mathematics

MDPI

Article

Two-Way Price Discovery Between a Cryptocurrency-Based Prediction Market and Federal Funds Futures: Evidence from FOMC Decisions

Marcus Alexis Giannopoulos and Giovanni Millo *

Department of Economics, Business, Mathematics and Statistics "Bruno de Finetti" (DEAMS), University of Trieste, 34127 Trieste, Italy; marcusalexis.giannopoulos@studenti.units.it
* Correspondence: giovanni.millo@deams.units.it

Abstract

Expectations about policy rates play a central role in financial and insurance markets. This paper asks whether Polymarket, an offshore cryptocurrency-collateralized prediction market, contributes independently to price discovery about Federal Open Market Committee (FOMC) interest-rate decisions or whether its prices merely follow the established Federal Funds Futures (ZQ) market. For twenty-nine FOMC meetings between March 2023 and September 2026, Polymarket event-contract prices and CME Federal Funds Futures prices are each reduced to a meeting-level expected basis-point change on a common time grid and compared along three dimensions: long-run co-movement, forecast accuracy, and the direction of price discovery. The two series are cointegrated in a majority of the meetings and in nearly all of the later more liquid sub-sample, where the long-run relationship is close to one-to-one; the forecast tests do not detect a statistically significant accuracy difference between the markets. Lead–lag tests based on pooled Granger causality and a single-equation error-correction model reveal frequency-dependent predictive relationships. Under strict freshness filtering, a wild cluster bootstrap supports Polymarket-to-futures prediction at 10, 30 and 60 min and the reverse direction at 60 min but neither direction at one minute. Because this longer-horizon pattern cannot be cleanly separated from a timing effect in which both markets respond to common public information at different speeds, the results stop short of establishing that Polymarket leads outright. The findings are consistent with Polymarket being a liquidity-dependent complement to Federal Funds Futures, but do not establish an independent contribution to price discovery.

Keywords: price discovery; prediction markets; cointegration; Granger causality; error-correction model; Federal Funds Futures; FOMC; market microstructure

MSC: 91B84

Check for updates

Academic Editor: Andrea Scozzari

Received: 29 September 2026
Revised: 21 October 2026
Accepted: 27 October 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Expectations about monetary policy play a central role in modern financial markets. In particular, predicting the policy rate is important in insurance markets because interest rates directly affect insurers' investment returns and the valuation of their liabilities. Accurate forecasts help insurers to make better decisions about asset allocation, pricing, and capital management. They also support the management of duration and interest-rate risk, particularly for life insurers with long-term obligations. Furthermore, expectations

about future policy rates can influence insurers' profitability, solvency, and overall financial stability.

Decisions made by the Federal Open Market Committee (FOMC) directly determine the target range for the federal funds rate, but their effects extend much further: bond yields, equity prices, exchange rates, credit conditions, and broader macroeconomic expectations all depend not only on what the Federal Reserve does today but what market participants expect it to do in the future. When markets correctly anticipate a policy decision, the announcement itself has only a limited effect on asset prices; when the decision differs from expectations, the surprise can produce sharp adjustments across financial markets [1].

Expectations about future Federal Reserve policy have traditionally been inferred from interest-rate derivatives, in particular the Federal Funds Futures traded on the Chicago Mercantile Exchange (CME) under the ZQ ticker. These contracts settle on the average effective federal funds rate over a calendar month, and the CME Group documents a methodology (FedWatch) for converting contract prices into implied meeting-level probabilities.

A newer class of markets has emerged as a potential complementary source of information: prediction markets, whose contracts settle directly on the outcome of a specific event rather than requiring indirect interpretation. In the context of monetary policy, traders can buy and sell contracts that are linked directly to specific FOMC outcomes—a cut, a hold, or a hike of a given size—from which a probabilistic forecast of each meeting can be recovered without the intermediate decomposition that a futures-based measure requires.

This paper focuses on Polymarket, one of the most prominent prediction-market platforms in recent years. Polymarket is a cryptocurrency-based decentralized exchange whose contracts are collateralized in a US-dollar stablecoin and settled without the oversight of a domestic derivatives regulator; throughout the sample period it was closed to US persons under a 2022 settlement with the Commodity Futures Trading Commission (CFTC). This stands in contrast to CFTC-regulated venues such as Kalshi, which are open to US retail and institutional participants. Prediction markets gained particular prominence during the 2024 US presidential election, where they outperformed polling aggregates and Polymarket-implied probabilities were found to predict subsequent poll movements rather than the reverse [2]. This paper asks whether an analogous informational advantage holds in a very different setting: a recurring scheduled monetary-policy decision that a deep and established futures market already prices continuously.

Following [3], price discovery is understood here as the process by which market prices incorporate new information about an underlying factor, in this case the FOMC's forthcoming decision. A market is said to contribute to price discovery if its price changes lead, rather than merely follow, those of a comparison market. The central research question is whether Polymarket contributes independently to price discovery about FOMC decisions or its prices primarily follow information that is already reflected in Federal Funds Futures. This is decomposed into three testable sub-questions: whether the two market-implied measures are aligned in the long run; whether one produces more accurate forecasts of the realized decision; and whether changes in one market tend to precede changes in the other.

The empirical strategy proceeds in three steps. First, Polymarket event-contract prices are converted into implied probabilities across possible FOMC outcomes and reduced to an expected basis-point change per meeting. Second, Federal Funds Futures prices are transformed into an analogous meeting-level expectation, correcting for the fact that a single contract settles on a monthly average rate that blends the periods before and after the decision. Third, the two series are compared through an Engle–Granger cointegration test, Mincer–Zarnowitz and Diebold–Mariano forecast-accuracy tests, and pooled Granger-causality and error-correction tests, the latter supported by a freshness filter that corrects for the two markets' markedly different trading frequencies.

The contribution of this paper is to extend the recent literature on macroeconomic prediction markets, so far concentrated on the regulated institutionally accessible Kalshi platform [4,5], to an offshore cryptocurrency-collateralized venue with a structurally different and highly concentrated participant base [6,7] and to benchmark it directly against the traditional interest-rate derivative that is used to measure the same expectation. The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 sketches the necessary institutional background to interpret the two price series. Section 4 describes the data and the construction of the two market-implied measures and the econometric methodology. Section 5 presents the empirical results. Section 6 discusses their interpretation and limitations, and Section 7 concludes.

2. Related Work

2.1. Prediction Markets and Information Aggregation

Prediction markets trade contracts whose prices are designed to reflect the probability of a specific future event. Refs. [8,9] sets out the theoretical framework in which informed traders arbitrage the gap between their beliefs and the traded price until prices converge to the underlying probability so that aggregate market prices act as an analogue of the efficient-markets hypothesis for probability forecasts. In practice this connection is looser than the theory suggests. Ref. [10] argues that most prediction-market contracts are incomplete markets with no replicating instrument so that prices converge to a range rather than a single correct value; analyzing over 290,000 contracts, that study finds prices systematically above true probabilities, with the gap widening in less-liquid and longer-duration markets and narrowing as contracts approach resolution. Ref. [11] traces an analogous favorite-longshot bias on Kalshi to a maker–taker microstructure in which informed traders post prices with a positive expected return that less-informed traders accept, with accuracy improving markedly as resolution nears. This concentration of information among a small set of participants is confirmed directly for Polymarket by [7], who find that roughly 3% of accounts generate the bulk of price discovery, and by [6], who analyze 588 million Polymarket trades and find that the top 1% of users capture 76.5% of the total profits while 69% of users lose money overall. For the present setting this matters directly: whether Polymarket leads Federal Funds Futures is, in effect, a question of whether this informed minority prices FOMC decisions faster than the institutional participants who dominate the futures market.

2.2. Prediction Markets in Macroeconomic Forecasting

Ref. [4] provides the most comprehensive evaluation to date of a regulated prediction market's macroeconomic forecasting properties, showing that Kalshi achieves comparable or better point-forecast accuracy than the Bloomberg consensus and the FRBNY Survey of Market Expectations across FOMC, CPI, GDP, and payrolls contracts while offering continuous intraday updating and a full outcome distribution rather than a periodic point estimate. Ref. [5] uses the same continuous updating to isolate the response of macroeconomic expectations to FOMC announcements and finds little evidence for the "Fed information effect" documented in lower-frequency survey data, attributing earlier findings of a contrarian response to the coarseness of the available forecast data rather than genuine information revealed by the policy action. Both studies, however, are centered on Kalshi, a CFTC-regulated US-accessible venue with institutional participation; whether the same informational content extends to Polymarket, whose structure differs on essentially every relevant dimension, is unexamined.

2.3. Federal Funds Futures as a Benchmark

Federal Funds Futures remain the longest-standing market-based gauge of Federal Reserve policy expectations. Ref. [12] shows that futures prices can decompose a rate change into expected and unexpected components, with only the latter moving asset prices, a result that is reinforced by [1], who confirm that the futures market is an efficient near-unbiased predictor of near-term policy outcomes. The specific methodology used to translate futures prices into meeting-level probabilities, the CME FedWatch tool, has received comparatively little scrutiny; Ref. [13] provides the first systematic assessment, finding 88% accuracy thirty days before a meeting (against 75% for a single-contract approach), which is attributable to FedWatch's binary-tree treatment of the post-2008 corridor system and positive economic value from FedWatch-based trading strategies. That methodology is replicated in Section 4.3 to recover meeting-level expectations from ZQ prices.

2.4. Price Discovery Across Markets

When an asset or event is priced simultaneously in more than one venue, prices share a common implicit efficient value to which each market contributes some share of price discovery [3]. This mechanism follows from market microstructure theory in which information enters prices through the trades of informed participants [14,15] and deeper lower-cost markets attract more such participants and adjust faster. Cross-listed equities confirm that price discovery concentrates where trading activity concentrates [16,17], while frictions such as capital controls or participation constraints can sustain persistent price gaps between markets pricing formally identical assets [18,19]. Polymarket and Federal Funds Futures differ in instrument design, collateral, regulation, trading hours, and participant base to a degree that makes direct arbitrage between them effectively impossible, so any observed co-movement or lead–lag pattern reflects information flow rather than mechanical linkage. The first direct evidence of price discovery within modern prediction markets comes from [2], who find that the Polymarket returns on the 2024 US presidential election significantly predict the subsequent poll levels (not the reverse), the law of one price fails between Polymarket and Kalshi, and Polymarket returns predict Kalshi returns but not vice versa, with the lead being the strongest during periods of high Polymarket volume.

