# PRICE DISCOVERY ACROSS POLITICAL PREDICTION MARKETS: EVIDENCE FROM THE 2024 U.S. PRESIDENTIAL ELECTION*

**EMREHAN AKTUĞ**†
*Sabancı University*

**ORHAN TORUL**‡
*Boğaziçi University*

April 2026

**Abstract**

We bring standard price discovery tools from the equity microstructure literature to nine political betting platforms during the 2024 U.S. presidential election, the first such analysis for prediction markets. Exchange-based venues dominate: PolyMarket and Betfair jointly account for nearly 85% of Hasbrouck information shares, and all sportsbooks are excluded from the Model Confidence Set. Leadership is frequency-dependent—PolyMarket's automated market maker leads at 5-minute frequency, while Betfair's limit order book leads at 1-hour frequency—a crossover confirmed by spectral connectedness decomposition and cross-market jump timing. The CFTC-regulated exchange Kalshi attains nontrivial information shares within weeks of its October 2024 launch, though formal tests cannot distinguish its performance from less informative venues. These findings speak to the regulatory debate: prediction markets organized as exchanges appear to serve a substantive price discovery function, even without the centralized linkages that discipline equity market fragmentation.

**Keywords**: Price Discovery; Prediction Markets; Information Shares; Market Microstructure; Political Betting; Electoral Probabilities

**JEL Classifications:** D72; D83; G14; G18

*We are grateful to Koray Şimşek, Yiğit Atılgan, Burak Saltoğlu, Şerif Aziz Şimşir, and Hüseyin Gülen for their valuable feedback. All remaining errors are ours.

†Corresponding author. Address: Sabancı University Department of Economics, 34956, Tuzla, Istanbul, Türkiye.
E-Mail: emrehan.aktug@sabanciuniv.edu

‡Address: Boğaziçi University, Department of Economics, 34342, Bebek, Istanbul, Türkiye.
E-Mail: orhan.torul@bogazici.edu.tr

# 1 Introduction

The 2024 U.S. presidential election was traded simultaneously on nine betting platforms spanning multiple continents and architecturally distinct market designs: a blockchain-based automated market maker (PolyMarket, \$3.6 billion in election-related volume), traditional limit order book exchanges (Betfair, Kalshi), and six sportsbooks. Kalshi, the only CFTC-regulated venue, began listing election contracts in October 2024 after a protracted legal battle over whether political event contracts serve a legitimate economic purpose. This proliferation of venues, trading the same binary event across incompatible jurisdictions and architectures, creates a natural laboratory for studying how market structure shapes price discovery.

What makes this setting distinctive is the severity of its frictions. In U.S. equity markets, fragmentation is disciplined by Regulation NMS, the National Best Bid and Offer, and high-frequency arbitrageurs who enforce the law of one price within milliseconds. None of these mechanisms exists in political prediction markets. U.S. residents cannot legally access Betfair; offshore crypto platforms are barred to U.S. persons; capital is immobilized across venues with no netting; and settlement horizons range from minutes (on-chain) to days (sportsbook withdrawals). The average cross-platform spread on the same contract is 6.2 percentage points. These frictions are analogous to the limits to arbitrage emphasized by Shleifer and Vishny (1997). They make the question of which venue *produces* the efficient probability both substantive and unresolved.

We bring the standard price discovery toolkit from the equity microstructure literature (the Hasbrouck (1995) information share, the Gonzalo and Granger (1995) component share, and the Putniņš (2013) information leadership share) to 5-minute probability data from these nine platforms over January–November 2024. To our knowledge, these decompositions have not previously been applied to prediction markets. Unlike simple cross-correlations, which cannot distinguish a market that *independently discovers* information from one that merely *replicates* price changes quickly, the VECM framework formally identifies each venue's contribution to the common efficient price.

Our analysis yields three main results. First, exchange-based venues dominate price discovery: PolyMarket and Betfair jointly account for nearly 85% of information shares in the clean panel, while no sportsbook exceeds 11%. Second, information leadership is frequency-dependent. PolyMarket's AMM leads at 5-minute frequency, while Betfair's limit order book leads at 1-hour frequency, a crossover consistent with a trade-off between execution latency and market depth. Different exchange architectures thus appear to serve complementary informational roles rather than simply competing for the same function. Third, the CFTC-regulated exchange Kalshi attains nontrivial information share point estimates within 32 days of its launch, consistent with the hypothesis that exchange architecture, rather than incumbency or accumulated liquidity, facilitates informed trading. The short window and confounding with election proximity preclude strong causal inference.

