects data availability and record-age choices. It is neither randomly selected nor representative by construction.

Table 1: Calendar and usable price histories

Release	Calendar	Delayed	Matched	Meetings
Employment	22	3	13	11
CPI	22	3	10	6
PPI	22	3	8	6

Calendar: releases with the original study's Polymarket series. Delayed: calendar releases classified as delayed. Matched: valid two-leg Kalshi midpoint and Polymarket thirty-minute changes with the primary age restrictions. Meetings: distinct FOMC contracts in that matched sample.

Table 2 characterizes the selection using Polymarket observations available for every calendar release. Excluded events have smaller average and median movements and occur farther from the next meeting. Of the 35 exclusions, 24 lack a history for at least one Kalshi leg, eight lack valid complete endpoints, and three fail the record-age limit. Delayed releases are retained at a rate of six of nine, compared with 25 of 57 regular releases. These descriptive differences do not identify agreement on the missing observations; weighting the retained sample cannot recover their unobserved Kalshi changes without additional assumptions.

Table 2: Matched and excluded calendar releases

Sample	Events	Mean $\lvert\Delta^P\rvert$	Median $\lvert\Delta^P\rvert$	Median P^P_-	Days to FOMC
Excluded	35	1.70	0.85	8.40	26.00
Matched	31	5.12	2.00	23.50	15.00

All 66 releases have a usable thirty-minute Polymarket change under the two-minute sample-age rule. Prices and changes are in percentage points. The pre-release column reports the median Polymarket change-price level, P^P_-. Days to FOMC is the calendar-day distance to the linked meeting. Matching status reflects Kalshi coverage and endpoint validity as well as the age restriction.

The primary matched sample contains 13 employment, ten CPI, and eight PPI releases, including six delayed releases. The underlying policy meetings are not 31 independent realizations: mul-

tiple announcements reprice the same terminal contract. I cluster the regression uncertainty and resample the disagreement statistics by FOMC meeting. With only 13 clusters, the resulting intervals remain approximate. Leave-one-meeting-out estimates and the full event-level results show how much individual contracts contribute.

Selection also differs within release type (Table 3). Employment movements average 8.50 points among matched dates and 4.23 among excluded dates; the corresponding pairs are 4.09 and 1.28 for CPI, and 0.92 and 0.43 for PPI. Polymarket's employment–CPI–PPI ordering therefore holds on the excluded dates as well. This diagnoses observed composition, not agreement with the missing Kalshi measurements.

Table 3: Selection within release type

Release	Excluded N	Matched N	Excluded mean	Matched mean	Std. diff.
CPI	12	10	1.28	4.09	0.81
Employment	9	13	4.23	8.50	0.56
PPI	14	8	0.43	0.92	0.39

Mean absolute thirty-minute Polymarket changes, in percentage points. Standardized difference divides matched minus excluded means by the square root of the average of their sample variances. It is descriptive, not a selection correction. The replication file additionally reports signed movements, pre-release levels, days to FOMC, delay status, calendar period, and top-quartile movement indicators.

4 Measuring agreement

4.1 Shared-clock changes

For each release time t_i, I take the last Polymarket sample strictly before the release and the last sample at or before $t_i + h$. Both must be no more than two minutes old. For Kalshi, the pre-release candle may end at t_i, because its interval precedes the announcement; the post-release candle must end by $t_i + h$. The benchmark requires both legs' endpoint records to be no more than thirty minutes old. The same information-time rule applies to bid and ask fields. The event change is

$$\Delta_i^v(h) = P_m^v(t_i + h) - P_m^v(t_i^-), \qquad v \in \{P, K\}, \tag{3}$$

with the superscript minus denoting the applicable pre-release observation rule. These are changes between selected records. Their realized time spans need not equal the nominal horizon, and each Kalshi leg can have different record times. Supplement L reports those spans and the cost of requiring synchronized legs. At short horizons the last available Kalshi endpoint can still precede the release; the horizon comparison describes the change observable in the archive, not necessarily a change between two newly updated quotes. The thirty-minute benchmark is a disclosed analysis choice, not a preregistered or estimated optimal horizon. I report one-, five-, fifteen-, sixty-, and 180-minute alternatives.

Agreement has several dimensions. The regression

$$\Delta_i^K(h) = \alpha_h + \beta_h \Delta_i^P(h) + u_{ih} \tag{4}$$

measures linear co-movement, using CR1 meeting-clustered standard errors, with variance multiplier $[G/(G-1)][(N-1)/(N-k)]$, a t reference distribution, and $G-1$ degrees of freedom. Here k is the number of fitted coefficients. For unit-slope hypothesis tests, the primary reported procedure is a restricted wild cluster bootstrap-t using every Rademacher sign vector across meetings (Cameron et al., 2008). Enumeration removes simulation error, not the finite-sample limitations of the bootstrap. It does not identify a structural response coefficient because both regressors and outcomes are market-based measurements. CR1 intervals remain descriptive uncertainty benchmarks; neither method supplies exact small-sample coverage. More directly, at the thirty-minute benchmark I report the distribution of $|\Delta_i^K(30) - \Delta_i^P(30)|$ and the fraction of events within one, two, and five percentage points. These thresholds describe practical accuracy tolerances; they are not market-efficiency nulls. Percentile intervals use 9,999 meeting-block bootstrap draws, preserving all releases associated with a sampled meeting and weighting events equally within each draw.

A nonsignificant test of $\beta_h = 1$ does not establish equivalence. Nor does a near-unit slope imply a small error for every event: a large common movement can generate a high correlation while smaller event-specific discrepancies remain important. The error distribution is therefore part of the main result, rather than an auxiliary robustness check.

4.2 Quote-implied ranges

Let $b_j(t)$ and $a_j(t)$ be Kalshi's recorded bid and ask for the lower ($j = L$) and upper ($j = U$) threshold contracts. The corresponding synthetic range for a change price is

$$B(t) = 100[1 - a_L(t) + b_U(t)], \qquad A(t) = 100[1 - b_L(t) + a_U(t)]. \tag{5}$$

The range of changes consistent with these endpoint quotes is

$$[B(t_i + h) - A(t_i^-),\ A(t_i + h) - B(t_i^-)]. \tag{6}$$

The Kalshi midpoint change is exactly the center of this interval. I ask whether the Polymarket change lies inside it; containing the Kalshi change itself supplies no corroboration. The calculation uses the same valid candles as the midpoint comparison. It does not truncate the synthetic quote range to $[0, 100]$ or assume execution in both legs at the displayed prices. Asynchronous legs, unavailable depth, fees, and absent historical Polymarket quotes preclude an executable arbitrage test. A wide interval may encompass another price simply because it is uninformative.

5 Cross-venue agreement

5.1 Co-movement and economically relevant errors

Figure 1 displays both the paired changes and their differences. The thirty-minute correlation is 0.940; the slope is 1.059, with a 95% interval of [0.943, 1.176]. The intercept is 0.347 percentage points, with interval [−0.668, 1.362]. The joint CR1 test of zero intercept and unit slope has $p = 0.527$; the restricted wild-bootstrap unit-slope test has $p = 0.456$. Neither is an equivalence test. The interval includes a unit slope, but it also admits deviations that could matter in applications. The leave-one-meeting-out slope ranges from 1.002 to 1.083 and the corresponding correlations range from 0.886 to 0.978. The meeting-block 95% interval for the correlation is [0.824, 0.989]. These are small-cluster descriptive estimates: omitting a meeting changes the correlation appreciably, even though it remains positive and high.

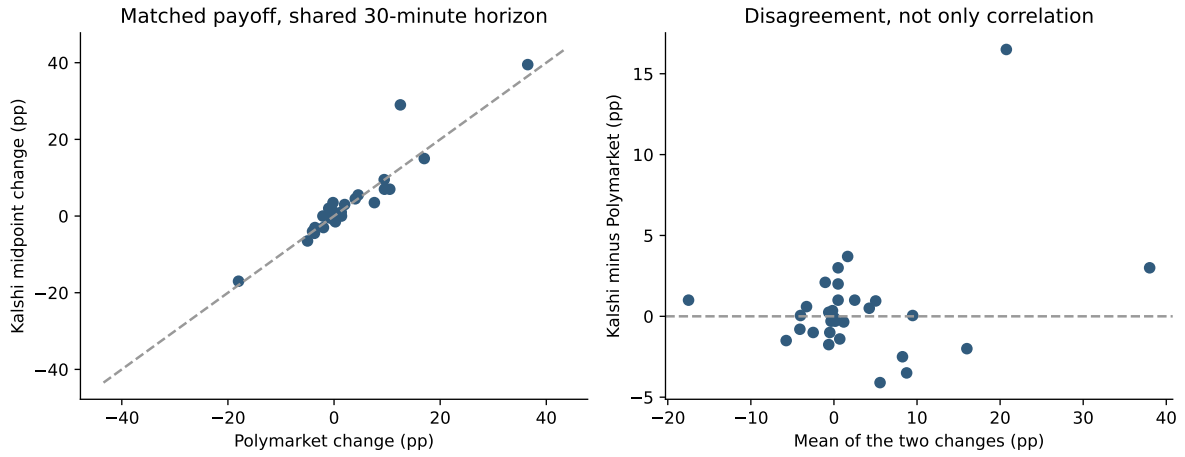


Figure 1: Matched changes and direct disagreement

Each point is a release in the 31-event benchmark sample. The left panel compares thirty-minute changes in the Polymarket complementary hold price and Kalshi's matched two-leg midpoint. The diagonal is equality. The right panel plots their difference against their mean, following Bland and Altman (1986). All events receive equal weight. The horizontal line is zero difference; no normal-reference limits are plotted.