2.5. Research Gap and Hypotheses

The literature establishing prediction markets as informative about macroeconomic events and benchmarking them against Federal Funds Futures is confined to Kalshi [4,5]; the only direct evidence that Polymarket leads another venue comes from the 2024 election markets [2], not from a recurring macroeconomic setting in which a deep professionally traded futures market already aggregates expectations. This paper addresses that gap by testing three hypotheses.

Hypothesis 1. *The Polymarket-implied and ZQ-implied expected rate changes reflect the same underlying expectation and are cointegrated.*

Hypothesis 2. *Both market-implied measures are rational forecasts of the realized decision, and neither is systematically more accurate than the other.*

Hypothesis 3. *Polymarket contributes independently to price discovery in that its price changes lead changes in Federal Funds Futures rather than merely following them.*

3. Institutional Background

The FOMC sets the federal funds rate as a 25-basis-point target range, communicated at eight scheduled meetings per year. The rate at which overnight federal funds transactions actually occur, the effective federal funds rate (EFFR), is a market outcome that is published daily by the Federal Reserve Bank of New York rather than a value the Committee sets directly; it is confined to the target range by a corridor system in which the Federal Reserve stands ready to lend and borrow in unlimited quantity at the range's upper and lower bounds, respectively, so arbitrage against these standing facilities keeps EFFR inside the range.

Federal Funds Futures, traded on the CME under the ZQ ticker since 1988, settle on 100 minus the average EFFR over the contract month; a contract written on a month containing an FOMC meeting therefore blends the pre- and post-decision rate, so a single price cannot be read directly as a meeting-level expectation (the inversion recovering that expectation is given in Section 4.3). Polymarket, by contrast, lists each FOMC outcome as a separate binary contract collateralized in a US-dollar stablecoin and settled via a decentralized optimistic-oracle mechanism (with disputes escalated to a token-holder vote) so that its price is read directly as an outcome probability at the cost of a resolution-related settlement risk that is absent from the futures contract. The two markets thus price the same underlying event through opposite designs, and, as Section 2 notes, differ on every dimension that is relevant to price formation: settlement mechanism, collateral, regulatory status, trading hours, and participant base. Because Polymarket was closed to US persons throughout the sample period, the large regulated institutions that dominate ZQ trading are structurally excluded from it, while Polymarket's own participation is concentrated in a small crypto-native informed minority [6,7]. No arbitrage links the two series, so whether Polymarket leads or merely follows the futures market is an empirical question rather than one settled by market structure alone.

4. Materials and Methods

4.1. Data Sources and Alignment

The analysis combines three data sources: Polymarket event-contract prices, CME Federal Funds Futures tick data, and Federal Reserve records of FOMC meeting dates and realized decisions. The sample covers 29 FOMC meetings from March 2023 to September 2026. Polymarket last-trade data were obtained via a publicly hosted subgraph indexing the platform's on-chain transaction history; ZQ tick data were obtained from a commercial market-data vendor at sub-second resolution (see the Data Availability Statement). All the timestamps are converted to a common time zone, the ZQ series is downsampled to one-second precision, and the two series are aligned on a regular one-second grid by forward filling. Because forward-filled values can misrepresent a stale quote as fresh information, binary update indicators are retained for each series to record whether an observation at a given timestamp corresponds to a genuine trade, which underlies the freshness filter described in Section 4.7. Because Polymarket trades continuously while CME futures trade only during exchange hours, the analysis is restricted to common CME trading time.

Figure 1 shows total Polymarket notional volume by FOMC meeting across the sample period.

4.2. Construction of Polymarket-Implied Expected Rate Changes

For a given FOMC meeting, each possible outcome k is associated with a rate change Δ_k (in percentage points), traded as a separate binary contract whose "Yes" price is read as the market-implied probability of that outcome. Trades on the "No" side are converted

to a "Yes"-equivalent price as one minus the observed price, and trades within the same second for a given outcome are collapsed into a volume-weighted average. Because independent trading of each outcome's contract leaves the sum of "Yes" prices only approximately equal to one, the prices are normalized: letting q_k denote the "Yes"-equivalent price of outcome k,

$$p_k = \frac{q_k}{\sum_j q_j}, \tag{1}$$

and the Polymarket-implied expected rate change for meeting m is

$$E_t^{\mathrm{PM}}(\Delta_m) = \sum_k p_k \Delta_k. \tag{2}$$

Additional details on the construction of the Polymarket measure are provided in Appendix C.

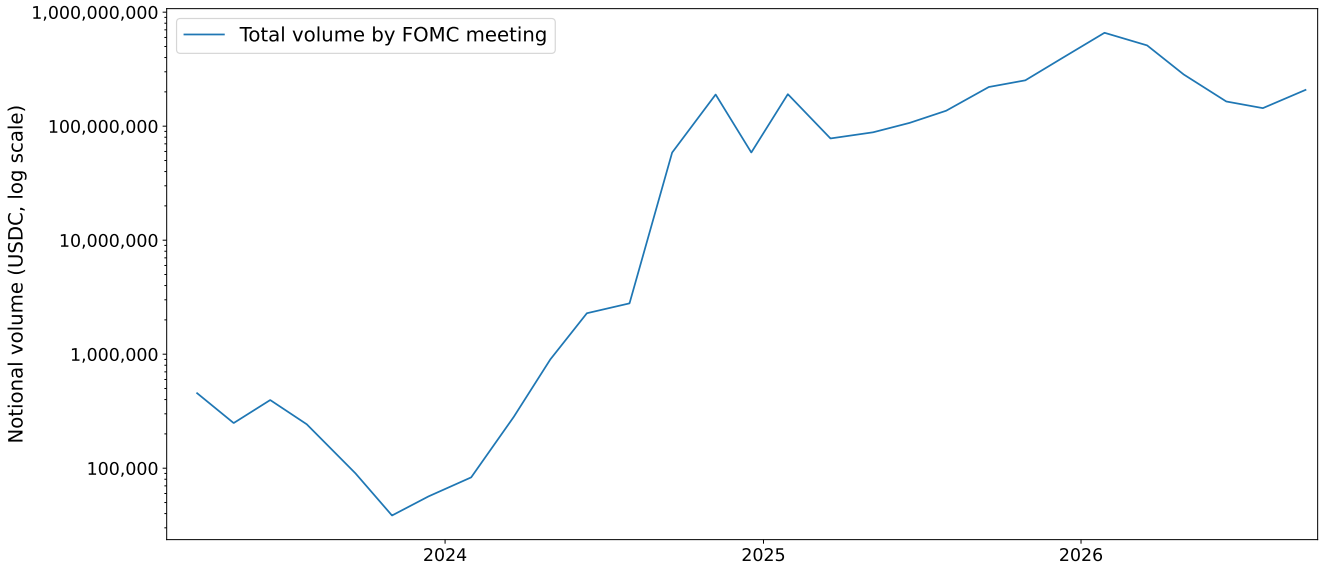


Figure 1. Total Polymarket notional volume (USDC, log scale) by FOMC meeting across the sample period. Source: Polymarket (polymarket.com).

4.3. Construction of ZQ-Implied Expected Rate Changes

Federal Funds Futures settle on 100 minus the average EFFR over the contract month. For the contract on month m with price P_m^{ZQ}, the implied monthly average rate is $\bar{r}_m = 100 - P_m^{\mathrm{ZQ}}$. Following the CME FedWatch methodology, the first subsequent month containing no FOMC meeting is designated the anchor month a, whose implied average rate is treated as the expected end-of-previous-month EFFR, $r_{a-1}^{\mathrm{end}} = \bar{r}_a$. A month containing a meeting has a pre-announcement rate r_m^{start} and a post-announcement rate r_m^{end}, whose day-weighted average equals $\bar{r}_m$:

$$\bar{r}_m = w_m\, r_m^{\mathrm{start}} + (1 - w_m)\, r_m^{\mathrm{end}}, \qquad w_m = \frac{N_m}{N_m + M_m}, \tag{3}$$

where N_m and M_m are the number of days on/before and after the announcement, respectively. Rearranging,

$$r_m^{\mathrm{start}} = \frac{\bar{r}_m - (1 - w_m)\, r_m^{\mathrm{end}}}{w_m}, \tag{4}$$

and applying this inversion month by month, from the anchor month back to the target meeting, recovers both r_m^{start} and r_m^{end}. The ZQ-implied expected rate change at meeting m is then

$$E_t^{\mathrm{ZQ}}(\Delta_m) = r_m^{\mathrm{end}} - r_m^{\mathrm{start}}. \tag{5}$$

Tick-level ZQ trades are aggregated to one-second precision, with a new implied change recorded whenever any contract in the relevant chain trades. Because FOMC decisions are assumed to occur in discrete 25-basis-point increments, $E_t^{\mathrm{ZQ}}(\Delta_m)$ is interpretable as a probability-weighted expectation conditional on only the two 25-basis-point outcomes adjacent to the implied value being possible; it also relies on EFFR moving proportionally with the target range, consistent with the FedWatch methodology, so that it is comparable with the target-range-based Polymarket measure of Section 4.2.