These results contribute to the literature on market fragmentation and price discovery (Hasbrouck, 1995; O'Hara and Ye, 2011; Makarov and Schoar, 2020). We extend the Hasbrouck–Gonzalo-Granger framework to an asset class where cross-venue arbitrage is severely constrained, and document a frequency-dependent complementarity between AMM and LOB architectures that has not, to our knowledge, been identified in prior work. The Kalshi episode, while resting on a short sample that permits only descriptive analysis, illustrates how a newly launched exchange can attain nontrivial information shares. Our analysis is descriptive throughout: the patterns are consistent with architecture playing an important role in price discovery, but we cannot rule out confounding factors such as differences in trader composition or the proximity to the election.

## 2 Institutional Background and Related Literature

### 2.1 Institutional Setting

Three architectural types coexist in this market. PolyMarket operates on the Polygon blockchain from a base in Antigua; its AMM-based trading mechanism reprices with each trade and settles on-chain within minutes.[1] Betfair Exchange, regulated by the UK Gambling Commission, and Kalshi, a CFTC-regulated designated contract market, both operate centralized limit order books where human participants post and revise limit orders. Kalshi's entry on October 4, 2024, following the District Court ruling in *Kalshi v. CFTC*, is the key institutional event in our sample: it introduced a U.S.-regulated exchange into an ecosystem previously accessible only through offshore platforms.

Six traditional sportsbooks complete the sample: Pinnacle, Bovada, and BetOnline (sharp bookmakers with tighter margins) and William Hill, Unibet, and Everygame (retail-focused bookmakers with wider margins). Unlike exchanges, sportsbooks set odds through internal risk models and adjust infrequently, creating the price staleness that is central to our empirical analysis. Table 1 in Section 3 provides a detailed comparison of all nine platforms.

### 2.2 Related Literature

*Price discovery and market fragmentation.* Hasbrouck (1995) introduced the information share (IS), measuring each market's contribution to the variance of the efficient price innovation from the moving-average representation of a VECM. Gonzalo and Granger (1995) proposed the component share (CS), an ordering-invariant alternative based on the error correction coefficients. Baillie et al. (2002) and De Jong (2002) compare the two, and Putniņš (2013) proposes the Information Leadership Share (ILS) to disentangle informational from speed advantages. These tools have been applied to equities (Hasbrouck, 1995), foreign exchange (Lehmann, 2002), futures (Baillie et al., 2002), ETFs, stock–option pairs (Chakravarty et al., 2004), and cryptocurrency exchanges (Makarov and Schoar, 2020), but not, to our knowledge, to prediction markets.

Our setting connects to the broader fragmentation literature. O'Hara and Ye (2011) show that fragmentation across U.S. equity venues improves market quality, while Foucault and Menkveld (2008) model how smart order routing affects price discovery across venues. Makarov and Schoar (2020) document large and persistent price differences across cryptocurrency exchanges, attributing them to capital controls and withdrawal frictions—an analogy that maps closely to our setting, where jurisdictional barriers and capital lockup

[1] PolyMarket's architecture has evolved over time and incorporates elements of both automated market making and order matching. We refer to it as an "AMM" throughout for expositional clarity, as its pricing mechanism remains algorithmically determined rather than set by human limit orders. In the data, PolyMarket updates in only 1.1% of 5-minute bins, reflecting its low baseline trading frequency outside of news events; this is substantially less than Betfair (5.4%), though still an order of magnitude above sportsbooks (0.1–0.4%).

similarly prevent cross-venue arbitrage.

*Exchange architecture.* Lehar and Parlour (2021) model how AMMs aggregate information differently from limit order books, showing that AMM prices can be informationally efficient but expose liquidity providers to adverse selection. Croxson and Reade (2014) study information incorporation in Betfair's betting exchange, documenting near-instantaneous price adjustment to public news. Our comparison of PolyMarket's AMM against Betfair's and Kalshi's LOBs provides a direct empirical test of how these architectural differences shape price discovery. In Appendix B, we develop a conceptual framework formalizing the key trade-off: AMMs offer zero latency but shallow depth, while LOBs have positive latency but greater depth, generating a frequency-dependent crossover in price discovery leadership.