Table 4 puts the discrepancies on the same percentage-point scale as the measured news. The median absolute Polymarket change in this sample is two points. Against that scale, a median disagreement of one point and a mean of 1.83 points are consequential. The 90th percentile of absolute disagreement is 3.5 points and the root mean squared difference is 3.44 points. Eighteen of the 31 paired changes agree within one point, 23 within two, and 30 within five. Supplement K reports event-level influence and regression-direction diagnostics. Omitting the largest discrepancy lowers MAE to 1.35 points and the root mean squared difference to 1.77. Exact sign agreement alone is not a stable summary when one recorded series is unchanged or either movement is close to zero.

The largest discrepancy occurs around the November 13, 2024 CPI release: Polymarket rises 12.5 points and the matched Kalshi midpoint rises 29 points. The endpoint records are at most one minute old, so the event also passes a five-minute record-age limit. Its synthetic pre-release quote

Table 4: How close are the thirty-minute changes?

Statistic	Estimate	95% block-bootstrap interval
Mean absolute disagreement (pp)	1.83	[1.03, 3.08]
Median absolute disagreement (pp)	1.00	[0.65, 2.00]
Share within 1 pp	0.58	[0.36, 0.79]
Share within 2 pp	0.74	[0.61, 0.86]
Share within 5 pp	0.97	[0.89, 1.00]

All estimates use the 31 matched releases and 13 FOMC contracts. The bootstrap resamples entire meetings 9,999 times; intervals describe sampling uncertainty conditional on this calendar and coverage. Shares use inclusive thresholds. They do not establish a population equivalence bound or independence across neighboring policy meetings.

range is 31 points wide and the post-release range is five points wide. Their summed width gives the implied change range [11, 47] points, containing the Polymarket change 1.5 points above its lower endpoint. The Kalshi change is the center by construction. This example illustrates uncertainty in the archived measurement rather than establishing a 16.5-point disagreement in underlying beliefs. It remains in the benchmark and in the age sensitivity because the wide quote is itself informative about the precision available to the researcher. Supplement D reports the prices and quote-implied change interval for every event.

5.2 Quotes qualify the interpretation

The median width of the synthetic Kalshi quote range is four points before the release and three points at the thirty-minute endpoint. Equation (6) contains the Polymarket change in 29 of 31 cases. Figure 2 shows these ranges alongside both point estimates. The median width of the change range is seven points, and the median ratio of range width to absolute Polymarket change is 3.75 among nonzero moves. These widths, rather than the containment count alone, define the resolution of this quote comparison.

Table 5 reports containment conditional on width. No range is two points wide or narrower. Across all 930 mismatched event pairs, 46.3% of Polymarket changes lie inside another event's range, compared with 93.5% for the actual pairs. Restricting mismatches to the same release type yields 46.7% containment; additionally matching days-to-meeting bins yields 48.9% across 90 pairs, versus 93.3% for the 30 eligible actual pairs. The bins are 0–14, 15–28, and more than 28 days. These are descriptive benchmarks, not permutation tests: events share meetings and differ even within strata. Intersecting each endpoint range with $[0,100]$ changes one of 62 ranges and leaves containment at 29. Neither calculation identifies efficient-price equality or the share of disagreement attributable to bid–ask effects.

5.3 Record age and horizon are part of the result

Supplement C reports five record-age limits. Requiring records no more than five minutes old leaves eight observations across seven meetings, with correlation 0.947 and mean absolute disagreement

Table 5: Containment and the width of quote-implied changes

Change-range width (pp)	Events	Contained	MAE (pp)
0–2	0	0	—
(2,5]	7	6	1.30
(5,10]	16	15	1.39
Above 10	8	8	3.19

Widths sum the two endpoint synthetic spreads. MAE is the absolute midpoint-change discrepancy in percentage points. The two noncontained Polymarket changes exceed their nearest range endpoint by 0.5 point: August 1 and November 20, 2025.

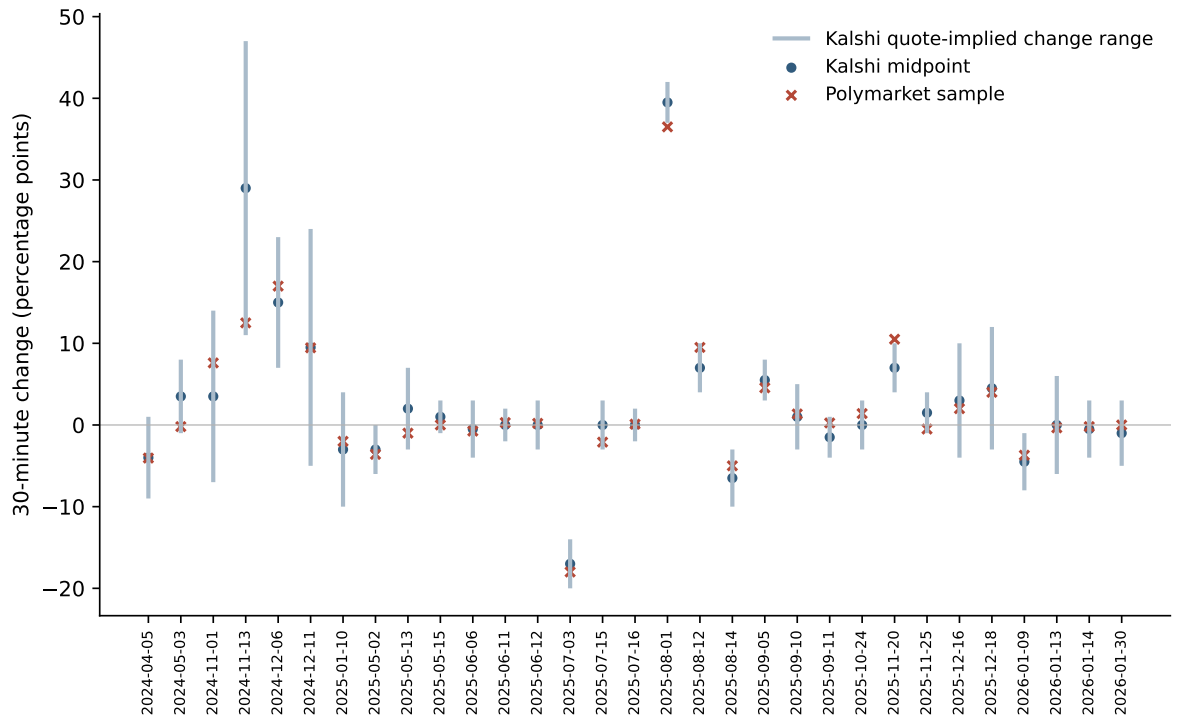


Figure 2: Point estimates within quote-implied change ranges

Vertical lines show equation (6), using recorded Kalshi endpoint bids and asks. Dots are the Kalshi midpoint changes; crosses are Polymarket changes. The ranges describe historical quote uncertainty, not executable trading opportunities. All 31 benchmark events are shown, including the largest discrepancy.

3.30 points. At four hours, the sample increases to 34 observations across 13 meetings, with correlation 0.939 and mean absolute disagreement 1.83 points. A stricter age rule does not mechanically improve estimated agreement, because it selects a different set of events. In particular, fresh records do not guarantee narrow quotes or small differences.

Horizon comparisons mix new records with price movement. In a balanced set of 21 releases linked to nine meetings, every Polymarket one-minute endpoint is a new post-release sample. Only six events have post-release records on both Kalshi legs; at five minutes, 13 do, and at thirty minutes all 21. Table 6 places these counts beside the agreement statistics and reports the stricter subsets separately. A new candle still does not guarantee a new underlying quote.

Table 6: Horizon sensitivity and post-release record coverage

Minutes	Corr., all 21	MAE, all 21	All post N	Corr., all post	MAE, all post
1	0.632	0.61	6	0.81	1.92
5	0.887	1.87	13	0.90	2.44
15	0.966	1.36	17	0.97	1.57
30	0.941	2.15	21	0.94	2.15
60	0.982	1.67	21	0.98	1.67
180	0.987	1.41	21	0.99	1.41

The first two statistics use the same 21 events and nine meetings. "All post" requires the selected post endpoint for Polymarket and both Kalshi legs to follow the release strictly; these columns change the event set. All rows retain the original record-age bounds. MAE is in percentage points. Short-horizon low error can reflect reused Kalshi records; the stricter subsets do not isolate a causal effect of freshness. Slope estimates and competing small-cluster tests are in the supplement.

The balanced-sample MAE rises from 0.61 to 1.87 points between one and five minutes, falls to 1.36 at fifteen, rises to 2.15 at thirty, and falls to 1.41 at three hours. The records cannot decompose this pattern into new information and measurement noise or identify an optimal horizon.

Hypothetically advancing or delaying all Polymarket timestamps by sixty seconds leaves all 31 benchmark events usable. Correlations range from 0.940 to 0.944 and MAE from 1.80 to 1.84 points. One-point classification disagreements fall from seven to six under either shift; five-point disagreements remain two. These are sensitivity calculations, not corrections to an unknown API convention. They show modest sensitivity of the pooled thirty-minute statistics but detectable sensitivity of a small-event indicator (Supplement N).