4.4. Cointegration

Long-run co-movement between $\mathrm{PM}_t \equiv E_t^{\mathrm{PM}}(\Delta_m)$ and $\mathrm{ZQ}_t \equiv E_t^{\mathrm{ZQ}}(\Delta_m)$ is assessed with the Engle–Granger two-step test. Each series is first tested for a unit root with the augmented Dickey–Fuller (ADF) test, with lag length selected by the Akaike information criterion. The long-run regression

$$\mathrm{PM}_t = \alpha + \beta\,\mathrm{ZQ}_t + \varepsilon_t \tag{6}$$

is estimated per meeting and per sampling grid, and the residual $\hat{\varepsilon}_t = \mathrm{PM}_t - \hat{\alpha} - \hat{\beta}\,\mathrm{ZQ}_t$ is tested for a unit root. The two series are treated as cointegrated when each is individually non-stationary in levels but $\hat{\varepsilon}_t$ is stationary. Tests are run separately for each meeting on regular 1-, 10-, 30-, and 60-min grids constructed by taking the last observation within each interval and carrying the previous value forward when no trade occurs.

4.5. Forecast Accuracy Comparison

For each meeting, the realized change in the target range's upper bound is Δ_m, forecast at a common pre-meeting horizon by $E^{\mathrm{PM}}(\Delta_m)$ and $E^{\mathrm{ZQ}}(\Delta_m)$. Forecast rationality is assessed with the Mincer–Zarnowitz regression [20],

$$\Delta_m = \alpha + \beta\,E^{\mathrm{PM}}(\Delta_m) + u_m, \tag{7}$$

(and analogously for ZQ), jointly testing $\alpha = 0$, $\beta = 1$ with an F-test; $\alpha \neq 0$ indicates systematic bias, and $\beta < 1$ ($\beta > 1$) indicates over-/under-reaction relative to outcomes. The two forecasts are compared directly with the Diebold–Mariano test [21]. Letting

$$d_m = \left(\Delta_m - E^{\mathrm{PM}}(\Delta_m)\right)^2 - \left(\Delta_m - E^{\mathrm{ZQ}}(\Delta_m)\right)^2, \tag{8}$$

the test statistic

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d})}} \tag{9}$$

uses an HAC estimate of the variance of $\bar{d}$ and is asymptotically standard normal under equal predictive accuracy; a positive value indicates Polymarket is the less accurate forecaster. Given the small sample (29 meetings), the Harvey–Leybourne–Newbold small-sample correction [22] is applied throughout, $\mathrm{DM}^* = \mathrm{DM} \times \sqrt{(n+1-2h+h(h-1)/n)/n}$ with $h = 1$, and the corrected statistic is compared against a Student-t rather than a normal distribution.

4.6. Price Discovery: Granger Causality and Error-Correction

The direction of price discovery is tested with a pooled bivariate Granger-causality regression estimated on first differences with three lags,

$$\Delta\mathrm{ZQ}_t = c + \sum_{i=1}^{3} \alpha_i\,\Delta\mathrm{ZQ}_{t-i} + \sum_{i=1}^{3} \beta_i\,\Delta\mathrm{PM}_{t-i} + u_t, \tag{10}$$

with the symmetric regression (dependent variable $\Delta\mathrm{PM}_t$) testing the reverse direction; standard errors are clustered by meeting. The null that Polymarket does not Granger-cause

ZQ, $\beta_1 = \beta_2 = \beta_3 = 0$, is tested with an F-statistic distributed $F(3, m-1)$, where m is the number of meetings contributing to the cell.

A single-equation error-correction model recovers the speed at which each market corrects deviations from the long-run relationship of Equation (6). Using the residual $\hat{\varepsilon}_t$ from that regression,

$$\Delta \mathrm{PM}_t = c_{\mathrm{PM}} + \gamma_{\mathrm{PM}}\, \hat{\varepsilon}_{t-1} + \phi_1\, \Delta \mathrm{PM}_{t-1} + \phi_2\, \Delta \mathrm{ZQ}_{t-1} + e_{\mathrm{PM},t}, \tag{11}$$

$$\Delta \mathrm{ZQ}_t = c_{\mathrm{ZQ}} + \gamma_{\mathrm{ZQ}}\, \hat{\varepsilon}_{t-1} + \psi_1\, \Delta \mathrm{PM}_{t-1} + \psi_2\, \Delta \mathrm{ZQ}_{t-1} + e_{\mathrm{ZQ},t}, \tag{12}$$

are estimated per meeting. A significant $\gamma_{\mathrm{PM}} < 0$ indicates that Polymarket is pulled back toward ZQ, and a significant $\gamma_{\mathrm{ZQ}} > 0$ indicates that ZQ is pulled toward Polymarket; the market with the larger $|\gamma|$ is interpreted as the follower. Both tests capture predictive precedence rather than structural causation and, by construction, are silent on information shared within the same sampling interval.

4.7. Robustness: A Trading-Frequency Freshness Filter

ZQ trades almost continuously and evenly through exchange hours, whereas Polymarket trades far less consistently, with activity concentrated close to each meeting and, in early meetings, orders of magnitude thinner than in later ones (959 Polymarket trades against 67,839 ZQ trades at the March 2023 meeting versus 97,989 against 89,182 before the September 2026 announcement). Because the two series are aligned by forward filling, a single interval can pair a current ZQ quote with a Polymarket quote that is stale by several minutes; a genuine ZQ move followed mechanically by a delayed Polymarket "catch-up" trade can then be misread by a naive regression as Polymarket following ZQ, while the same staleness fills the Polymarket differenced series with spurious zeros that can mask a genuine ZQ-to-Polymarket lead.

The lead–lag tests are therefore re-estimated on a freshness-filtered sample using the update indicators of Section 4 to reconstruct, at each grid point, the time elapsed since each series last traded. Two variants are reported: a *strict* filter requiring both series to be fresh across the entire lag window and a *light* filter requiring freshness only at the prediction point and the most recent lag. Both variants are applied only from the meeting in mid-2024 onward, excluding the early thinly traded meetings so that the filtered results describe the period in which Polymarket is liquid enough to be tested without forward-fill contamination. As a second robustness check, all the tests are repeated across the 1-, 10-, 30-, and 60-min sampling grids.

Appendix D (Table A3) reports the recorded included and excluded trades under each freshness specification and gives the precise thresholds.

5. Results

5.1. Forecast Accuracy

Table 1 reports the Mincer–Zarnowitz and Diebold–Mariano results by horizon. At the 90-, 60-, 30-, 14-, 7-, and 1-day horizons, the joint null of an unbiased well-calibrated forecast is not rejected for either market, with slopes close to one. At the longest 90-day horizon, the joint-test p-values are 0.12 for Polymarket and 0.11 for ZQ, with slopes of 1.05 and 0.92, respectively; this horizon rests on only fourteen of the twenty-nine meetings and should be read cautiously. The Diebold–Mariano statistic is statistically indistinguishable from zero at every horizon (corrected statistic between -0.70 and 0.26, $p \in [0.49, 0.98]$), and its sign is mixed across horizons, so neither market is consistently the more accurate forecaster. The tests do not reject the calibration restrictions or detect a statistically significant difference in squared forecast errors within the available sample; they do not establish equal accuracy or forecast rationality (H2 is not rejected).

Table 1. Forecast accuracy by horizon. $\hat{\beta}$ is the Mincer–Zarnowitz slope for each market, with the p-value of the joint test $\alpha = 0, \beta = 1$ in parentheses (a p-value above 0.05 indicates rationality is not rejected). DM* is the Harvey–Leybourne–Newbold-corrected Diebold–Mariano statistic (positive values indicate Polymarket is the less accurate), with its p-value in parentheses. n is the number of meetings.

Horizon	n	$\hat{\beta}_{\mathrm{PM}}(p)$	$\hat{\beta}_{\mathrm{ZQ}}(p)$	$\mathrm{DM}^*(p)$
90 days	14	1.05 (0.12)	0.92 (0.11)	−0.03 (0.98)
60 days	15	1.24 (0.08)	0.99 (0.21)	0.24 (0.82)
30 days	27	1.28 (0.19)	1.19 (0.44)	−0.70 (0.49)
14 days	29	1.06 (0.31)	1.02 (0.39)	0.23 (0.82)
7 days	29	1.18 (0.24)	1.11 (0.63)	−0.05 (0.96)
1 day	29	1.10 (0.17)	1.06 (0.48)	0.26 (0.80)

Appendix E reports the forecast errors and their uncertainty (Table A4) and confidence intervals for the Mincer–Zarnowitz intercepts and slopes (Table A5). Figure 2 shows the mean absolute forecast errors for the two markets across forecast horizons.

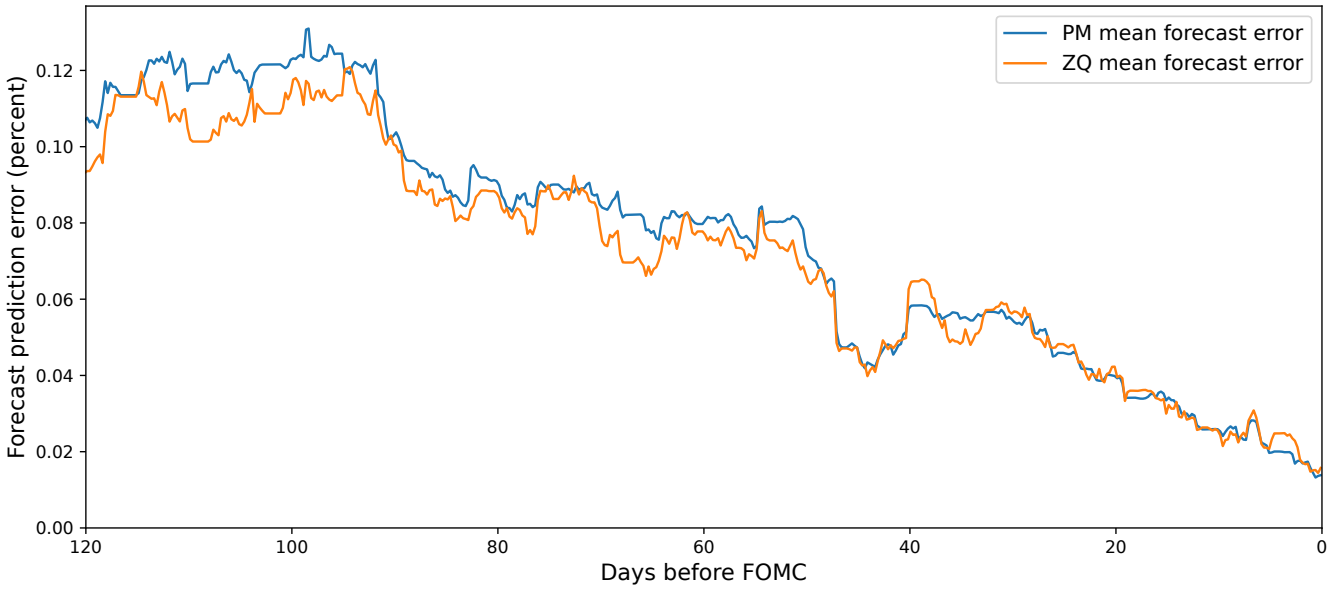


Figure 2. Mean absolute forecast error of the Polymarket-implied (PM) and ZQ-implied expected rate changes against the realized decision by horizon before the FOMC announcement averaged across the 29 meetings.