*Prediction markets.* A large literature examines prediction market accuracy, calibration, and design (Wolfers and Zitzewitz, 2008; Arrow et al., 2008; Berg et al., 2008; Manski, 2006; Snowberg et al., 2007; Rothschild, 2009; Hanson et al., 2006). Our paper differs in focus: rather than asking whether prediction markets are accurate, we ask which *type* of prediction market venue is informationally dominant. Most closely related are Flynn and Tarkom (2025), who document that betting markets incorporate political information faster than equity markets during the 2024 election, and Aktuğ and Torul (2026), who use prediction market probabilities to estimate electoral risk premia for partisan firms. Both take prediction market prices as given; we study the prediction markets *themselves* to understand which venue produces the efficient probability.

### 2.3 Hypotheses

The institutional differences across our nine platforms generate testable predictions about the cross-section and dynamics of information shares. We formalize these predictions in a conceptual framework of venue choice (Appendix B); here we state the hypotheses informally and note the corresponding theoretical result.

**H1: Exchange architecture dominates sportsbooks in price discovery.** Exchange-based venues (PolyMarket, Betfair, Kalshi) feature transparent order flow, continuous pricing, and attract sophisticated participants. Sportsbooks set prices through internal models and adjust infrequently, tracking the efficient price set elsewhere with a lag (Proposition 4). The market microstructure literature predicts that venues with continuous, transparent price formation should lead price discovery (Madhavan, 2000; Hasbrouck, 1995).

**H2: A regulated exchange can rapidly attain nontrivial information shares.** If exchange architecture facilitates price discovery, then a newly launched exchange (Kalshi) should attain nontrivial information shares within a short period, provided it offers comparable trading infrastructure. The framework shows that a regulated exchange that unlocks a previously excluded pool of informed traders can gain meaningful informational presence even without matching the incumbent's volume (Proposition 3).

**H3: Information leadership shifts from traditional venues to exchange-based platforms as information intensity increases.** If information shares reflect the equilibrium allocation of informed traders across venues, then the rise in election salience and trading volume over the campaign should drive a reallocation of price discovery from traditional platforms (which dominated the low-volatility early period) to exchange-based venues (which attract sophisticated, high-frequency participants). We predict a regime shift around the July 2024 political shocks, when the campaign's salience and the volume of informed order flow increased sharply. Information shares should also respond discontinuously to major political shocks that alter participation patterns.

**H4: The time horizon of price discovery depends on the price-setting mechanism.** PolyMarket's AMM provides instantaneous algorithmic price adjustment (zero latency), while Betfair and Kalshi rely on human participants posting and revising limit orders (positive latency but greater depth). The framework predicts that low-latency, thin-depth AMMs should dominate at high frequencies, while deeper LOBs should dominate at lower frequencies as cumulative informed order flow aggregates (Proposition 2).

# 3 Data

## 3.1 Sources and Sample Construction

Table 1 summarizes the nine platforms in our sample. We obtain 5-minute snapshots of Trump victory implied probabilities from each platform over their respective coverage periods.

**Table 1:** Data Summary

| Platform | Architecture | Obs | Period | Coverage (%) | Prob Range |
|---|---|---|---|---|---|
| PolyMarket | Prediction Market (DEX) | 188,406 | Jan 2023–Nov 2024 | 100.0 | 0.44–0.75 |
| Betfair | Betting Exchange (P2P) | 89,080 | Dec 2023–Nov 2024 | 99.8 | 0.38–0.71 |
| Pinnacle | Sharp Sportsbook | 35,459 | Jul 2024–Nov 2024 | 39.8 | 0.47–0.76 |
| Bovada | Sharp Sportsbook | 88,866 | Dec 2023–Nov 2024 | 99.7 | 0.42–0.79 |
| BetOnline | Sharp Sportsbook | 71,195 | Dec 2023–Nov 2024 | 79.9 | 0.45–0.75 |
| William Hill | Retail Sportsbook | 53,993 | Dec 2023–Nov 2024 | 60.8 | 0.45–0.75 |
| Unibet | Retail Sportsbook | 85,785 | Dec 2023–Nov 2024 | 96.2 | 0.44–0.80 |
| Everygame | Retail Sportsbook | 47,519 | Dec 2023–Nov 2024 | 53.5 | 0.27–0.67 |
| Kalshi | Regulated Exchange (CFTC) | 9,358 | Oct 4–Nov 5, 2024 | — | 0.49–0.58 |

*Notes:* Coverage indicates the percentage of 5-minute grid periods with valid observations during the primary sample period. Sportsbook decimal odds are converted to implied probabilities via $p = 1/\text{odds}$. PolyMarket and Kalshi report implied probabilities directly. Pinnacle and Kalshi data reflect their respective mid-sample entries.