6 What changes when the researcher changes venue?

6.1 Event selection across the threshold range

Consider a researcher selecting announcements whose absolute response exceeds a threshold c. For the thirty-minute measurements, define

$$D(c) = N^{-1} \sum_i \mathbf{1}\{\mathbf{1}(|\Delta_i^P(30)| \geq c) \neq \mathbf{1}(|\Delta_i^K(30)| \geq c)\}. \tag{7}$$

Figure 3 reports this disagreement rate on a quarter-point grid from zero to twenty points; Table 7 gives reference cutoffs. The thresholds are exploratory sensitivity parameters, not externally validated definitions of policy relevance. Their analysis was developed after inspecting earlier results. Reporting the grid avoids selecting a single favorable cutoff, but does not make the exercise preregistered.

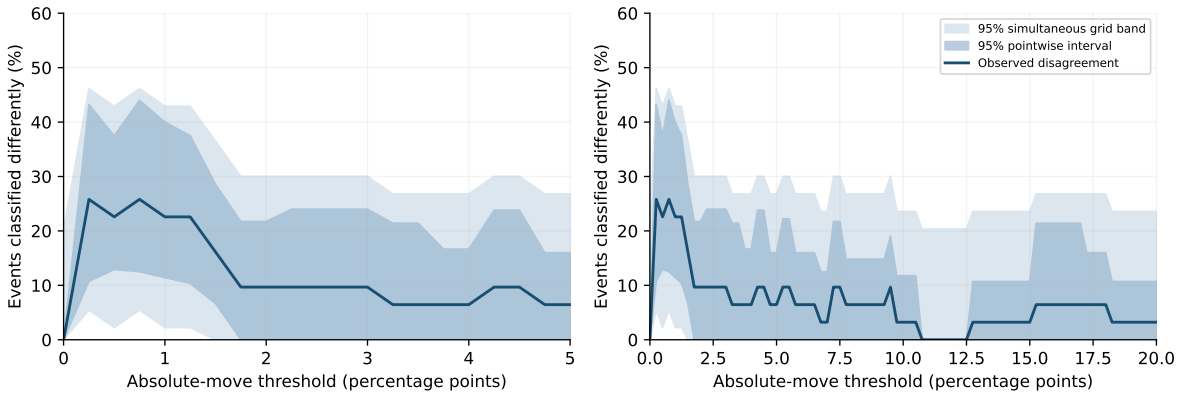


Figure 3: Venue choice and event-selection thresholds

Same 31 events throughout. The left panel enlarges the low-threshold region. Shading uses 9,999 meeting-block resamples: darker pointwise percentile intervals and a lighter simultaneous band formed from the 95th percentile of the maximum absolute bootstrap deviation over the displayed grid. Bands are approximate with 13 meetings, conditional on coverage, and do not correct for exploratory design. The underlying empirical curve is a step function; exact breakpoints accompany the grid in the replication files.

Table 7: Classification disagreement and quote ambiguity

Cutoff	PM only	K only	Disagree	Rate [95% CI]	Ambiguous K range
0.5	2	5	7	22.6 [12.9, 37.5]	22
1	2	5	7	22.6 [11.4, 40.0]	22
2	1	2	3	9.7 [0.0, 21.7]	23
3	0	3	3	9.7 [0.0, 24.0]	21
5	1	1	2	6.5 [0.0, 16.0]	17
10	1	0	1	3.2 [0.0, 11.8]	9

"PM only" and "K only" count events at or above the cutoff on just that venue. Rates and pointwise meeting-block intervals are percentages of 31. An ambiguous Kalshi range contains both a change below the absolute cutoff and one at or above it. No midpoint disagreement at the displayed cutoffs is forced by the entire Kalshi range. These intervals describe recorded quotes, not a confidence set for a latent true response.

At one point, the observed disagreement rate is 22.6%, with pointwise interval [11.4%, 40.0%]; at five points it is 6.5%, with interval [0.0%, 16.0%]. Seven and two different selections are observed facts, while these wide intervals limit extrapolation. The five-point switches are November 1, 2024 (7.60 points on Polymarket, 3.50 on Kalshi) and September 5, 2025 (4.55 and 5.50). The supplement lists all one- and five-point switches. The nearest endpoint is 1.50 points from the threshold in the first case and 0.45 in the second: switching is not confined to numerical equality at the cutoff.

Quote ranges further qualify these selections. Let $[L_i, U_i]$ be equation (6). Its implied absolute-movement range is $[0, \max(|L_i|, |U_i|)]$ if it spans zero and $[\min(|L_i|, |U_i|), \max(|L_i|, |U_i|)]$ otherwise. At one and five points, 22 and 17 such ranges permit both classifications. Every observed midpoint

switch is compatible with the Kalshi range. This is a partial description of recorded-price precision, not a decomposition of belief differences and microstructure noise; historical Polymarket spreads are unavailable.

Selecting a fixed number of large responses supplies a second application without assigning an economic meaning to a particular point threshold. The venues share two of their top three events, four of their top five, and seven of their top eight. The top-five difference replaces November 20, 2025 with December 11, 2024 when moving from Polymarket to Kalshi. For a requested top ten, Kalshi has a tie at its cutoff: retaining all ties gives eleven events versus ten on Polymarket, with nine shared. The absolute-movement Spearman rank correlation is 0.825, with meeting-block interval [0.516, 0.933]. These are observed changes in an event-selection rule, not evidence that either selected set is correct.

For the 14 events with $|\Delta^P| < 2$ points, the signed-change correlation is -0.19 [meeting-block interval -0.48, 0.35] and MAE is 1.12 points. Conditioning instead on $|\Delta^K| < 2$ gives correlation 0.08; requiring both to be below two gives 0.11. These outcome-selected groups restrict variation and cannot establish a breakdown of latent agreement. They do show why the pooled 0.94 correlation cannot validate small-event classification. Among the 20 events where both absolute changes are at least 0.5 point, 18 agree in sign; this outcome screen is another sensitivity, not a general directional-accuracy rate.

6.2 Release-type summaries and composition

A complementary application compares raw average movements across announcement types. I regress absolute thirty-minute changes on employment, CPI, and PPI indicators without an intercept, separately for each venue. A third regression uses the paired difference in absolute changes, preserving cross-venue covariance. Table 8 reports every pairwise release-type contrast, including uncertainty for each venue. These are raw movement summaries, not coefficients per unit of news surprise.

Table 8: Venue replacement in release-type means and contrasts

Release / contrast	Polymarket [95% CI]	Kalshi [95% CI]	K–P [95% CI]
Jobs	8.50 [2.54, 14.45]	8.38 [2.16, 14.61]	-0.11 [-1.58, 1.36]
CPI	4.09 [0.26, 7.91]	5.35 [-1.88, 12.58]	1.26 [-2.36, 4.89]
PPI	0.92 [-0.56, 2.39]	1.44 [-0.12, 2.99]	0.52 [0.23, 0.81]
Jobs minus CPI	4.41 [-1.10, 9.92]	3.03 [-5.57, 11.64]	-1.38 [-5.74, 2.98]
Jobs minus PPI	7.58 [2.59, 12.56]	6.95 [1.83, 12.07]	-0.63 [-2.17, 0.90]
CPI minus PPI	3.17 [-0.53, 6.86]	3.91 [-3.30, 11.12]	0.75 [-2.84, 4.33]

31 identical releases, 13 FOMC meetings; percentage points. All three pairwise contrasts are reported. Intervals use CR1 meeting clustering and t_{12}, including the paired-difference regression for K–P. They remain small-sample approximations and are not adjusted for multiple comparisons. Negative lower bounds for nonnegative means reflect the linear reference approximation. Differences are calculated before rounding. Signed-change estimates appear in Supplement N.

The observed means rank employment above CPI above PPI on both venues. Employment exceeds PPI by 7.58 points on Polymarket [2.59, 12.56] and 6.95 on Kalshi [1.83, 12.07]. The paired venue difference is -0.63 [-2.17, 0.90]. Employment–CPI contrasts are 4.41 and 3.03, but their respective

intervals [−1.10, 9.92] and [−5.57, 11.64] include zero. Thus neither the full population ordering nor equivalence of contrast magnitudes follows from the sample ordering. The paired intervals already show the venue effects the data cannot exclude; post-hoc power based on the observed effect would not resolve that uncertainty.

Medians reduce the employment–PPI gap to 3.85 points on Polymarket and 3.00 on Kalshi. The PPI mean absolute change is 0.52 points larger on Kalshi, with CR1 interval [0.23, 0.81]. This is a difference in measured magnitudes, not evidence of systematically higher change beliefs: its signed mean difference is −0.06 [−0.94, 0.83], and unequal measurement noise can itself change absolute-value means.

The ranking also survives deleting November 13, 2024 or August 1, 2025 individually and restricting to the four meetings with all three release types represented (19 events). In that common-meeting subset, employment–PPI differences are 9.78 and 10.17 points. Equal weighting of meeting-specific means within each release type also preserves the observed ordering. These checks diagnose composition and influence; they neither recover excluded Kalshi responses nor establish portability to another calendar. The supplement reports all subset means and deletion contrasts.

7 Implications for empirical event studies

The evidence favors using the two prediction venues as complementary measurements, with their disagreement reported as part of the empirical result. For an application concerned with a large policy repricing, the common component can be clear. For an application estimating a one-point response, the cross-venue discrepancy and quote range can be of the same order as the quantity of interest. The acceptable tolerance depends on the research question. A high correlation alone cannot determine it.

More generally, an announcement window identifies price movement near a release, not the marginal effect of a single headline number. As emphasized by Gürkaynak et al. (2020), a report can contain several relevant pieces of news; other news can arrive during a longer window. This paper consequently avoids interpreting absolute employment-versus-inflation movements as response coefficients to equally sized surprises. It also avoids using limited pre-release drift as proof that traders learned nothing before publication.