5.2. Co-Movement and Cointegration

The two market-implied series move together closely, with a correlation in levels averaging 0.80 across the 29 meetings; because two non-stationary series can appear correlated without a genuine relationship, cointegration is the more informative test. Table 2 reports the Engle–Granger results by sub-period and grid. At the coarser 10-, 30-, and 60-min grids used for the lead–lag analysis, the average residual p-value falls from around 0.05–0.11 in 2023–2024 to around 0.01–0.03 in 2025–2026; at the 10-min grid, the count of meetings satisfying the full cointegration criterion rises from eight of fifteen to eleven of fourteen. The estimated slope $\hat{\beta}$ is positive at every meeting and, in the later sample, sits close to one (0.99 at the September 2024 meeting; 1.01 at March 2026), with departures concentrated in the earliest most thinly traded meetings (an implausible slope of about 5.1 at March 2024; no cointegration at any grid for January 2024). Both the frequency of cointegration and the estimated slope therefore support H1: the two markets price the same underlying

expectation and are tied together in the long run, but the link was loose and unstable while Polymarket was thinly traded and tightened materially as the platform matured.

Table 2. Engle–Granger cointegration by sub-period and sampling frequency. Each cell reports the average residual *p*-value across meetings in that period with, in parentheses, the number of meetings satisfying the full cointegration criterion (both series non-stationary in levels; residual spread stationary at the 5% level). Lower *p*-values indicate a stronger long-run relationship.

Period	Meetings	1 min	10 min	30 min	60 min
2023–2024	15	0.006 (6/15)	0.052 (8/15)	0.092 (8/15)	0.105 (7/15)
2025–2026	14	0.000 (11/14)	0.012 (11/14)	0.018 (10/14)	0.032 (11/14)

Appendix A provides plots of the two market-implied series and the realized decision for three representative meetings (Figures A1–A3).

Appendix F (Table A6) reports the per-meeting unit-root and residual-test results and the ten-minute long-run intercepts and slopes with uncertainty ranges for all 29 meetings. Tables A1 and A2 in Appendix B also report the individual-series unit-root results.

5.3. Lead–Lag and Price-Discovery Results

On the unfiltered grid the Granger tests give a mixed picture (Table 3). Past Polymarket changes significantly predict subsequent ZQ changes at the 10-, 30-, and 60-min grids; the reverse direction is also significant at 10 min, where it is in fact the stronger of the two, but it is insignificant at 30 and 60 min. Meanwhile, at 1 min, neither direction is significant because the forward-filled Polymarket series carries almost no genuine variation between consecutive observations.

The error-correction estimates (Table 4) agree in direction: both adjustment coefficients carry their equilibrium-restoring sign and are significant for a large majority of the meetings, but γ_{ZQ} is consistently larger in absolute value than γ_{PM} (e.g., $\gamma_{ZQ} = +0.036$ against $\gamma_{PM} = -0.008$ at 10 min), so ZQ behaves more as the follower on the unfiltered grid.

These unfiltered results are, however, partly an artifact of Polymarket's staleness. Under the strict freshness filter, the Polymarket-to-ZQ relationship strengthens at 10, 30, and 60 min rather than weakening, confirming that this lead is not simply a product of stale quotes; at the same time the ZQ-to-Polymarket direction, previously significant only at 10 min, becomes significant at every grid. The relationship is therefore bidirectional rather than one-directional: each market's own past changes carry information about the other's next move. A consistent pattern emerges across frequencies: at 1 min both directions are significant with joint *F*-statistics of 3.24 for Polymarket-to-ZQ and 3.85 for ZQ-to-Polymarket; at 10, 30 and 60 min both directions remain significant with larger joint *F*-statistics in the Polymarket-to-ZQ direction. The filtered error-correction estimates agree: the asymmetry favoring ZQ as the faster adjuster is the largest at 10 min (about three times the Polymarket adjustment speed) and narrows to near-parity by 60 min.

Appendix G reports the size and uncertainty of the strict-filter Granger relationships (Table A7) and filtered and unfiltered estimates over the same meetings and periods (Table A9). Table A8 reports Granger tests with two and four lags to check sensitivity to the three-lag specification. Appendix H (Table A10) reports the standard and wild cluster bootstrap *p*-values under strict freshness filtering.

At the 5% level, the wild cluster bootstrap retains Polymarket-to-ZQ prediction at 10, 30 and 60 min and ZQ-to-Polymarket prediction only at 60 min; neither direction remains significant at one minute. Thus, the evidence for bidirectionality at every grid is sensitive to the inference procedure, and larger *F*-statistics alone do not measure a larger economic contribution to price discovery.

Table 3. Pooled bivariate Granger causality on first differences: joint F-statistic on the three cross-lags by direction and grid, unfiltered and under the strict and light freshness filters. Standard errors are clustered by meeting; each test is distributed $F(3, m-1)$, with m (the number of meetings entering the cell) shown in parentheses. Meetings with fewer than 50 valid observations under a given grid and filter are omitted. * $p < 0.05$, ** $p < 0.01$.

Grid	Direction	No Filter	Strict Filter	Light Filter
1 min	PM → ZQ	0.33 (29)	3.24 * (17)	5.49 ** (20)
	ZQ → PM	2.66 (29)	3.85 * (17)	8.75 ** (20)
10 min	PM → ZQ	7.55 ** (29)	26.69 ** (18)	39.72 ** (22)
	ZQ → PM	9.21 ** (29)	6.12 ** (18)	5.79 ** (22)
30 min	PM → ZQ	8.68 ** (29)	27.38 ** (19)	43.24 ** (22)
	ZQ → PM	2.87 (29)	4.03 * (19)	3.06 (22)
60 min	PM → ZQ	10.04 ** (29)	41.03 ** (19)	28.21 ** (23)
	ZQ → PM	2.09 (29)	6.90 ** (19)	3.73 * (23)

Table 4. Error-correction adjustment speeds before and after the strict freshness filter: mean coefficient on the lagged equilibrium error by grid. The market with the larger $|\gamma|$ is the follower.

Grid	Before Filter		After Filter	
	γ_{PM}	γ_{ZQ}	γ_{PM}	γ_{ZQ}
10 min	−0.008	+0.036	−0.023	+0.082
30 min	−0.018	+0.053	−0.057	+0.089
60 min	−0.032	+0.063	−0.102	+0.131

6. Discussion

The two series are closely aligned and cointegrated in a majority of the meetings and nearly all of the more liquid later sub-sample, with a long-run slope near one, so Polymarket and Federal Funds Futures measure the same underlying expectation rather than drifting apart (H1 supported). The forecast tests do not detect a statistically significant accuracy difference at any tested horizon within the available sample (H2 is not rejected). The substantive answer to the paper's research question lies in the direction of price discovery, where the evidence is of a two-way rather than one-directional relationship. There is a clear dependence on sampling frequency: conventional clustered tests find both directions significant at the one-minute horizon under the strict filter, including a larger joint F-statistic from futures to Polymarket. The results do not establish the exclusive futures lead suggested by the microstructure prediction, which holds that the deeper more continuously traded professionally populated market incorporates new information the fastest [3,14,15]. At the ten-, thirty- and sixty-minute horizons, both directions remain significant with larger joint F-statistics in the Polymarket-to-ZQ direction. The error-correction estimates show the futures price doing more of the adjusting back toward the long-run relationship at these horizons. Polymarket's past price changes help to predict subsequent futures changes at some horizons, but this does not by itself distinguish an independent informational contribution from different responses to common public information.

Two features of the setting make such a longer-horizon contribution possible without by themselves establishing outright leadership. First, the absence of arbitrage between the two venues (Section 3) is a permissive rather than causal condition: it allows a genuine lead to persist beyond the few seconds over which it would normally be arbitraged away in integrated markets without generating that lead itself. Second, and more cautiously, the longer-horizon result admits two readings that this design cannot fully separate. On one reading it is genuine: an unconstrained speculative Polymarket pool moves to a directional view faster than institutional ZQ participants, whose repricing is slowed by mandates, risk

limits, and hedging motives, which is consistent with ZQ doing more of the error-correcting adjustment. On the other reading it is largely a timing artifact in which both venues respond to the same public information at different speeds—ZQ within seconds and Polymarket more slowly—so that, at the thirty- and sixty-minute grids, a lagged Polymarket print lines up against ZQ's next-bar change and a common shock arriving at different speeds is recorded as a Granger causality. The freshness filter removes stale unchanging quotes but not this difference in the timing of genuine updates. The conservative conclusion is that the markets display frequency-dependent predictive relationships, while an independent contribution by Polymarket remains unestablished (H3 is supported only in its predictive component).