PolyMarket probabilities reflect the marginal cost of acquiring outcome tokens from its AMM. Kalshi probabilities are last transaction prices on its limit order book. Sportsbook decimal odds are converted to implied probabilities via $p = 1/d$.[2]

We align the series to a 5-minute UTC grid using last-observation resampling and forward-fill gaps of up to 30 minutes. We define four analysis panels over the full January–November 2024 period (Kalshi, which enters on October 4, enters the nine-market panel only): (i) a *clean panel* of six markets (excluding Pinnacle, which enters in July, and Everygame, which exhibits a staleness artifact documented in Section 5.2); (ii) a *common panel* of seven markets spanning the full sample period (excluding Pinnacle); (iii) an *eight-market panel* including Pinnacle from July onward; and (iv) a *Kalshi-window panel* of nine markets during October 4 to November 5. The clean panel serves as our primary specification; the other panels are used for robustness. Two caveats deserve attention. First, sportsbook odds are obtained from periodic snapshots of publicly posted

[2] This conversion does not remove the bookmaker's overround (vig). Since all sportsbooks apply a similar markup and we study *changes* in probabilities rather than levels, the vig does not materially affect the price discovery analysis.

odds; the exact timing of oddsmaker updates is unobserved, and some degree of stale pricing is inherent in this data-collection approach. Second, the forward-fill procedure mechanically introduces staleness for markets that update infrequently: if a venue does not post a new price during a 5-minute interval, the previous price is carried forward, producing a zero innovation in the VECM residual. When the venue eventually updates, the accumulated catch-up adjustment manifests as a single large jump whose squared magnitude can exceed the sum of the many small squared innovations that would have occurred under continuous updating. As documented in Section 5.2, this mechanism can inflate information shares for infrequently updating venues (particularly Everygame) and should be borne in mind when interpreting the cross-section of IS estimates.

### 3.2 Logit Transformation

Implied probabilities are bounded on $[0,1]$ and exhibit trend-like dynamics over the campaign period. To ensure the innovation terms are not bounded and to improve the defensibility of the unit root assumptions, we apply the logit transformation $\ell_{i,t} = \ln(p_{i,t}/(1-p_{i,t}))$. All subsequent econometric analysis is conducted on the logit-transformed series. Section 6 demonstrates that our primary results are robust to the use of raw probabilities.

### 3.3 Descriptive Statistics

Table 2 reports summary statistics for the 5-minute first differences of the probabilities. The statistics are computed over the 10,491 complete-case observations in the eight-market panel (July 2024 onward, when Pinnacle enters); Pinnacle is omitted from the table for brevity. The VECM estimation in Section 5 uses the seven-market common panel (16,460 observations), which excludes Pinnacle to maintain a balanced sample from January onward.

**Table 2:** Descriptive Statistics of 5-Minute Probability Changes

| Market | Mean (bp) | Std (bp) | Skewness | Kurtosis | Updates (%) | Roll Spread (bp) |
|---|---|---|---|---|---|---|
| PolyMarket | −0.07 | 27.0 | −2.25 | 2500 | 1.1 | 11.5 |
| Betfair | −0.06 | 13.6 | −0.57 | 25.4 | 5.4 | 7.3 |
| Bovada | 0.03 | 9.1 | 0.54 | 430 | 0.2 | — |
| BetOnline | −0.09 | 11.8 | −3.48 | 204 | 0.4 | — |
| William Hill | −0.03 | 11.8 | 4.78 | 465 | 0.2 | — |
| Unibet | −0.06 | 11.9 | −2.08 | 359 | 0.2 | — |
| Everygame | −0.04 | 51.5 | −0.52 | 4712 | 0.1 | 33.2 |

*Notes:* Changes are 5-minute first differences of raw (untransformed) implied probabilities, in basis points. The VECM estimation in Section 5 uses logit-transformed series, which alter the variance structure. Updates (%) represents the fraction of periods in which the price changed. Roll Spread is the Hasbrouck-style effective spread. Sample: 10,491 complete-case observations from the eight-market panel (July 2024 onward); Pinnacle omitted from table for brevity.

Two features of the data are central to the econometric analysis. First, update frequencies differ by more than an order of magnitude: Betfair updates in 5.4% of 5-minute bins, PolyMarket in 1.1%, and sportsbooks in just 0.1–0.4%. This heterogeneity means that in any given 5-minute window, most sportsbook “innovations” are mechanically zero, a stale price carried forward, which directly affects the VECM residual covariance and hence the estimated information shares. We address this concern through four complementary tests in Sections 5.2–6.6: exclusion of the most stale venue, a fresh-quote filter that conditions on active updating, Monte Carlo simulation of the staleness bias, and re-estimation at lower sampling frequencies where staleness is less severe. Second, Everygame's standard deviation (51.5 bp) is nearly twice that of the next-highest venue (PolyMarket, 27.0 bp), reflecting infrequent but large discrete adjustments. Its residuals exhibit serial correlation (Ljung-Box $p < 0.01$), unlike all other markets, foreshadowing the estimation artifact discussed in Section 5.2.