Three measurement choices should travel with any reported prediction-market event response. First, the terminal payoff should be written explicitly, including the distinction between a cut and any policy change. Second, the information-time convention should separate a sample timestamp from a bar's start or end and from the age of its underlying quote. Third, evidence of agreement should include direct errors, coverage, and quote uncertainty. Where an archive cannot supply quote-update age or historical spreads, the researcher should report that absence rather than imply the requirement has been met. Kalshi supplies candle bid/ask fields but not their underlying update ages; the Polymarket histories supply neither historical spreads nor update ages. These checks have a real

coverage cost, documented here, even when the calculations are simple.

8 Conclusion

Venue choice changes event selection even when raw release-type mean rankings agree. On 31 matched releases, the venues share four of their five largest movements. At a one-point rule, seven selections differ (22.6%; meeting-block interval 11.4–40.0%); at five points, two differ. Kalshi quote ranges permit both classifications for many more events, so midpoint switches should not be interpreted as differences in beliefs. For the aggregate employment–PPI contrast, the paired venue effect is -0.63 points with interval $[-2.17, 0.90]$, leaving economically relevant magnitude differences unresolved.

These conclusions apply to the retained 2024–2026 records and their observed coverage conditions. Retained releases have larger movements and lie closer to the linked meeting than excluded releases. Neither small-cluster intervals nor composition checks justify extrapolation to the missing observations, other platforms, or another policy and liquidity environment. An empirical application should report the events selected under each measurement, the full threshold sensitivity, and the available quote ranges alongside its aggregate estimates. More complete synchronized quotes are needed to distinguish record conventions from economic differences between venues.

Data availability

The analysis uses archived public Polymarket and Kalshi responses and previously acquired commercial CME data supplied through Databento. The accompanying replication materials document the endpoints, transformations, input hashes, and output tables. Redistribution of commercial inputs is subject to the data provider's terms. The code separates public-history acquisition from offline analysis. Public-history replication files accompany the manuscript; commercial input files are excluded from the shareable archive. The accompanying README identifies the required futures inputs and the parts that reproduce without them.

Declaration of generative AI use

During preparation of this work, the author used Claude Opus and OpenAI Codex to assist with drafting, coding, data checks, and revision. The author is responsible for the final submitted text, code, citations, and analyses.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Vega, C. (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review*, 93(1):38–62.

Bland, J. M. and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310.

Bürgi, C., Deng, W., and Whelan, K. (2026). Makers and takers: The economics of the Kalshi prediction market. Working paper, University College Dublin. January 2026 version.

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

de Jong, F. and Schotman, P. C. (2010). Price discovery in fragmented markets. *Journal of Financial Econometrics*, 8(1):1–28.

Diercks, A. M., Katz, J. D., and Wright, J. H. (2026). Kalshi and the rise of macro markets. Finance and Economics Discussion Series 2026-010, Board of Governors of the Federal Reserve System. Also NBER Working Paper 34702.

Ederington, L. H. and Lee, J. H. (1993). How markets process information: News releases and volatility. *Journal of Finance*, 48(4):1161–1191.

Eichengreen, B., Viswanath-Natraj, G., Wang, J., and Wang, Z. (2025). Under pressure? central bank independence meets blockchain prediction markets. Working paper, Social Science Research Network. SSRN Working Paper 5366862.

Fleming, M. J. and Remolona, E. M. (1999). Price formation and liquidity in the U.S. Treasury Market: The response to public information. *Journal of Finance*, 54(5):1901–1915.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gürkaynak, R. S., Kısacıkoğlu, B., and Wright, J. H. (2020). Missing events in event studies: Identifying the effects of partially measured news surprises. *American Economic Review*, 110(12):3871–3912.

Gürkaynak, R. S., Sack, B., and Swanson, E. T. (2005). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*, 1(1):55–93.

Gürkaynak, R. S. and Wolfers, J. (2007). Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk. In Frankel, J. A. and Pissarides, C. A., editors, *NBER International Seminar on Macroeconomics 2005*, pages 11–50. MIT Press.

Gürkaynak, R. S. and Wright, J. H. (2012). Macroeconomics and the term structure. *Journal of Economic Literature*, 50(2):331–367.

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

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the Fed Funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

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

Putniņš, T. J. (2013). What do price discovery metrics really measure? *Journal of Empirical Finance*, 23:68–83.

Snowberg, E., Wolfers, J., and Zitzewitz, E. (2013). Prediction markets for economic forecasting. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2. Elsevier.

Swanson, E. T., Wang, R., and Wu, Y. (2025). The effects of monetary policy on macroeconomic expectations: High-frequency evidence from traded event contracts. Working paper, University of California, Irvine; CEIBS; University of Florida. March 2025 draft.

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

Wolfers, J. and Zitzewitz, E. (2006). Interpreting prediction market prices as probabilities. Working Paper 12200, National Bureau of Economic Research.

Two Venues, One Policy Event:
Supplementary Analyses and Data Audit

This supplement documents auxiliary exercises and the construction checks supporting the main article. Its timing, forecast-score, and regime analyses are separate from the main venue-replacement estimand.

Two-year Treasury futures observations use the project's previously acquired Databento CME histories. One-second bars support the five-minute timing exercise; available one-minute histories extend the employment comparison to thirty minutes. Databento labels these bars by interval start.[1] I assign each close to interval end, adding one second or one minute as appropriate. This prevents the closing price of the first post-release bar from being treated as a price available at the release instant. Missing or old observations remain missing; no series is backward-filled.

A Auxiliary recorded-path comparison with an unmatched payoff

A.1 Directional half-crossings

For observed price path $p_i^v(s)$ and horizon H, define the terminal change $d_i^v(H) = p_i^v(H) - p_i^v(0^-)$ and the directional half-crossing

$$\tau_i^v(H) = \inf\{s > 0 : \operatorname{sign}(d_i^v(H))[p_i^v(s) - p_i^v(0^-)] \geq |d_i^v(H)|/2\}. \tag{S1}$$

The statistic is undefined for a zero terminal change. An excursion in the opposite direction does not count as completing half the eventual move. A crossing also does not imply that the price remains beyond the threshold; reversals are allowed. Changing H changes both the barrier and potentially its direction, so half-crossings at different horizons are not nested. Computation allows a 10^{-10} price-unit tolerance at exact half-barrier equality to prevent binary floating-point roundoff from deferring a crossing.

I compare matched five-minute horizons first and use only pairs for which both half-crossings are defined. The thirty-minute comparison is a separate employment-only extension. Eligibility requires nonzero terminal changes in both series, without a further magnitude cutoff. Nevertheless, a half-crossing is an ex-post path statistic indexed by its terminal change. Removing an outcome-size cutoff does not eliminate that dependence. The unnormalized price-change paths and a same-clock

[1] See the provider's schema definition: https://databento.com/docs/schemas-and-data-formats/ohlcv.

sensitivity provide complementary descriptions without claiming to recover a latent adjustment process.

Table S1 compares directional half-crossings on common event and endpoint definitions. The main five-minute comparison has 26 paired observations: 12 employment and 14 CPI releases. Median futures and Polymarket half-crossings are two and 70.5 seconds, respectively; the median within-event difference is 67 seconds. Every usable pair has a later Polymarket crossing. These pairs are a subset of the 30 events with valid futures coverage because a zero terminal change leaves the half-crossing undefined.

Table S1: Paired directional half-crossings

Release	Horizon	Pairs	ZT (s)	PM (s)	Median gap (s)
All	5	26	2.0	70.5	67.0
Employment	5	12	2.5	68.5	65.0
Employment	30	14	2.0	191.0	188.0
CPI	5	14	2.0	128.0	115.0

Horizon is in minutes; half-crossings and gaps are in seconds after the release. Each row uses pairs with both statistics defined, so marginal and paired sample medians need not coincide. Futures bars are timestamped at interval end. The thirty-minute sample uses employment releases only because the cached longer futures windows cover that exercise; no corresponding CPI estimate is available. The median gap is the median of paired differences, not the difference of medians.

The native data have different observation frequencies, so I repeat the five-minute comparison using futures prices observed only at Polymarket's actual sample timestamps. At each timestamp I use the most recent futures close, subject to the same sixty-second age limit. I define both terminal changes at the last shared sample before five minutes. The ordering survives: 28 pairs have median observed half-crossings of 8.5 seconds for futures and 73.5 seconds for Polymarket. Polymarket is later in 26 pairs and ties in two. The pair count changes because the shared terminal times can turn a small measured move into zero, making its crossing undefined.

The ordering persists on the shared observation grid. Historical price construction and unmatched payoffs still limit interpretation: a two-year future reflects a broader policy path and other components of bond pricing, whereas the binary contract concerns one decision. First trade times, common-trader activity, order-book revisions, and information shares remain outside the evidence.

The thirty-minute futures path uses one-second closes through minute 15 and one-minute closes thereafter. For employment releases with longer futures histories, the thirty-minute paired sample contains 14 events. The one-minute futures endpoint may be at most two minutes old in this extension; the pre-release baseline retains the sixty-second limit. Its median half-crossings are two seconds for futures and 191 seconds for Polymarket. The median within-event gap is 188 seconds. These values are conditional on nonzero measured terminal changes and usable histories; they should not be extrapolated to missing dates or to CPI at thirty minutes. Appendix E shows the shared-clock comparison and unnormalized Polymarket paths. The latter retain the scale of the observed moves rather than dividing every early movement by an ex-post terminal movement.