This extends to a structurally very different offshore cryptocurrency-collateralized platform that is closed to regulated institutional participants regarding the recent finding that regulated prediction markets such as Kalshi are informative about macroeconomic events [4,5]: informational value in a prediction market need not depend on a regulated venue or an institutional participant base. It does, however, appear to depend on liquidity: Polymarket's contribution was weak in its early thinly traded period and strengthened as the platform matured. Therefore, prediction-market signals should be weighted by the depth of trading behind them rather than treated as uniformly informative. The two markets are best viewed as complements rather than substitutes.

Limitations

The asynchronous and uneven nature of Polymarket's trading is the most important limitation. The freshness filter removes the resulting staleness at the cost of discarding most observations and all of the early meetings, so the clean lead–lag results describe only Polymarket's mature more liquid period and the most actively traded intervals within each meeting. The unfiltered and filtered results should be read together, the former showing the relationship that the raw data suggest and the latter the relationship that survives once staleness is controlled. A more direct remedy would estimate co-movement from the irregular tick data itself; the Hayashi–Yoshida estimator [23] measures covariation between non-synchronously observed price processes without a fixed grid or forward filling, avoiding both the staleness artifact and the loss of observations that the freshness filter incurs, which is a natural avenue for future robustness checks. The sample is also small—twenty-nine meetings, falling further after filtering and at the longest horizons—which limits statistical power, particularly in the one-minute Granger results and the ninety-day Mincer–Zarnowitz tests (resting on fourteen meetings, with neither market rejecting at 5%). Granger causality and the error-correction model capture predictive precedence rather than structural causation and, by construction, measure only lagged relationships; a full vector error-correction model with Hasbrouck information shares [3] would more precisely apportion the contribution of each market. Finally, the constructed series rest on the assumptions in Sections 4.2 and 4.3, in particular the FedWatch anchoring logic and the proportional treatment of EFFR and the target range; the sample period spans a single tightening and easing cycle with relatively few policy surprises, and the specific settlement and oracle risk to Polymarket is not modeled.

Some further limitations are also worth mentioning both as a word of caution about our results and meaningful directions for future improvement. Firstly, as an anonymous reviewer notes, the freshness thresholds involve a judgment, and filtering may select more active trading periods as well as remove stale prices. Furthermore, our exploratory checks around CPI, unemployment and other macroeconomic releases had too few usable observations to infer meaningful relationships between these announcements and each market's price response. Lastly, a validation against historical CME FedWatch probabilities at matching dates and times, together with a sample-specific comparison of EFFR and target-range

changes, would be desirable and remains for future research; the present transformation relies on the cited FedWatch methodology.

7. Conclusions

The three hypotheses receive differing degrees of support. The two market-implied measures show evidence of long-run co-movement (H1 supported), while the forecast tests do not detect a statistically significant difference in accuracy within the available sample (H2 not rejected); this does not establish equivalence or forecast rationality. The weight of the paper falls on H3, where the answer is neither what the raw lead–lag statistics suggest at face value nor the proposed null hypothesis. The unfiltered Granger tests show predictive precedence from Polymarket at 10, 30 and 60 min; under strict freshness filtering, conventional clustered tests find significance in both directions at every tested grid, whereas the wild cluster bootstrap retains Polymarket-to-ZQ prediction at 10, 30 and 60 min and ZQ-to-Polymarket prediction only at 60 min. These results do not establish outright leadership or rule out a timing explanation in which the markets respond to common public information at different speeds. The evidence supports predictive relationships whose significance depends on sampling frequency, liquidity and the inference procedure. Quantifying Polymarket's exact share of price discovery is the natural next step. This would be conducted via a vector error-correction model with Hasbrouck information shares or a Hayashi–Yoshida estimator applied directly to the irregular tick data. More broadly, the results suggest that informational value in a prediction market need not require regulation, an institutional participant base, or onshore access: it requires liquidity, which is why Polymarket's contribution grew over the sample and why its signal should be weighted accordingly. Polymarket therefore appears to be informative in this setting with predictive relationships that depend on liquidity and sampling frequency, while an independent contribution to price discovery remains unestablished.

Author Contributions: Conceptualization, M.A.G. and G.M.; methodology, M.A.G.; software, M.A.G.; formal analysis, M.A.G.; data curation, M.A.G.; writing—original draft preparation, M.A.G.; writing—review and editing, M.A.G. and G.M.; supervision, G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. The data were obtained from Databento and Goldsky and are available at their respective URLs: https://databento.com/ (accessed on 25 September 2026) and https://docs.goldsky.com/chains/polymarket (accessed on 25 September 2026), with the permission of Goldsky and Databento.

Acknowledgments: This article is adapted from the first author's undergraduate thesis, which was supervised by the second author at the Università degli Studi di Trieste.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FOMC	Federal Open Market Committee
EFFR	Effective federal funds rate
CME	Chicago Mercantile Exchange
CFTC	Commodity Futures Trading Commission

ZQ	CME 30-Day Federal Funds Futures ticker
PM	Polymarket-implied expected rate change
ADF	Augmented Dickey–Fuller (test)
ECM	Error-correction model
HAC	Heteroskedasticity- and autocorrelation-consistent

Appendix A. Additional Figures

Figures A1–A3 plot the two market-implied series against the realized decision for three representative meetings: a large surprise (September 2024), a standard cut (September 2025), and a hold (April 2026).

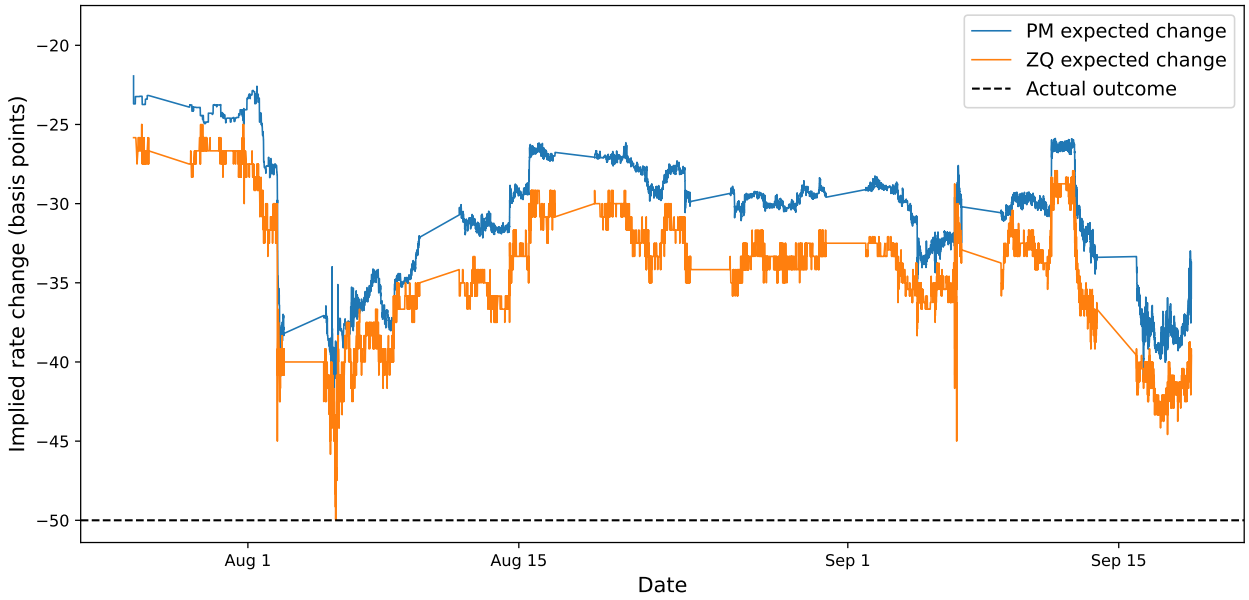


Figure A1. Polymarket-implied and ZQ-implied expected rate change ahead of the 18 September 2024 FOMC meeting. The dashed line marks the realized decision, a 50-basis-point cut (−50 bp). Both markets were uncertain of the outcome, with Polymarket pricing in a higher probability of only a 25-basis-point cut.

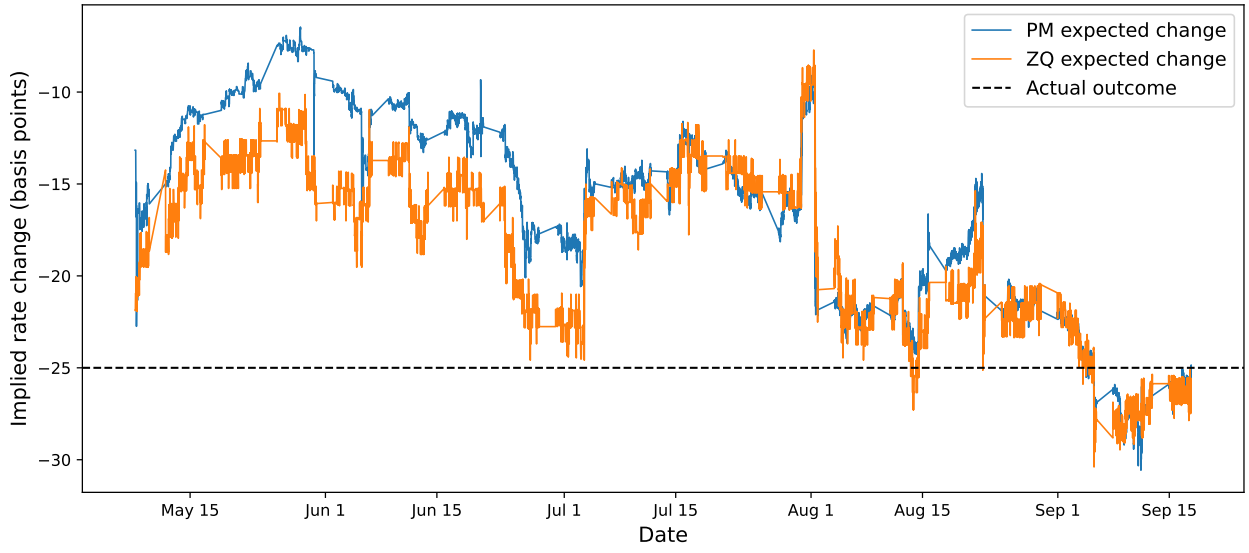


Figure A2. Polymarket-implied and ZQ-implied expected rate change ahead of the 17 September 2025 FOMC meeting. The dashed line marks the realized decision, a 25-basis-point cut (−25 bp), toward which both series converge.