### 3.4 Data Limitations

Two data limitations should be noted. First, we lack platform-level trading volume. While PolyMarket's aggregate volume ($3.6 billion) is publicly available, comparable figures for other venues are proprietary or unreported at the contract level. Our interpretations of the mechanisms underlying information shares, particularly the role of liquidity in attracting informed traders, are therefore based on institutional knowledge rather than direct measurement.

Second, Kalshi's probability range during its 32-day window (0.49–0.58) is considerably narrower than other venues. Near $p = 0.5$, the logit transformation is approximately linear; at the extremes observed for other platforms ($p \approx 0.38$ or 0.80), the logit substantially amplifies variation. This difference in dynamic range should be borne in mind when comparing Kalshi-window IS estimates to full-sample estimates.

### 3.5 Key Political Events

Five events anchor our event-window analysis. Four are political shocks, each generating probability movements of 5 percentage points or more within 48 hours: (i) the Biden–Trump debate (June 27), which triggered Biden's collapse from ~45% to ~35%; (ii) the Trump assassination attempt (July 13); (iii) the Biden withdrawal and Harris endorsement (July 21); and (iv) Election Day (November 5). We also include (v) the Kalshi market launch (October 4) as a market structure event that altered the venue landscape. Figure 1 plots the implied probabilities across the full sample with vertical lines at each event.

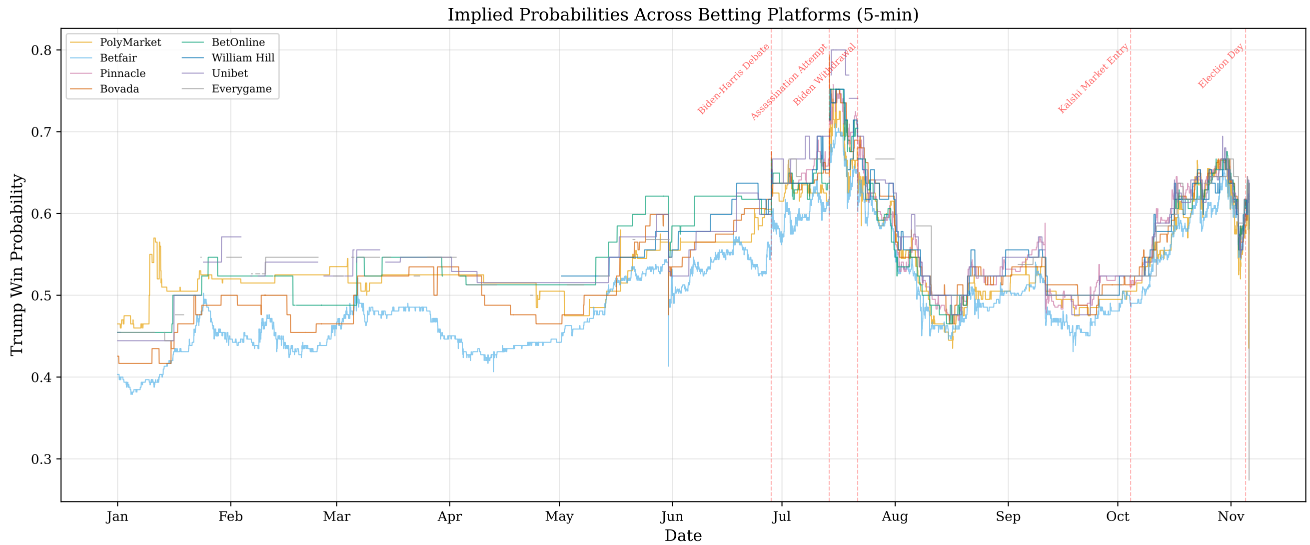


**Figure 1:** Implied Trump victory probabilities across eight platforms at 5-minute frequency, January–November 2024 (Kalshi excluded due to its shorter sample; Pinnacle enters in July). Vertical lines indicate key political events.