B Auxiliary fixed-horizon forecast comparison

A second venue is a useful benchmark, but it is not the truth. I therefore score the matched change prices against the 16 policy meetings' realized outcomes. To avoid treating adjacent days as independent realizations, each forecast horizon contributes at most one observation per meeting. Forecasts are taken at 20:00 UTC on the calendar day one, seven, or fourteen days before the meeting. These are calendar offsets, not exact multiples of 24 hours before the announcement.

The exercise uses separately archived hourly Polymarket histories and Kalshi quotes, with the same two-leg payoff mapping. A Polymarket sample must be no more than two hours old. I report Kalshi record-age limits of 24 and 72 hours. Prices outside $[0, 1]$ are excluded rather than clipped. The Brier loss is $(p_m - y_m)^2$, where y_m is one for a realized change and zero for a hold. Differences are computed on the same meetings for both venues, with equal meeting weights.

At the seven-day offset and 24-hour age bound, ten meetings are usable. Mean Brier losses are 0.00223 for Polymarket and 0.00359 for Kalshi. The difference, Kalshi minus Polymarket, is 0.00136; a paired-meeting 95% t interval is $[-0.00028, 0.00299]$. The one- and fourteen-day comparisons also have small point differences and intervals containing zero (Appendix G). The results neither establish equal forecasting accuracy nor estimate calibration reliably across the probability distribution. There are few terminal outcomes, and the scored meetings are selected by historical coverage. Very small losses near the meeting should also be interpreted in light of decisions that may already be highly anticipated.

Forecast scores answer a different question from the announcement regressions. Two venues may forecast the same eventual outcome accurately while producing different changes around one release. Conversely, similar event changes can occur even if their probability levels share a bias. Following the distinction between probabilistic scoring and cross-price comparison (Gneiting and Raftery, 2007), the scoring exercise prevents a high event correlation from being promoted into an unsupported claim about probabilistic accuracy or the absence of favorite–longshot bias.

With a symmetric two-hour record-age bound at all three forecast components, ten meetings remain at each horizon (Table S2). The seven-day comparison is unchanged from the 24-hour Kalshi-bound result. Scored changes versus holds are four versus six, five versus five, and four versus six at the one-, seven-, and fourteen-day offsets. The replication outputs also report realized ages, invalid-price exclusions, and clipping sensitivity. A realized-sample base rate is supplied only as an in-sample descriptive benchmark, not an available real-time forecast.

Table S2: Forecast scores with the same two-hour record-age bound

Days	Meetings	Changes	PM Brier	Kalshi Brier	K–P	95% CI
1	10	4	0.00040	0.00053	0.00014	[-0.00010, 0.00037]
7	10	5	0.00223	0.00359	0.00136	[-0.00028, 0.00299]
14	10	4	0.00693	0.00717	0.00024	[-0.00193, 0.00241]

At most one observation per meeting and horizon. Two hours applies to Polymarket and both Kalshi legs. Differences and paired intervals use the same definitions as the original age-grid comparison.

C Record-age and event-level sensitivity

Table S3 varies the maximum permitted Kalshi record age while keeping the thirty-minute horizon and matched payoff. It exposes the coverage cost of tighter restrictions. Quote-update age is not observed, so even the tightest bound is a record-age screen rather than a guarantee of synchronized underlying updates. The complete output also contains a cut-only comparison and transaction-close alternatives. Cut-only comparisons omit the hike leg and therefore do not define the same payoff without an additional assumption. They are not substitutes for the benchmark.

Table S3: Record-age sensitivity at thirty minutes

Max. age (min)	Events	Meetings	Correlation	Slope	95% CI	MAE
5	8	7	0.947	1.09	[0.88, 1.30]	3.30
15	21	9	0.945	1.10	[0.88, 1.33]	2.05
30	31	13	0.940	1.06	[0.94, 1.18]	1.83
60	32	13	0.939	1.04	[0.92, 1.16]	1.90
240	34	13	0.939	1.04	[0.92, 1.16]	1.83

MAE is in percentage points. Slopes and intervals use FOMC-meeting clustering. The four-hour row admits every complete valid matched observation available in the archived event windows; no future observation fills an earlier endpoint.

D Individual matched releases

Table S4: All benchmark event changes and quote-implied ranges

Date	Release	PM (pp)	Kalshi (pp)	Difference	Quote range
2024-04-05	Jobs	-4.05	-4.00	0.05	[-9.0, 1.0]
2024-05-03	Jobs	-0.20	3.50	3.70	[-1.0, 8.0]
2024-11-01	Jobs	7.60	3.50	-4.10	[-7.0, 14.0]
2024-11-13	CPI	12.50	29.00	16.50	[11.0, 47.0]
2024-12-06	Jobs	17.00	15.00	-2.00	[7.0, 23.0]
2024-12-11	CPI	9.45	9.50	0.05	[-5.0, 24.0]
2025-01-10	Jobs	-2.00	-3.00	-1.00	[-10.0, 4.0]
2025-05-02	Jobs	-3.60	-3.00	0.60	[-6.0, -0.0]
2025-05-13	CPI	-1.00	2.00	3.00	[-3.0, 7.0]
2025-05-15	PPI	0.00	1.00	1.00	[-1.0, 3.0]
2025-06-06	Jobs	-0.75	-0.50	0.25	[-4.0, 3.0]
2025-06-11	CPI	0.30	0.00	-0.30	[-2.0, 2.0]
2025-06-12	PPI	0.20	0.00	-0.20	[-3.0, 3.0]
2025-07-03	Jobs	-18.00	-17.00	1.00	[-20.0, -14.0]
2025-07-15	CPI	-2.10	0.00	2.10	[-3.0, 3.0]
2025-07-16	PPI	0.10	0.00	-0.10	[-2.0, 2.0]
2025-08-01	Jobs	36.50	39.50	3.00	[37.0, 42.0]
2025-08-12	CPI	9.50	7.00	-2.50	[4.0, 10.0]
2025-08-14	PPI	-5.00	-6.50	-1.50	[-10.0, -3.0]
2025-09-05	Jobs	4.55	5.50	0.95	[3.0, 8.0]
2025-09-10	PPI	1.35	1.00	-0.35	[-3.0, 5.0]
2025-09-11	CPI	0.25	-1.50	-1.75	[-4.0, 1.0]
2025-10-24	CPI	1.40	0.00	-1.40	[-3.0, 3.0]
2025-11-20	Jobs	10.50	7.00	-3.50	[4.0, 10.0]
2025-11-25	PPI	-0.50	1.50	2.00	[-1.0, 4.0]
2025-12-16	Jobs	2.00	3.00	1.00	[-4.0, 10.0]
2025-12-18	CPI	4.00	4.50	0.50	[-3.0, 12.0]
2026-01-09	Jobs	-3.70	-4.50	-0.80	[-8.0, -1.0]
2026-01-13	CPI	-0.35	0.00	0.35	[-6.0, 6.0]
2026-01-14	PPI	-0.20	-0.50	-0.30	[-4.0, 3.0]
2026-01-30	PPI	0.00	-1.00	-1.00	[-5.0, 3.0]

All changes and ranges are in percentage points. Difference is Kalshi minus Polymarket. Quote range is equation (6); it can be wide and is not an executable arbitrage bound. "Jobs" denotes the Employment Situation. This table retains all benchmark observations.

E Timing diagnostics

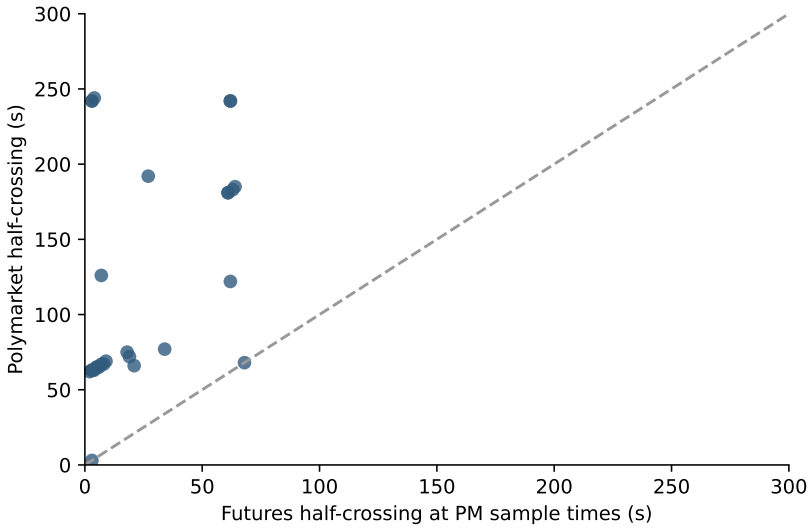


Figure S1: Five-minute half-crossings on the same observation clock

Both series are observed at actual Polymarket timestamps, and their endpoints use the last such timestamp before five minutes. The diagonal is equality. There are 28 valid pairs; 26 are above the diagonal and two are ties. Observed sampling times do not establish the update time of the historical price field.