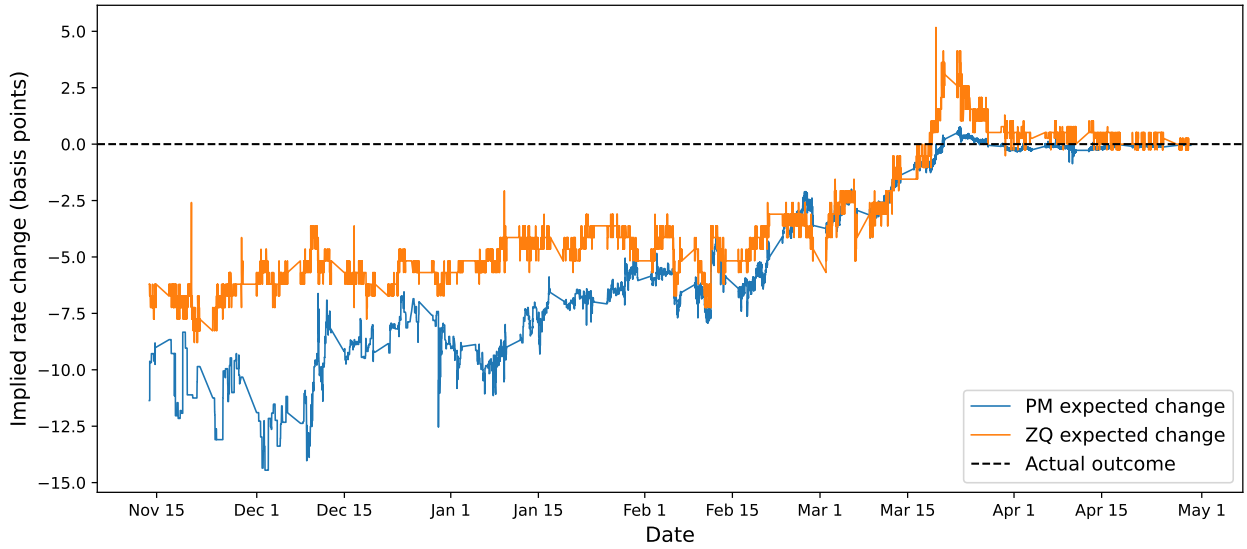


Figure A3. Polymarket-implied and ZQ-implied expected rate change ahead of the 29 April 2026 FOMC meeting. The dashed line marks the realized decision, no change (0 bp), toward which both series converge.

Appendix B. Additional Tables

Tables A1 and A2 report the per-meeting augmented Dickey–Fuller unit-root test p-values underlying the cointegration analysis of Section 4.4 (Table 2) for the Polymarket-implied and ZQ-implied series, respectively. The lag length is selected by AIC. A large p-value indicates a unit root (non-stationarity in levels), the precondition for cointegration; an asterisk marks $p < 0.05$, where the series is instead stationary.

Table A1. ADF p-values for the Polymarket-implied change (PM_t) in levels by meeting and sampling frequency. An asterisk (*) denotes rejection of the unit-root null at the 5% level ($p < 0.05$).

Meeting	1 min	10 min	30 min	60 min
22 March 2023	0.238	0.319	0.105	0.475
3 May 2023	0.530	0.589	0.596	0.579
14 June 2023	0.140	0.139	0.213	0.278
26 July 2023	0.312	0.212	0.217	0.220
20 September 2023	0.068	0.110	0.107	0.105
1 November 2023	0.265	0.322	0.297	0.281
13 December 2023	0.001 *	0.001 *	<0.001 *	0.056
31 January 2024	0.183	0.183	0.378	0.378
20 March 2024	0.667	0.669	0.676	0.677
1 May 2024	0.096	0.277	0.455	0.098
12 June 2024	<0.001 *	0.016 *	0.089	0.077
31 July 2024	<0.001 *	0.127	0.173	0.219
18 September 2024	0.227	0.301	0.192	0.195
7 November 2024	0.506	0.632	0.598	0.635
18 December 2024	0.364	0.453	0.452	0.420
29 January 2025	0.062	0.029 *	0.019 *	0.285
19 March 2025	0.469	0.720	0.704	0.713
7 May 2025	0.070	0.167	0.440	0.639
18 June 2025	0.844	0.849	0.840	0.828

Table A1. *Cont.*

Meeting	1 min	10 min	30 min	60 min
30 July 2025	0.724	0.777	0.777	0.788
17 September 2025	0.473	0.651	0.558	0.582
29 October 2025	0.678	0.855	0.767	0.830
10 December 2025	0.067	0.180	0.114	0.209
28 January 2026	0.456	0.821	0.890	0.882
18 March 2026	0.626	0.636	0.680	0.682
29 April 2026	0.366	0.623	0.721	0.828
17 June 2026	<0.001 *	0.084	0.096	0.105
29 July 2026	0.134	0.143	0.198	0.223
16 September 2026	0.787	0.796	0.784	0.769

Table A2. ADF *p*-values for the ZQ-implied change (ZQ_t) in levels by meeting and sampling frequency. An asterisk (*) denotes rejection of the unit-root null at the 5% level ($p < 0.05$).

Meeting	1 min	10 min	30 min	60 min
22 March 2023	0.433	0.448	0.418	0.397
3 May 2023	<0.001 *	0.098	0.110	0.079
14 June 2023	0.086	0.245	0.341	0.343
26 July 2023	0.326	0.736	0.715	0.659
20 September 2023	<0.001 *	0.238	0.362	0.463
1 November 2023	0.345	0.507	0.667	0.684
13 December 2023	<0.001 *	0.001 *	0.011 *	0.009 *
31 January 2024	<0.001 *	0.001 *	0.006 *	0.082
20 March 2024	<0.001 *	0.081	0.179	0.166
1 May 2024	<0.001 *	0.003 *	0.028 *	0.018 *
12 June 2024	<0.001 *	0.001 *	0.126	0.081
31 July 2024	<0.001 *	0.163	0.295	0.382
18 September 2024	0.057	0.138	0.178	0.101
7 November 2024	<0.001 *	0.072	0.177	0.131
18 December 2024	0.244	0.437	0.427	0.420
29 January 2025	0.144	0.357	0.354	0.363
19 March 2025	0.248	0.718	0.702	0.700
7 May 2025	<0.001 *	<0.001 *	<0.001 *	<0.001 *
18 June 2025	0.499	0.677	0.699	0.814
30 July 2025	0.711	0.796	0.796	0.841
17 September 2025	0.054	0.305	0.271	0.300
29 October 2025	0.316	0.509	0.490	0.513
10 December 2025	<0.001 *	0.011 *	0.025 *	0.018 *
28 January 2026	0.422	0.697	0.677	0.654
18 March 2026	0.147	0.753	0.812	0.865
29 April 2026	0.064	0.586	0.637	0.617
17 June 2026	<0.001 *	0.143	0.459	0.521
29 July 2026	0.114	0.300	0.304	0.293
16 September 2026	0.055	0.641	0.726	0.723

Appendix C. Construction of Data

For 29 FOMC meetings from March 2023 to September 2026, the code combines Polymarket outcome contracts into a meeting-level implied rate change. The total sample contains 115 contracts.

Yes prices are used directly and No prices are converted to 1 minus the observed price. Within each outcome, records sharing a transaction hash are consolidated; token-volume-weighted prices are then calculated for each UTC second. Each outcome's last price is carried forward until its next trade.

Open-ended categories use their stated boundary (e.g., a 50+ bp cut is coded as −50 bp) rather than an estimated within-tail mean.

The series ends at the FOMC announcement of interest rate changes. Its opening period is trimmed to the first date when cumulative recorded token volume reaches 0.1% of its full-archive total. The series is aligned with futures at observed UTC seconds, restricted to the coded CME trading-hour schedule, and sampled at 1, 10, 30 and 60 min.

At each forecast timestamp, the code uses the latest recorded PM and ZQ values at or before that time, carrying earlier prices forward as needed. Aggregating trades to seconds can introduce a timing difference of less than one second. The initial volume cutoff uses full-sample volume, and some cleaning uses later observations to identify unusual prices, so preparation of the historical dataset is not fully real-time.

Appendix D. Freshness Filtering: Included and Excluded Trades

Table A3. How many recorded trades are included after freshness filtering.

Model	Grid	Rule	Total Trades	Excluded Trades	Included Trades	Included Trades (%)
Granger	1 min	none	3,229,177	0	3,229,177	100.00
Granger	1 min	light	3,229,177	2,795,571	433,606	13.43
Granger	1 min	strict	3,229,177	3,115,962	113,215	3.51
Granger	10 min	none	3,229,177	0	3,229,177	100.00
Granger	10 min	light	3,229,177	1,516,222	1,712,955	53.05
Granger	10 min	strict	3,229,177	2,263,917	965,260	29.89
Granger	30 min	none	3,229,177	0	3,229,177	100.00
Granger	30 min	light	3,229,177	926,930	2,302,247	71.30
Granger	30 min	strict	3,229,177	1,588,949	1,640,228	50.79
Granger	60 min	none	3,229,177	0	3,229,177	100.00
Granger	60 min	light	3,229,177	594,588	2,634,589	81.59
Granger	60 min	strict	3,229,177	1,233,652	1,995,525	61.80
ECM	10 min	none	3,229,177	0	3,229,177	100.00
ECM	10 min	strict	3,229,177	2,056,599	1,172,578	36.31
ECM	30 min	none	3,229,177	0	3,229,177	100.00
ECM	30 min	strict	3,229,177	1,407,761	1,821,416	56.40
ECM	60 min	none	3,229,177	0	3,229,177	100.00
ECM	60 min	strict	3,229,177	1,042,781	2,186,396	67.71

Total trades: 3,229,177. Each recorded PM or ZQ update counts as one trade; a CSV row with updates in both markets counts as two. A trade is included when its time interval supplies a current or past price needed by a fitted equation. It is counted once per model/filter even if several equations use it. The models take the last price in each interval; these counts describe the trading activity covered by their inputs. Meeting histories can overlap. A price is fresh if its market measure updated within 30 s, 5 min, 15 min or 30 min on the 1-, 10-, 30- or 60-min grid. Both markets must qualify. Light filtering checks the current and previous intervals. Strict Granger filtering checks the current and four earlier intervals; strict ECM checks the current and two earlier intervals. ECM counts cover its adjustment equations; its long-run relationship is still estimated using the full grid.