# 4 Econometric Methodology

## 4.1 Unit Root and Cointegration Tests

We test each logit-transformed series for a unit root using the Augmented Dickey–Fuller (ADF; Dickey and Fuller, 1979) test with automatic lag selection and the KPSS test (which has stationarity as its null). We test for cointegration using the Johansen (1991) trace and maximum-eigenvalue tests. The logit transformation is essential here: raw probabilities are bounded on $[0,1]$ and cannot be unit root processes in the strict sense, whereas logit-transformed series are unbounded and can exhibit the persistent, mean-reverting-only-through-error-correction dynamics that the VECM framework requires.

With $n$ markets tracking the same underlying binary event, the theoretical prediction is $n-1$ cointegrating relations (one common stochastic trend: the efficient probability). Deviations from this rank would indicate that some markets track systematically different information sets.

## 4.2 Vector Error Correction Model

Let $\boldsymbol{\ell}_t = (\ell_{1t}, \ldots, \ell_{nt})'$ denote the vector of logit-transformed probabilities. If the series are cointegrated with rank $r$, the VECM representation is:

$$\Delta \boldsymbol{\ell}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\ell}_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{\ell}_{t-i} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim (0, \boldsymbol{\Sigma}) \tag{1}$$

where $\boldsymbol{\alpha}$ $(n \times r)$ contains the error correction (adjustment) coefficients, $\boldsymbol{\beta}$ $(n \times r)$ the cointegrating vectors, and $\boldsymbol{\Gamma}_i$ the short-run dynamics. The lag order $k$ is selected by the Bayesian Information Criterion (BIC) applied to the unrestricted VAR on first differences. The BIC selects $k = 17$ for the common seven-market panel, corresponding to 85 minutes of dynamics. This relatively high lag order reflects the stale-price dynamics generated by infrequently updating sportsbooks: when several venues update only in 0.1–0.2% of 5-minute bins (Table 2), the autocorrelation structure of their residuals extends over many lags, requiring a rich lag structure to achieve white-noise innovations. Section 6 confirms that the IS ranking is stable across $k \in \{13, 15, 17, 19, 21\}$.

### 4.3 Hasbrouck Information Shares

The Granger Representation Theorem implies the moving-average form:

$$\Delta \boldsymbol{\ell}_t = \boldsymbol{\Psi}(1)\boldsymbol{\varepsilon}_t + (1-L)\boldsymbol{\Psi}^*(L)\boldsymbol{\varepsilon}_t \tag{2}$$

where $\boldsymbol{\Psi}(1) = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}'_\perp \boldsymbol{\Xi} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}'_\perp$ is the long-run impact matrix, with $\boldsymbol{\Xi} = \boldsymbol{I} - \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i$. When $r = n-1$, all rows of $\boldsymbol{\Psi}(1)$ are proportional, and we define $\boldsymbol{\psi}$ as the (normalized) common row vector representing each market's weight in the efficient price innovation.

The Hasbrouck (1995) information share of market $j$ is:

$$\mathrm{IS}_j = \frac{(\boldsymbol{\psi}'\mathbf{F}_{\cdot j})^2}{\boldsymbol{\psi}'\boldsymbol{\Sigma}\boldsymbol{\psi}} \tag{3}$$

where $\mathbf{F}$ is the lower-triangular Cholesky factor of the residual covariance matrix, $\boldsymbol{\Sigma} = \mathbf{F}\mathbf{F}'$, and $\mathbf{F}_{\cdot j}$ denotes its $j$th column. Because the Cholesky factorization depends on the ordering of variables, we compute $\mathrm{IS}_j$ for all $n!$ orderings and report upper and lower bounds, as well as the midpoint (Hasbrouck, 1995; Baillie et al., 2002). For the nine-market Kalshi-window system (9! = 362,880), we sample 10,000 random orderings to approximate the bounds.

### 4.4 Gonzalo–Granger Component Shares

The Gonzalo and Granger (1995) permanent–transitory decomposition defines the permanent (efficient price) component as $\boldsymbol{\alpha}'_\perp \boldsymbol{\ell}_t$. The component share of market $j$ is:

$$\mathrm{CS}_j = \frac{(\boldsymbol{\alpha}_\perp)_j}{\sum_{i=1}^{n} (\boldsymbol{\alpha}_\perp)_i} \tag{4}$$

This measure is point-identified (no ordering dependence) but can be negative, indicating systematic overshooting. It uses only the error correction coefficients and is thus independent of the residual covariance structure (Gonzalo and Granger, 1995; De Jong, 2002; Putniņš, 2013). To disentangle the two channels through which a market can lead—receiving information first versus adjusting more rapidly—we also compute the Putniņš (2013) Information Leadership Share, $\mathrm{ILS}_j = \mathrm{IS}_j/\mathrm{CS}_j$. Markets with ILS > 1 have an *informational* advantage (they contribute more to innovation variance than their error-correction weight implies), while ILS < 1 indicates a *speed* advantage (rapid adjustment without proportional informational contribution). The ILS is defined only for markets with $\mathrm{CS}_j > 0.01$; markets with near-zero CS produce unstable ratios.