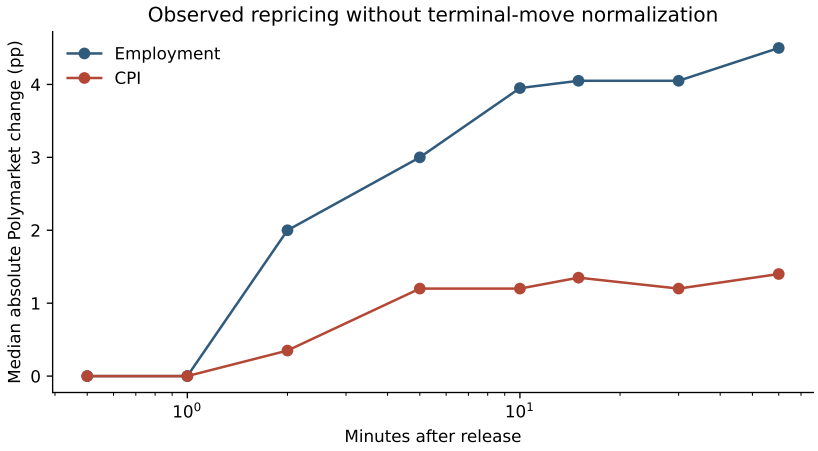


Figure S2: Unnormalized Polymarket changes in the futures-coverage sample

Medians of absolute changes, using the 15 employment and 15 CPI releases with valid futures coverage, subject to endpoint Polymarket availability. No terminal-move threshold is imposed. The horizontal axis is logarithmic. These are medians of event magnitudes, not a single monotone cumulative adjustment path.

F Futures observations excluded by the coverage rule

Table S5: Events without valid pre-release and five-minute futures endpoints

Date	Release	Pre age (s)	End age (s)
2024-06-12	CPI	Missing	235
2024-09-11	CPI	Missing	Missing
2024-12-11	CPI	Missing	63
2025-03-12	CPI	Missing	94
2025-06-11	CPI	Missing	Missing
2025-09-11	CPI	Missing	97
2025-12-18	CPI	Missing	Missing
2024-06-07	Jobs	Missing	101
2024-09-06	Jobs	518	40
2024-12-06	Jobs	Missing	Missing
2025-03-07	Jobs	Missing	158
2025-06-06	Jobs	Missing	Missing
2025-09-05	Jobs	52	149
2025-12-16	Jobs	362	662

A valid endpoint requires a close no more than sixty seconds old. Missing denotes no eligible past record, including empty histories. Bar closes are first assigned to interval end. The listed exclusions concern data coverage, not a minimum reaction size. The 30 events passing this rule can still have undefined half-crossings if their terminal change is zero.

G Fixed-horizon forecast scores

Table S6: Paired Brier losses, one forecast per meeting and horizon

Days	Age (h)	Meetings	PM Brier	Kalshi Brier	Difference	95% CI
1	24	12	0.00035	0.00046	0.00011	[-0.00008, 0.00030]
1	72	14	0.00035	0.00040	0.00005	[-0.00014, 0.00024]
7	24	10	0.00223	0.00359	0.00136	[-0.00028, 0.00299]
7	72	13	0.00183	0.00284	0.00101	[-0.00026, 0.00227]
14	24	13	0.00548	0.00572	0.00024	[-0.00135, 0.00183]
14	72	13	0.00548	0.00572	0.00024	[-0.00135, 0.00183]

Days are calendar-day offsets at 20:00 UTC. Age is the maximum Kalshi endpoint record age; Polymarket samples are at most two hours old. Difference is Kalshi minus Polymarket. Intervals use paired meeting differences and a t reference distribution with $N-1$ degrees of freedom. Positive differences favor Polymarket on this score. The small number of meetings and nonrandom coverage limit generalization. Comparisons across rows also change the sample.

H Complete release calendar

The calendar records the actual publication date and the next linked scheduled FOMC meeting. A release on an FOMC date precedes the policy announcement. "Matched" identifies the benchmark prediction-market sample. Delay classifications describe the original calendar and are not used to select the primary agreement sample or to identify a causal coefficient.

Table S7: Release calendar, first half

Date	Release	FOMC	Status	Matched
2024-04-05	Jobs	2024-05-01	ontime	Yes
2024-04-10	CPI	2024-05-01	ontime	No
2024-04-11	PPI	2024-05-01	ontime	No
2024-05-03	Jobs	2024-06-12	ontime	Yes
2024-05-14	PPI	2024-06-12	ontime	No
2024-05-15	CPI	2024-06-12	ontime	No
2024-06-07	Jobs	2024-06-12	ontime	No
2024-06-12	CPI	2024-06-12	ontime	No
2024-06-13	PPI	2024-07-31	ontime	No
2024-07-05	Jobs	2024-07-31	ontime	No
2024-07-11	CPI	2024-07-31	ontime	No
2024-07-12	PPI	2024-07-31	ontime	No
2024-08-02	Jobs	2024-09-18	ontime	No
2024-08-13	PPI	2024-09-18	ontime	No
2024-08-14	CPI	2024-09-18	ontime	No
2024-09-06	Jobs	2024-09-18	ontime	No
2024-09-11	CPI	2024-09-18	ontime	No
2024-09-12	PPI	2024-09-18	ontime	No
2024-10-04	Jobs	2024-11-07	ontime	No
2024-10-10	CPI	2024-11-07	ontime	No
2024-10-11	PPI	2024-11-07	ontime	No
2024-11-01	Jobs	2024-11-07	ontime	Yes
2024-11-13	CPI	2024-12-18	ontime	Yes
2024-11-14	PPI	2024-12-18	ontime	No
2024-12-06	Jobs	2024-12-18	ontime	Yes
2024-12-11	CPI	2024-12-18	ontime	Yes
2024-12-12	PPI	2024-12-18	ontime	No
2025-01-10	Jobs	2025-01-29	ontime	Yes
2025-01-14	PPI	2025-01-29	ontime	No
2025-01-15	CPI	2025-01-29	ontime	No
2025-02-07	Jobs	2025-03-19	ontime	No
2025-02-12	CPI	2025-03-19	ontime	No
2025-02-13	PPI	2025-03-19	ontime	No

Table S8: Release calendar, second half

Date	Release	FOMC	Status	Matched
2025-03-07	Jobs	2025-03-19	ontime	No
2025-03-12	CPI	2025-03-19	ontime	No
2025-03-13	PPI	2025-03-19	ontime	No
2025-04-04	Jobs	2025-05-07	ontime	No
2025-04-10	CPI	2025-05-07	ontime	No
2025-04-11	PPI	2025-05-07	ontime	No
2025-05-02	Jobs	2025-05-07	ontime	Yes
2025-05-13	CPI	2025-06-18	ontime	Yes
2025-05-15	PPI	2025-06-18	ontime	Yes
2025-06-06	Jobs	2025-06-18	ontime	Yes
2025-06-11	CPI	2025-06-18	ontime	Yes
2025-06-12	PPI	2025-06-18	ontime	Yes
2025-07-03	Jobs	2025-07-30	ontime	Yes
2025-07-15	CPI	2025-07-30	ontime	Yes
2025-07-16	PPI	2025-07-30	ontime	Yes
2025-08-01	Jobs	2025-09-17	ontime	Yes
2025-08-12	CPI	2025-09-17	ontime	Yes
2025-08-14	PPI	2025-09-17	ontime	Yes
2025-09-05	Jobs	2025-09-17	ontime	Yes
2025-09-10	PPI	2025-09-17	ontime	Yes
2025-09-11	CPI	2025-09-17	ontime	Yes
2025-10-24	CPI	2025-10-29	delayed	Yes
2025-11-20	Jobs	2025-12-10	delayed	Yes
2025-11-25	PPI	2025-12-10	delayed	Yes
2025-12-16	Jobs	2026-01-28	delayed	Yes
2025-12-18	CPI	2026-01-28	delayed	Yes
2026-01-09	Jobs	2026-01-28	ontime	Yes
2026-01-13	CPI	2026-01-28	ontime	Yes
2026-01-14	PPI	2026-01-28	delayed	Yes
2026-01-30	PPI	2026-03-18	ontime	Yes
2026-02-11	Jobs	2026-03-18	delayed	No
2026-02-13	CPI	2026-03-18	delayed	No
2026-02-27	PPI	2026-03-18	delayed	No

I Historical regime extension: a coverage audit

An exploratory extension constructs the no-change price from hourly Kalshi ladder records from late 2022 onward and compares absolute three-hour pre/post mean changes across employment, CPI, and PPI releases. The initial construction backward-filled missing early values and carried records forward without an age limit. It labeled periods using the subsequently identified last hike and first cut. Those labels are an ex-post cycle description, not information known at the beginning of each period.

An ablation that removes only backward filling changes the employment-versus-CPI tightening/easing interaction from 6.40 to 6.35 percentage points. The original normal-reference, meeting-clustered p-value moves from 0.098 to 0.107. This comparison shows that backward filling alone does not explain the point estimate; it does not validate unrestricted forward filling or a causal regime interpretation.

A separate reconstruction uses only past records and applies maximum ages to both legs at each sampling point. At least two valid pre-release and two valid post-release hourly observations are required. With one- or three-hour age limits, no tightening-period observations survive. With a 24-hour limit, the ex-post tightening group contains 13 releases linked to five meetings; with a seven-day limit it contains 18 releases linked to six meetings. These counts show the coverage–staleness tradeoff in the archived transaction-price histories. The caches also mix transaction closes with fallback fields, so they are not interchangeable with the full bid/ask candle responses used in the main analysis.

I do not promote this extension into an explanation of differences with Diercks et al. (2026). Their announcement regressions concern responses scaled by the news surprise and a different measure of policy expectations. A reversal in raw mean absolute movements need not imply a reversal in the response to an equally sized surprise. Extending the main result across regimes requires more complete historical quotes and a consistent outcome and news normalization. The replication audit retains the ablation, record-age grid, and a separate classification based on the last observed policy action so these choices remain inspectable.