Appendix E. Forecast Accuracy and Calibration Uncertainty

Table A4. Average forecast errors and their uncertainty at each forecast horizon.

Days	n	PM MAE [95%]	ZQ MAE [95%]	PM−ZQ MAE [95%]
90	14	10.395 [8.094, 12.576]	9.853 [7.003, 12.731]	0.542 [−1.057, 2.141]
60	15	7.972 [5.669, 10.350]	7.777 [5.566, 10.114]	0.195 [−1.533, 1.923]
30	27	5.457 [3.552, 7.695]	6.238 [4.005, 9.073]	−0.781 [−2.747, 1.185]
14	29	3.201 [1.757, 4.984]	3.114 [1.730, 4.829]	0.087 [−0.724, 0.898]
7	29	2.846 [1.290, 4.919]	2.915 [1.313, 4.918]	−0.069 [−0.610, 0.472]
1	29	1.462 [0.726, 2.434]	1.639 [0.946, 2.454]	−0.177 [−0.746, 0.391]

Days means days before the meeting; n is the number of available meetings. MAE is the average size of the forecast error, ignoring its sign: smaller is better. Errors are in basis points (1 bp = 0.01 percentage point). A negative PM−ZQ difference favors PM; a positive difference favors ZQ. Brackets show 95% uncertainty intervals. Every difference interval includes zero, so this sample does not clearly favor either market; it does not prove equal accuracy. The individual-market intervals use resampled meetings; difference intervals compare the two errors within each meeting. The paper's separate Diebold–Mariano test still uses squared errors.

Table A5. Checks of forecast bias and calibration (Mincer–Zarnowitz regressions).

Days	n	PM α [95%]	PM β [95%]	ZQ α [95%]	ZQ β [95%]
90	14	6.77 [−1.22, 14.75]	1.052 [0.689, 1.415]	5.00 [−2.71, 12.71]	0.920 [0.581, 1.258]
60	15	6.83 [0.33, 13.33]	1.235 [0.696, 1.775]	3.82 [−1.74, 9.39]	0.991 [0.496, 1.486]
30	27	1.32 [−1.64, 4.27]	1.280 [0.967, 1.594]	0.22 [−3.39, 3.83]	1.191 [0.887, 1.495]
14	29	−1.25 [−3.42, 0.91]	1.058 [0.817, 1.299]	−1.17 [−3.38, 1.04]	1.022 [0.742, 1.302]
7	29	0.51 [−1.36, 2.38]	1.177 [0.940, 1.415]	0.08 [−2.06, 2.22]	1.106 [0.881, 1.330]
1	29	−0.15 [−0.96, 0.66]	1.099 [0.992, 1.207]	0.02 [−0.93, 0.96]	1.056 [0.961, 1.151]

These regressions compare forecasts with the actual meeting decisions. The intercept α describes an offset (in bp), and the slope β describes the scale of the relationship. The calibration targets are $\alpha = 0$ and $\beta = 1$. Brackets show 95% confidence intervals, allowing the size of the errors to differ across meetings.

https://doi.org/10.3390/math14214071

Appendix F. Checks Behind the Error-Correction Model

Table A6. Checks of the long-run relationship for all 29 meetings.

Meeting	10 m PM	ZQ	EG	30 m PM	ZQ	EG	60 m PM	ZQ	EG	10 m α [Range]	10 m β [Range]
22 March 2023	0.319	0.448	<0.001	0.105	0.418	0.004	0.475	0.397	0.003	7.62 [6.95, 8.30]	0.689 [0.666, 0.713]
3 May 2023	0.589	0.098	0.141	0.596	0.110	0.142	0.579	0.079	0.433	13.43 [13.16, 13.71]	0.302 [0.289, 0.315]
14 June 2023	0.139	0.245	0.008	0.213	0.341	0.043	0.278	0.343	0.082	1.44 [1.03, 1.86]	0.576 [0.502, 0.649]
26 July 2023	0.212	0.736	<0.001	0.217	0.715	<0.001	0.220	0.659	<0.001	−4.13 [−4.66, −3.59]	1.172 [1.147, 1.196]
20 September 2023	0.110	0.238	<0.001	0.107	0.362	<0.001	0.105	0.463	<0.001	0.16 [0.01, 0.30]	1.022 [0.964, 1.080]
1 November 2023	0.322	0.507	0.005	0.297	0.667	0.020	0.281	0.684	0.009	0.70 [0.61, 0.80]	0.729 [0.701, 0.758]
13 December 2023	0.001	0.001	<0.001	<0.001	0.011	<0.001	0.056	0.009	0.024	0.50 [0.42, 0.58]	0.324 [0.289, 0.358]
31 January 2024	0.183	0.001	0.186	0.378	0.006	0.135	0.378	0.082	0.143	0.83 [0.33, 1.33]	1.868 [1.359, 2.377]
20 March 2024	0.669	0.081	0.123	0.676	0.179	0.498	0.677	0.166	0.274	−0.47 [−0.83, −0.10]	5.103 [4.739, 5.467]
1 May 2024	0.277	0.003	0.189	0.455	0.028	0.194	0.098	0.018	0.414	−0.11 [−0.21, −0.01]	0.754 [0.654, 0.855]
12 June 2024	0.016	0.001	0.003	0.089	0.126	0.060	0.077	0.081	0.057	−0.76 [−0.79, −0.73]	0.115 [0.093, 0.137]
31 July 2024	0.127	0.163	0.122	0.173	0.295	0.278	0.219	0.382	0.112	−1.48 [−1.55, −1.41]	0.232 [0.196, 0.267]
18 September 2024	0.301	0.138	<0.001	0.192	0.178	<0.001	0.195	0.101	<0.001	3.06 [2.46, 3.66]	0.991 [0.973, 1.009]
7 November 2024	0.632	0.072	<0.001	0.598	0.177	<0.001	0.635	0.131	<0.001	−9.12 [−9.64, −8.59]	0.671 [0.652, 0.690]
18 December 2024	0.453	0.437	<0.001	0.452	0.427	<0.001	0.420	0.420	0.027	−6.39 [−6.62, −6.15]	0.754 [0.743, 0.764]
29 January 2025	0.029	0.357	<0.001	0.019	0.354	0.002	0.285	0.363	0.002	0.40 [0.30, 0.50]	1.672 [1.641, 1.703]
19 March 2025	0.720	0.718	<0.001	0.704	0.702	0.001	0.713	0.700	<0.001	−0.22 [−0.26, −0.17]	0.892 [0.882, 0.901]
7 May 2025	0.167	<0.001	0.002	0.440	<0.001	0.017	0.639	<0.001	0.159	−3.90 [−4.06, −3.73]	0.569 [0.540, 0.599]
18 June 2025	0.849	0.677	<0.001	0.840	0.699	<0.001	0.828	0.814	0.001	−0.97 [−1.08, −0.85]	0.814 [0.800, 0.828]
30 July 2025	0.777	0.796	<0.001	0.777	0.796	<0.001	0.788	0.841	<0.001	−0.89 [−0.96, −0.82]	0.795 [0.784, 0.806]
17 September 2025	0.651	0.305	0.021	0.558	0.271	0.048	0.582	0.300	0.088	3.65 [3.22, 4.07]	1.096 [1.076, 1.117]
29 October 2025	0.855	0.509	0.009	0.767	0.490	0.010	0.830	0.513	0.005	−4.15 [−4.37, −3.92]	0.803 [0.792, 0.815]
10 December 2025	0.180	0.011	0.090	0.114	0.025	0.097	0.209	0.018	0.119	−5.51 [−6.15, −4.87]	0.638 [0.609, 0.666]
28 January 2026	0.821	0.697	<0.001	0.890	0.677	<0.001	0.882	0.654	<0.001	0.40 [0.29, 0.52]	1.195 [1.178, 1.211]
18 March 2026	0.636	0.753	<0.001	0.680	0.812	<0.001	0.682	0.865	0.005	−0.26 [−0.36, −0.16]	1.006 [0.987, 1.024]
29 April 2026	0.623	0.586	0.005	0.721	0.637	0.005	0.828	0.617	0.005	−0.80 [−0.90, −0.69]	1.348 [1.321, 1.375]
17 June 2026	0.084	0.143	0.009	0.096	0.459	0.011	0.105	0.521	0.019	−0.34 [−0.35, −0.32]	0.138 [0.127, 0.150]
29 July 2026	0.143	0.300	<0.001	0.198	0.304	0.002	0.223	0.293	0.003	−0.90 [−0.96, −0.83]	0.703 [0.689, 0.717]
16 September 2026	0.796	0.641	0.038	0.784	0.726	0.058	0.769	0.723	0.044	−4.18 [−4.54, −3.83]	1.033 [1.007, 1.059]

PM and ZQ give the ADF test p-values for the individual market series; EG gives the Engle–Granger test p-value for their long-run relationship. The cointegration criterion requires both individual p-values to be at least 0.05 and the EG p-value to be below 0.05. These checks describe support for the model's assumptions. All 29 meetings enter the unfiltered ECM. The last two columns describe the ten-minute fitted line, PM = $\alpha + \beta$ ZQ + gap: α is the offset in bp and β is the slope. The ranges describe approximate uncertainty, allowing for changing error size and related errors over time.