### 4.5 Modified Information Shares

The ordering dependence of Hasbrouck IS is a well-known limitation. Following Lien and Shrestha (2009), we also compute the *Modified Information Share* (MIS), which replaces the covariance matrix $\mathbf{\Sigma}$ with the correlation matrix $\mathbf{R}$ in the Cholesky decomposition:

$$\mathrm{MIS}_j = \frac{\left(\boldsymbol{\psi}'\mathbf{D}^{1/2}\mathbf{M}_{\cdot j}\right)^2}{\sum_{k=1}^{n}\left(\boldsymbol{\psi}'\mathbf{D}^{1/2}\mathbf{M}_{\cdot k}\right)^2} \tag{5}$$

where $\mathbf{D}^{1/2} = \mathrm{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}})$ and $\mathbf{M}$ is the lower-triangular Cholesky factor of the correlation matrix $\mathbf{R} = (\mathbf{D}^{1/2})^{-1}\mathbf{\Sigma}(\mathbf{D}^{1/2})^{-1}$. The MIS is approximately ordering-invariant when cross-market residual correlations are small (Lien and Shrestha, 2009); in our data, where contemporaneous correlations of 5-minute changes are near zero (0.002–0.051), the approximation is tight and MIS values are nearly identical across orderings. We use MIS as an alternative numerator in the Putniņš (2013) ILS computation.

### 4.6 Weak Exogeneity Tests

A market is weakly exogenous in the VECM if its row of $\boldsymbol{\alpha}$ is zero—meaning it does not adjust to cointegrating errors and is thus a pure price leader. For each market $j$, we test $H_0\colon \alpha_{j,1} = \cdots = \alpha_{j,r} = 0$ using a Wald $\chi^2$ test with $r$ degrees of freedom. Rejection indicates the market participates in error correction; the magnitude of the test statistic reveals the *intensity* of adjustment, providing a formal complement to the IS and CS measures.

### 4.7 Connectedness Analysis

We complement the VECM-based price discovery measures with the Diebold and Yılmaz (2012, 2014) connectedness framework, which constructs a *directed information flow network* from generalized forecast error variance decompositions (GFEVD; Pesaran and Shin, 1998). Unlike the IS, which identifies each market's share of the efficient price innovation, the connectedness table shows *who transmits information to whom.* For each market $j$, we compute directional spillovers: FROM (share of $j$'s forecast error variance due to other markets), TO (share of other markets' variance due to $j$), and NET = TO − FROM. A positive NET identifies net information transmitters. The total connectedness index (TCI) summarizes system-wide integration. We estimate the GFEVD at a 12-period (1-hour) forecast horizon using the VMA representation derived from the VECM.

Following Baruník and Křehlík (2018), we decompose the connectedness measures into frequency bands—short-run (<30 minutes), medium-run (30 minutes to 2 hours), and long-run (>2 hours)—using spectral methods applied to the VMA representation. This provides a within-model test of frequency-dependent leadership, without requiring re-estimation at different sampling frequencies.

### 4.8 Bootstrap Confidence Intervals

We construct confidence intervals using two complementary approaches. Our primary method is the stationary bootstrap of Politis and Romano (1994) with automatic block length selection following Politis and White (2004) and Patton et al. (2009). Unlike the fixed-block bootstrap, the stationary bootstrap draws geometrically distributed block lengths, yielding a strictly stationary resampling scheme. The data-driven optimal block length (22 periods for the common panel) adapts to the dependence structure of each series. We apply the bias-corrected accelerated (BCa) correction of Efron and Tibshirani (1993) to address the first-order bias arising from the bounded $[0,1]$ support of IS estimates.

As a robustness check for conditional heteroskedasticity (volatility spikes around political events), we also implement a recursive-design wild bootstrap (Cavaliere and Taylor, 2008; Gonçalves and Kilian, 2004). Each replication multiplies VECM residuals by Rademacher random variables ($\pm 1$), preserving the conditional variance profile, then recursively constructs bootstrap levels using the estimated VECM dynamics.

