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

Orhan Torul

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

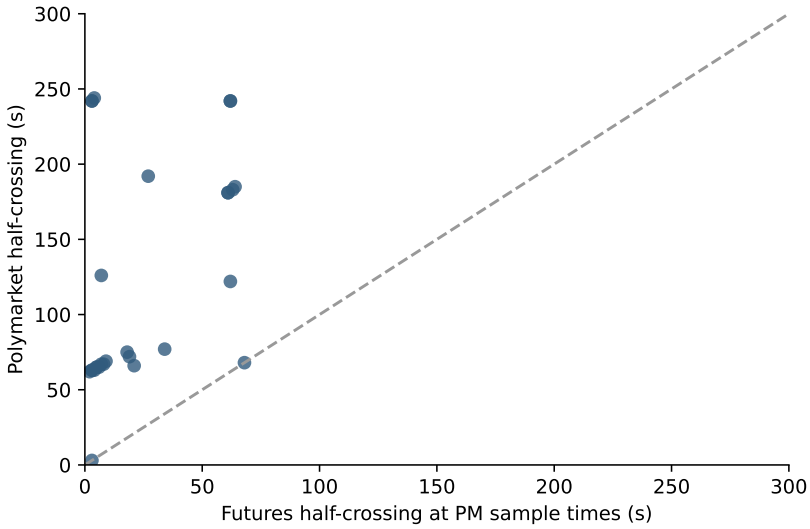


Figure S1: Five-minute half-crossings on the same observation clock

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.

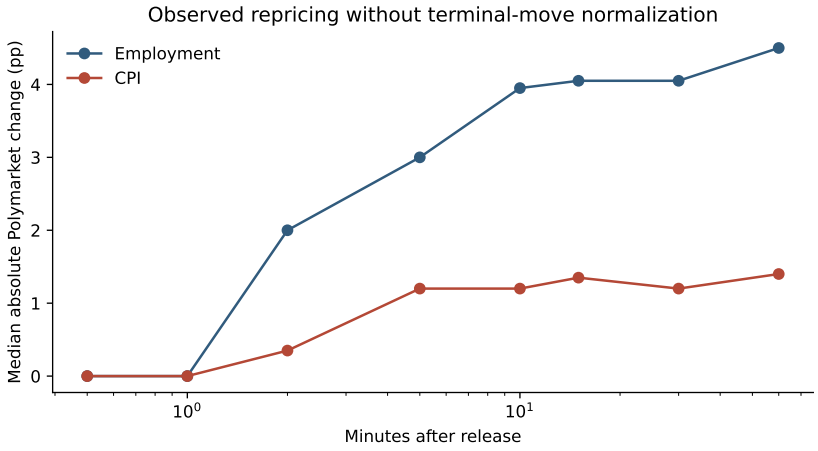


Figure S2: Unnormalized Polymarket changes in the futures-coverage sample

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.