Appendix G. Granger Results and Robustness Checks

Table A7. Size and uncertainty of the Granger relationships.

Grid	Direction	Meetings	F	Cross-Lag Sum [Cluster 95%]	Partial R^2 (%)
1 min	PM → ZQ	17	3.24 *	0.184 [−0.191, 0.559]	0.60
1 min	ZQ → PM	17	3.85 *	0.073 [0.014, 0.132]	1.24
10 min	PM → ZQ	18	26.69 **	0.640 [0.452, 0.827]	1.37
10 min	ZQ → PM	18	6.12 **	0.060 [0.011, 0.108]	0.76
30 min	PM → ZQ	19	27.38 **	0.610 [0.457, 0.762]	1.63
30 min	ZQ → PM	19	4.03 *	0.076 [0.027, 0.125]	1.01
60 min	PM → ZQ	19	41.03 **	0.628 [0.458, 0.797]	2.13
60 min	ZQ → PM	19	6.90 **	0.121 [0.027, 0.214]	1.48

This expands the strict-filter results in the paper's Table 3. The *F*-statistics and meeting counts are repeated for reference. The cross-lag sum describes the combined effect of the other market's past changes: each is raised by 1 bp while own-market past changes are held fixed. Brackets show its 95% confidence interval. Partial R^2 measures the percentage reduction in the model's remaining squared error when the other market's past changes are added. It is not an information share. Uncertainty allows observations from the same meeting to be related. The coefficient sum and *F*-test test different claims, so their significance can differ. These results describe the fitted sample, not new forecasts. * $p < 0.05$, ** $p < 0.01$.

Table A8. Granger tests with two and four lags.

Grid	Direction	2 lags: None	Strict	Light	4 lags: None	Strict	Light
1 min	PM → ZQ	0.46 (29)	5.25 * (17)	7.84 ** (20)	0.30 (29)	3.99 * (15)	9.10 ** (20)
1 min	ZQ → PM	1.93 (29)	5.12 * (17)	7.90 ** (20)	2.41 (29)	13.43 ** (15)	6.12 ** (20)
10 min	PM → ZQ	7.20 ** (29)	34.78 ** (18)	47.77 ** (22)	8.12 ** (29)	32.26 ** (18)	33.51 ** (22)
10 min	ZQ → PM	13.30 ** (29)	2.27 (18)	2.08 (22)	7.99 ** (29)	4.84 ** (18)	5.27 ** (22)
30 min	PM → ZQ	11.67 ** (29)	22.71 ** (20)	28.93 ** (22)	8.59 ** (29)	33.97 ** (18)	39.27 ** (22)
30 min	ZQ → PM	2.33 (29)	3.86 * (20)	4.29 * (22)	3.02 * (29)	6.70 ** (18)	2.76 (22)
60 min	PM → ZQ	14.40 ** (29)	36.88 ** (20)	38.59 ** (23)	8.48 ** (29)	28.96 ** (18)	22.69 ** (23)
60 min	ZQ → PM	2.48 (29)	8.28 ** (20)	4.10 * (23)	2.37 (29)	4.69 ** (18)	2.64 (23)

This repeats the method of the paper's Table 3, changing the lag count to two or four. It checks whether the findings depend on choosing three past changes. Each cell shows the *F*-statistic with the number of contributing meetings in parentheses. Stars mark statistical significance (* $p < 0.05$, ** $p < 0.01$). Light checks the current and previous intervals; strict checks the current and three earlier intervals for two lags or the current and five earlier intervals for four lags.

Table A9. Filtered and unfiltered Granger results for the same meetings and periods.

Grid	Meetings	PM → ZQ: No Filter	Strict	ZQ → PM: No Filter	Strict
1 min	17	2.19	3.24 *	2.48	3.85 *
10 min	18	16.99 **	26.69 **	3.05	6.12 **
30 min	19	39.10 **	27.38 **	2.88	4.03 *
60 min	19	27.02 **	41.03 **	3.76 *	6.90 **

This addresses the reviewer's request to compare estimates over the same meetings and calendar windows. Both versions use the meetings admitted by the strict filter. For each meeting, the unfiltered model uses the same first and last times as the strict sample, keeping earlier values only to form the required past changes. The method and lag count are the same as paper Table 3; its strict results are repeated here. Cells show *F*-statistics; stars as above.

Appendix H. Bootstrap Robustness Check

We test the joint null that the other market's three lagged changes have zero coefficients using a restricted wild cluster bootstrap. Webb six-point weights are applied to residuals at the meeting level with 99,999 draws. The bootstrap p-value is

$$p_{\text{boot}} = \frac{1 + \#\{F^* \geq F_{\text{observed}}\}}{99{,}999 + 1},$$

where F^* is the bootstrap test statistic and F_{observed} is the statistic from the original sample.

Table A10. Strict freshness filtering: standard and bootstrap results.

Grid	Direction	Meetings	F	Standard p	Bootstrap p
1 min	PM → ZQ	17	3.24	0.04992	0.16226
1 min	ZQ → PM	17	3.85	0.02990	0.21243
10 min	PM → ZQ	18	26.69	<0.00001	0.00244
10 min	ZQ → PM	18	6.12	0.00514	0.21525
30 min	PM → ZQ	19	27.38	<0.00001	0.00006
30 min	ZQ → PM	19	4.03	0.02343	0.17156
60 min	PM → ZQ	19	41.03	<0.00001	0.00002
60 min	ZQ → PM	19	6.90	0.00273	0.00738

For methodological details, see MacKinnon et al. [24].

References

1. Gürkaynak, R.S.; Sack, B.; Swanson, E.T. Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. *Int. J. Cent. Bank.* **2005**, *1*, 55–93.
2. Ng, H.; Peng, L.; Tao, Y.; Zhou, D. *Price Discovery and Trading in Modern Prediction Markets*; SSRN Working Paper No. 5331995; SSRN: Rochester, NY, USA, 2026. [CrossRef]
3. Hasbrouck, J. One security, many markets: Determining the contributions to price discovery. *J. Finance* **1995**, *50*, 1175–1199. [CrossRef]
4. Diercks, A.M.; Katz, J.D.; Wright, J.H. Kalshi and the Rise of Macro Markets. In *Finance and Economics Discussion Series 2026-010*; Board of Governors of the Federal Reserve System: Washington, DC, USA, 2026. [CrossRef]
5. Swanson, E.T.; Wang, R.; Wu, Y. *The Effects of Monetary Policy on Macroeconomic Expectations: High-Frequency Evidence from Prediction Markets*; SSRN Working Paper No. 6073727; SSRN: Rochester, NY, USA, 2025. [CrossRef]
6. Akey, P.; Grégoire, V.; Harvie, N.; Martineau, C. *Who Wins and Who Loses in Prediction Markets? Evidence from Polymarket*; SSRN Working Paper No. 6443103; SSRN: Rochester, NY, USA, 2026. [CrossRef]
7. Gómez-Cram, R.; Guo, Y.; Jensen, T.I.; Kung, H. *Prediction Market Accuracy: Crowd Wisdom or Informed Minority?*; SSRN Working Paper No. 6617059; SSRN: Rochester, NY, USA, 2026. [CrossRef]
8. Wolfers, J.; Zitzewitz, E. Prediction markets. *J. Econ. Perspect.* **2004**, *18*, 107–126. [CrossRef]
9. Wolfers, J.; Zitzewitz, E. *Interpreting Prediction Market Prices as Probabilities*; NBER Working Paper No. 12200; National Bureau of Economic Research: Cambridge, MA, USA, 2006.
10. Yang, Y. *Pricing Prediction Markets: Incomplete Markets, Selection Rules, and Risk Premia*; SSRN Working Paper No. 6468338; SSRN: Rochester, NY, USA, 2026. [CrossRef]
11. Bürgi, C.; Deng, W.; Whelan, K. *Makers and Takers: The Economics of the Kalshi Prediction Market*; CEPR Discussion Paper No. 20631; Centre for Economic Policy Research: London, UK, 2025.
12. Kuttner, K.N. Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. *J. Monet. Econ.* **2001**, *47*, 523–544. [CrossRef]
13. Bonini, S.; Huang, S.; Simaan, M. Watching the FedWatch. *J. Futur. Mark.* **2026**, *46*, 675–697. [CrossRef]
14. Glosten, L.R.; Milgrom, P.R. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *J. Financ. Econ.* **1985**, *14*, 71–100. [CrossRef]
15. Kyle, A.S. Continuous auctions and insider trading. *Econometrica* **1985**, *53*, 1315–1335. [CrossRef]
16. Eun, C.S.; Sabherwal, S. Cross-border listings and price discovery: Evidence from U.S.-listed Canadian stocks. *J. Finance* **2003**, *58*, 549–575. [CrossRef]

17. Karolyi, G.A. The world of cross-listings and cross-listings of the world: Challenging conventional wisdom. *Rev. Finance* **2006**, *10*, 99–152. [CrossRef]
18. Froot, K.A.; Dabora, E.M. How are stock prices affected by the location of trade? *J. Financ. Econ.* **1999**, *53*, 189–216. [CrossRef]
19. Makarov, I.; Schoar, A. Trading and arbitrage in cryptocurrency markets. *J. Financ. Econ.* **2020**, *135*, 293–319. [CrossRef]
20. Mincer, J.; Zarnowitz, V. The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*; Mincer, J., Ed.; National Bureau of Economic Research: New York, NY, USA, 1969; pp. 3–46.
21. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [CrossRef]
22. Harvey, D.; Leybourne, S.; Newbold, P. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* **1997**, *13*, 281–291. [CrossRef]
23. Hayashi, T.; Yoshida, N. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **2005**, *11*, 359–379. [CrossRef]
24. MacKinnon, J.G.; Nielsen, M.Ø.; Webb, M.D. Fast and reliable jackknife and bootstrap methods for cluster-robust inference. *J. Appl. Econom.* **2023**, *38*, 671–694. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.