J Reproduction details

The September 18 numerical audit corrected floating-point comparison at exact half-barrier equality. The thirty-minute Polymarket median crossing changed from 281.5 to 191 seconds; the paired gap changed from 253 to 188 seconds. This was an arithmetic correction, separate from the documented futures bar-end convention. Independent rational-arithmetic reconstruction verifies 57 Polymarket crossings across the native and shared-clock comparisons.

Public acquisition scripts archive full responses and metadata without overwriting source files. Offline analysis constructs the event panel and diagnostics, retaining endpoint ages and missing observations. The exhibit builder generates numerical tables and figures directly from those outputs.

The primary uncertainty calculation resamples FOMC meetings using seed 20260917. It preserves dependence among releases pricing the same decision but does not establish independence across neighboring meetings or policy regimes. The analysis is an exploratory revision following inspection of prior data and drafts. Its horizon grid, age restrictions, and practical error thresholds are reported transparently rather than described as preregistered choices.

The commercial futures histories are the principal replication access constraint. The public-price agreement and fixed-horizon forecast comparisons can be rebuilt from their archived public

responses. Rebuilding the futures timing results requires the corresponding licensed histories, including their timestamps and missing records. Independent archival quote data on both prediction venues would improve the ability to distinguish sampling conventions from genuine differences in price adjustment.

K Influence, regression direction, and precision

Table S9: Influence and movement-size diagnostics

Sample	Events	Correlation	MAE	RMS difference
All events	31	0.94	1.83	3.44
Omit Nov. 13, 2024	30	0.98	1.35	1.77
Omit two largest PM moves	29	0.85	1.82	3.51
PM absolute move below 2 pp	14	-0.19	1.12	1.56
PM absolute move at least 5 pp	9	0.93	3.79	6.00

All errors are in percentage points. Deletions are diagnostics, not replacement benchmark samples. Movement-size groups condition on the observed Polymarket outcome, which can itself affect the correlation; they are not independent tests of latent measurement reliability.

The November 13, 2024 observation contributes 29.0% of the sum of absolute discrepancies and 74.3% of squared discrepancies. Omitting it lowers MAE from 1.83 to 1.35 and the root mean squared difference from 3.44 to 1.77. Omitting the two largest absolute Polymarket movements instead lowers correlation to 0.854. Thus both common large movements and an influential discrepancy matter; the full-sample median absolute error remains one point.

Table S10: Regression direction changes the slope

Dependent / regressor	Slope	95% CI	Intercept	95% CI
Kalshi / PM	1.06	[0.94, 1.18]	0.35	[-0.67, 1.36]
PM / Kalshi	0.83	[0.62, 1.05]	-0.01	[-0.94, 0.92]

Same 31 releases and 13 meeting clusters. CR1 intervals use t_{12}. Both prices are measurements, so neither slope is structural or an equivalence statistic.

The reverse-regression interval also includes one. The classical homoskedastic interval for the original slope is [0.914, 1.205], compared with the clustered [0.943, 1.176]; cluster adjustment need not increase estimated variance. Under assumed Kalshi-to-Polymarket error-variance ratios of 0.25, one, and four, Deming slopes are 1.175, 1.135, and 1.090. These are sensitivities to imposed ratios, not estimates of the error variances. Correlation after demeaning by meeting is 0.928; the correlation between meeting means is 0.931.

The mean signed Kalshi-minus-Polymarket discrepancy is 0.492 points, with meeting-block interval $[-0.421, 1.663]$. The conventional normal-reference spread, mean difference plus or minus

1.96 sample standard deviations, is [−6.29, 7.27]. Its upper and lower endpoints are sensitive to the large CPI discrepancy; skewness, dependence, and sample selection prevent interpreting it as a validated 95% prediction interval for a future release. The difference-versus-mean plot borrows the comparison principle of Bland and Altman (1986), not a claim that their normal-reference coverage conditions hold here.

The bootstrap 95% upper percentile for the mean absolute discrepancy is 2.85 points. This supplies a scale for uncertainty about average error; it cannot certify that individual events lie within a tolerance. A one-point interchangeability claim would also conflict with the 13 observed discrepancies exceeding one point. Average-error and event-level tolerance statements concern different objects.

Table S11: Record-age sensitivity with meeting-block uncertainty

Age (min)	Events	Correlation [95% CI]	MAE [95% CI]
5	8	0.947 [0.800, 1.000]	3.30 [0.89, 8.01]
15	21	0.945 [0.821, 0.994]	2.05 [1.03, 4.36]
30	31	0.940 [0.827, 0.989]	1.83 [1.05, 3.04]
60	32	0.939 [0.836, 0.988]	1.90 [1.15, 3.09]
240	34	0.939 [0.841, 0.988]	1.83 [1.10, 2.96]

4,999 meeting-block resamples per row, percentile intervals, seed 20260917. These are small-cluster approximations. The rows have different event sets; differences across rows cannot isolate a causal effect of freshness.

L Actual endpoint times and payoff construction

The nominal horizon does not imply identical realized measurement spans. Across the 31 benchmark events, Polymarket's endpoint separation ranges from 1,798 to 1,802 seconds. The Kalshi cut leg spans 780–2,280 seconds and the hike leg 120–3,540 seconds. Both legs' post endpoints occur after the release in every benchmark event, but the maximum inter-leg timestamp difference at an endpoint is 1,620 seconds. Requiring both legs to lie within sixty seconds of each other at both endpoints leaves only five events across four meetings. Their correlation is 0.937 and MAE 4.12 points, a highly selected sample that still contains the large CPI discrepancy.

Table S12: Distinct records and genuinely post-release endpoints

Horizon	Balanced N	New PM	New cut	New hike	All post-release
1	21	21	9	7	6
5	21	21	17	14	13
15	21	21	20	18	17
30	21	21	21	21	21
60	21	21	21	21	21
180	21	21	21	21	21

The same balanced set of 21 events as the horizon table. "New" means the selected endpoint record differs from its pre-release record, not that the underlying quote was updated. "All post-release" additionally requires the Polymarket and both Kalshi post endpoints to be strictly later than the release. At one, five, and fifteen minutes, this stricter rule leaves six, 13, and 17 observations, so imposing it changes the sample.

No Polymarket endpoint is mechanically reused in this balanced comparison. Kalshi reuse is frequent at short horizons, especially in the hike leg. This is a feature of the as-of-record estimator, and it limits interpreting short-window changes as newly observed market reactions. The complete endpoint audit records times, spans, and inter-leg differences for each event.

All 124 selected leg-by-endpoint pairs have two present, noncrossed sides; none is a full-width zero-to-one quote. Missing and crossed pairs cannot enter under the stated midpoint rule. A candle's two closing fields need not have been updated at the same instant. One of 62 synthetic endpoint ranges extends above 100; capping the ranges leaves the containment count unchanged. This does not make the asynchronous bundle executable.

Dropping the hike leg changes the Kalshi event response by more than one point in three benchmark events, with a maximum difference of two points and a median absolute difference of zero. The correction is small on many easing-sample dates but not identically zero. Only two benchmark events satisfy the same record-age rules with transaction closes in both legs. On those two dates the average absolute midpoint-versus-transaction change difference is 2.75 points. That selected pair cannot identify how much of the full-sample cross-venue discrepancy comes from price-field choice; Polymarket's archived field is not independently verified as a transaction close.

Before the release, the median Kalshi-minus-Polymarket level difference is 1.50 points, with interquartile range $[-0.50, 2.00]$ and mean 0.95. These repeated contract observations describe the measured basis. They do not identify collateral costs, belief differences, or favorite–longshot bias.

M Sampling and numerical checks for the auxiliary timing exercise

Across all 30 events with futures endpoint coverage, the first Polymarket sample after the release arrives at a median four seconds; none lacks a sample in the first sixty seconds. The median within-event inter-sample gap is sixty seconds. Immediate repricing observed on those same sample schedules would therefore first become visible near the first sample, but an unknown age of the value

carried in each sample remains unresolved. Sample spacing alone does not identify the underlying repricing time.

At five minutes, two of the 30 events have zero measured Polymarket changes and two have zero futures changes, leaving 26 paired crossings. Requiring an absolute Polymarket move greater than 1.5 points leaves 15 pairs; their median half-crossings are two seconds for futures and 68 seconds for Polymarket, with a 65-second median paired gap. This outcome-dependent restriction is a sensitivity, not an improved representative sample.

An artificial sixty-second shift of all Polymarket timestamps illustrates the unresolved convention's importance. Advancing the stamps by sixty seconds gives 25 defined pairs and a median gap of nine seconds; delaying them gives 24 pairs and a median gap of 124.5 seconds. Each shift also changes the selected baseline and endpoint. Neither is an asserted correction to the API. The exercise demonstrates why the reported times cannot be interpreted as trader latency.

The five- and thirty-minute employment comparisons share eleven events. The latter splices one-second futures observations through minute 15 to one-minute observations thereafter. All fourteen futures crossings occur before that splice; two Polymarket crossings occur after minute 15. Consequently the two-second futures median is compatible with the thirty-minute endpoint convention. The longer-horizon statistic has a different barrier and sample and is not an estimate that adjustment becomes slower as the window expands.

The binary change probability has no globally valid hawkish/dovish sign mapping: an increased probability of a cut and an increased probability of a hike both raise it. Even in an easing sample, conditioning on a sign agreement with a two-year futures price would impose an additional economic interpretation. I therefore retain the reported ordering as a comparison of each instrument's own recorded movement and do not label it common-payoff information leadership. Historical trading tick sizes for every Polymarket observation are not available in these archives; observed decimal resolution is not a defensible substitute for such a tick schedule.

N Additional venue-replacement diagnostics

These analyses use the frozen public responses and were added after reviewing the original estimates. They diagnose the same selected sample; none supplies random assignment, a calibrated measurement-error model, or observations for missing Kalshi endpoints.

N.1 Named switches, ranks, and precision

Table S13: Events selected on only one venue at one and five points

Cutoff	Date	Release	PM change	K change	Kalshi range
1	2024-05-03	Jobs	-0.20	3.50	[-1.00, 8.00]
1	2025-05-15	PPI	0.00	1.00	[-1.00, 3.00]
1	2025-07-15	CPI	-2.10	0.00	[-3.00, 3.00]
1	2025-09-11	CPI	0.25	-1.50	[-4.00, 1.00]
1	2025-10-24	CPI	1.40	0.00	[-3.00, 3.00]
1	2025-11-25	PPI	-0.50	1.50	[-1.00, 4.00]
1	2026-01-30	PPI	0.00	-1.00	[-5.00, 3.00]
5	2024-11-01	Jobs	7.60	3.50	[-7.00, 14.00]
5	2025-09-05	Jobs	4.55	5.50	[3.00, 8.00]

Signed thirty-minute changes and Kalshi quote-implied change ranges, in percentage points. Selection uses absolute changes at or above the cutoff. All rows are retained in the benchmark. A midpoint switch does not imply that every change inside the quote range produces that switch.

Table S14: Overlap of large-response event sets

Requested top k	PM selected	K selected	Shared	Union
3	3	3	2	4
5	5	5	4	6
8	8	8	7	9
10	10	11	9	12

Descending absolute thirty-minute movements. All ties at the requested rank's cutoff are retained; thus the requested top ten contains eleven Kalshi events. No arbitrary date ordering breaks ties.

The integral of the classification-disagreement function over all nonnegative thresholds equals the mean absolute difference between the two absolute movements, $N^{-1}\sum_i ||\Delta_i^K| - |\Delta_i^P||$, or 1.708 points here. This identity follows because each event disagrees exactly between its two absolute magnitudes. It supplies a threshold-independent summary of the curve, but not an economically calibrated tolerance. The displayed grid ends at twenty points; the identity covers the full support.

Table S15: Outcome-conditioned size groups

Condition	Events	Meetings	Correlation [95% CI]	MAE
PM below 2	14	8	-0.19 [-0.48, 0.35]	1.12
Kalshi below 2	13	7	0.08 [-0.27, 0.47]	0.85
Both below 2	12	7	0.11 [-0.49, 0.62]	0.75
PM at least 5	9	5	0.93 [0.47, 1.00]	3.79
Both at least 0.5	20	10	0.94 [0.82, 0.99]	2.23

Meeting-block percentile intervals from 9,999 resamples within each fixed subgroup. Correlation is undefined in one draw for each of the "Kalshi below 2" and "Both below 2" groups and ten draws for "PM at least 5"; those draws are omitted from correlation quantiles. The groups are selected on measured outcomes and restrict their variance. They cannot identify error variances or latent reliability. MAE is in percentage points.

In the Polymarket-below-two group, four events have strictly opposite signed changes and seven have a zero on at least one venue. This separates reversals from unchanged records. These counts do not establish a sign-error rate against a true response. The relative-discrepancy output also reports $|\Delta_i^K - \Delta_i^P| / [(|\Delta_i^K| + |\Delta_i^P|)/2]$, leaving the ratio undefined when both are zero.

N.2 Composition and signed responses

Table S16: Release-type contrasts under composition and influence checks

Sample	Events	Jobs–PPI, PM	Jobs–PPI, K	Jobs–CPI, PM	Jobs–CPI, K
All 31	31	7.58	6.95	4.41	3.03
Common meetings	19	9.78	10.17	8.42	9.52
Omit Nov. 13, 2024	30	7.58	6.95	5.35	5.66
Omit Aug. 1, 2025	30	5.24	4.35	2.08	0.44
Omit Aug. 14, 2025	30	8.16	7.67	4.41	3.03

Point estimates in percentage points. "Common meetings" retains all events in the four meetings with at least one observation of every release type: six employment, seven CPI, six PPI. It equalizes meeting support, not the number or dates of releases. Individual deletions diagnose influence and are not preferred samples.

Equal weighting of the meeting-specific type means gives employment, CPI, and PPI absolute means of 7.92, 3.70, and 0.68 points for Polymarket and 7.52, 4.46, and 1.21 for Kalshi. A single grand mean per meeting would erase the release-type regressor; these type-specific summaries preserve the comparison. All-calendar Polymarket means are 6.75, 2.56, and 0.61 points, respectively. None of these computations reconstructs excluded Kalshi changes.

Table S17: Signed thirty-minute release-type means and contrasts

Release / contrast	Polymarket [95% CI]	Kalshi [95% CI]	K–P [95% CI]
Jobs	3.53 [-4.77, 11.83]	3.46 [-5.08, 12.01]	-0.07 [-1.52, 1.39]
CPI	3.40 [-0.93, 7.72]	5.05 [-2.28, 12.38]	1.66 [-2.04, 5.35]
PPI	-0.51 [-1.39, 0.38]	-0.56 [-2.13, 1.00]	-0.06 [-0.94, 0.83]
Jobs minus CPI	0.13 [-6.70, 6.96]	-1.59 [-11.51, 8.33]	-1.72 [-6.26, 2.82]
Jobs minus PPI	4.03 [-4.99, 13.06]	4.02 [-5.70, 13.75]	-0.01 [-2.10, 2.09]
CPI minus PPI	3.90 [-0.83, 8.63]	5.61 [-1.89, 13.11]	1.71 [-1.85, 5.27]

Same 31 events, three release indicators, and CR1 meeting-clustered t_{12} intervals as the absolute-change table. All differences precede rounding. A signed increase raises the measured probability of either a cut or a hike; it is not a universally hawkish or dovish response.

The full balance output reports matched and excluded means, mean differences, and standardized differences both pooled and within release type for movement, pre-release level, days to FOMC, delayed status, 2024 membership, and a full-calendar upper-quartile movement indicator. Standardization is descriptive and does not solve nonrandom coverage.

N.3 Timestamp sensitivity, quotes, and regression diagnostics

Table S18: Thirty-minute results under hypothetical Polymarket timestamp shifts

PM timestamp shift (s)	Events	Correlation	MAE	Flips at 1 pp	Flips at 5 pp
-60	31	0.94	1.84	6	2
0	31	0.94	1.83	7	2
60	31	0.94	1.80	6	2

All 31 benchmark events remain eligible. Both Polymarket endpoints are reselected after shifting all timestamps, retaining the two-minute age rule; Kalshi is unchanged. Negative means earlier stamps. This is not an asserted correction or a pair of independent measurements of a fixed efficient price.

Table S19: Deming slopes conditional on imposed error-variance ratios

Assumed error-variance ratio	Slope	95% interval
0.25	1.17	[0.87, 1.58]
1.0	1.13	[0.82, 1.47]
4.0	1.09	[0.77, 1.30]

Kalshi-to-Polymarket error-variance ratios are imposed, not estimated. Percentile intervals use 9,999 meeting-block resamples. These conditional sensitivities do not identify latent reliability or establish equivalence.

An exploratory regression of absolute venue disagreement on the sum of endpoint synthetic quote widths has slope 0.248 [CR1 interval 0.123, 0.374] and $R^2 = 0.381$. Quote widths and midpoint differences share underlying measurements and influential events. Their association cannot identify the fraction of discrepancy caused by liquidity or separate a latent news signal from noise. A two-minute

Kalshi record-age limit leaves four events in four meetings, with MAE 5.10 points; this is another highly selected sensitivity, not a stale-quote placebo or proof of a structural wedge.

Table S20: Balanced-horizon slopes retained as diagnostic comparisons

Horizon (min)	Events	Correlation	Slope	95% CI	MAE
1	21	0.632	3.11	[-1.79, 8.01]	0.61
5	21	0.887	0.68	[0.49, 0.86]	1.87
15	21	0.966	1.02	[0.77, 1.28]	1.36
30	21	0.941	1.09	[0.87, 1.31]	2.15
60	21	0.982	1.09	[0.91, 1.28]	1.67
180	21	0.987	0.97	[0.85, 1.08]	1.41

Same 21 releases and nine meetings. The main article reports post-release endpoint counts and stricter subsets alongside horizon errors. At five minutes CR1 excludes a unit slope, while the primary restricted wild-bootstrap unit-slope test gives $p = 0.113$. The bootstrap does not reject at five percent; this neither proves a unit slope nor makes CR1 an invalid descriptive benchmark.

The difference-versus-mean figure in the main article plots zero difference, not estimated limits of agreement. The supplementary normal-reference spread uses the raw event-level sample standard deviation; it does not substitute a clustered standard error for cross-event dispersion. Meeting-block intervals quantify uncertainty in the mean discrepancy separately.

The event-level contract crosswalk records the Polymarket market and token IDs, prevailing upper target bound, Kalshi lower and upper strikes and tickers, full archived primary and secondary rules, and terminal settlements. The existing Polymarket metadata file preserves its full resolution description. These files permit inspection of settlement definitions and the grid assumption rather than relying only on similarly worded market titles.

References

Bland, J. M. and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310.

Diercks, A. M., Katz, J. D., and Wright, J. H. (2026). Kalshi and the rise of macro markets. Finance and Economics Discussion Series 2026-010, Board of Governors of the Federal Reserve System. Also NBER Working Paper 34702.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.