### 4.9 Time-Varying and Event-Window Analysis

To examine how information leadership evolves, we estimate rolling-window VECMs with a 2-week window (4,032 five-minute periods) stepped by 1 day (288 periods), yielding approximately 44 overlapping windows across the common sample. The 2-week window represents a trade-off: shorter windows would capture finer dynamics but yield imprecise estimates in a 7-market VECM with 17 lags (119 short-run parameters per equation, 833 system-wide); longer windows would smooth over the event-driven regime shifts that are central to our analysis. For event-window analysis, we estimate separate VECMs in windows of 7 or 14 days centered on each of the five key political events.

## 5 Results

### 5.1 Preliminary Tests

#### 5.1.1 Unit Roots

Table 3 reports unit root test results for the seven common-panel markets. At the 5% level, all logit-transformed series fail to reject the ADF null of a unit root ($p$-values 0.08–0.32) and reject the KPSS null of stationarity ($p$-values $< 0.01$). First differences uniformly reject the ADF null ($p$-values $< 0.001$) and fail to reject KPSS. We conclude that all series are integrated of order one, $I(1)$.

#### 5.1.2 Cointegration

The Johansen trace test applied to the seven-market system (with BIC-selected lag order $k = 17$) rejects the null of rank $\leq r$ for $r = 0,1,2,3,4$ at the 5% level, consistent with at least five cointegrating relations. Given that all seven markets track the same binary event, we impose the theoretically motivated rank $r = n-1 = 6$ (one common stochastic trend). This choice warrants discussion: the statistical evidence supports rank 5, not 6. The gap could reflect genuine informational segmentation—some markets may track partially distinct information sets—or finite-sample distortions of the trace test in a system with stale-price dynamics. We maintain $r = n-1$ as our baseline because it corresponds to the strongest theoretical prior (a single underlying binary event), but Section 6 demonstrates that results with $r = 5$ preserve the qualitative ranking of venues.

**Table 3:** Unit Root Tests on Logit-Transformed Probabilities

| | Levels | | | First Differences | | |
|---|---|---|---|---|---|---|
| Market | ADF Stat | ADF $p$ | KPSS Stat | ADF Stat | ADF $p$ | KPSS Stat |
| PolyMarket | −1.922 | 0.322 | 5.203** | −20.261 | 0.000 | 0.152 |
| Betfair | −2.446 | 0.129 | 6.707** | −35.274 | 0.000 | 0.150 |
| Bovada | −2.544 | 0.105 | 5.208** | −93.890 | 0.000 | 0.160 |
| BetOnline | −2.540 | 0.106 | 2.902** | −54.953 | 0.000 | 0.184 |
| William Hill | −2.670 | 0.079 | 4.231** | −128.285 | 0.000 | 0.118 |
| Unibet | −2.616 | 0.090 | 3.360** | −53.153 | 0.000 | 0.154 |
| Everygame | −2.525 | 0.110 | 5.085** | −24.483 | 0.000 | 0.134 |

*Notes:* ADF tests include an intercept (no trend) with automatic lag selection (AIC). KPSS tests use the Newey–West automatic bandwidth with a level specification. ** indicates rejection at the 1% level. The KPSS critical value at 5% is 0.463. Sample: 16,460 complete-case observations at 5-minute frequency.

For the nine-market Kalshi-window system ($k = 3$, BIC-selected), both trace and maximum-eigenvalue tests indicate rank = 8, exactly $n - 1$.

## 5.2 Full-Sample Information Shares

Table 4 reports Hasbrouck information shares and Gonzalo–Granger component shares for the six-market clean panel (excluding Everygame), estimated on 16,460 complete-case observations.[3]

**Table 4:** Information Shares and Component Shares: Full Sample, Clean Panel (January–November 2024)

| | Hasbrouck IS | | | | |
|---|---|---|---|---|---|
| Market | Lower | Mid | Upper | CS | Arch. |
| PolyMarket | 0.375 | **0.475** | 0.576 | 0.370 | DEX |
| Betfair | 0.267 | 0.371 | 0.474 | 0.414 | Exchange |
| BetOnline | 0.027 | 0.110 | 0.193 | 0.175 | Sharp |
| Bovada | 0.006 | 0.059 | 0.112 | 0.111 | Sharp |
| Unibet | 0.004 | 0.044 | 0.084 | 0.078 | Retail |
| William Hill | 0.000 | 0.024 | 0.048 | −0.148 | Retail |

*Notes:* 6-market VECM excluding Everygame. Hasbrouck bounds computed over all 6! = 720 Cholesky orderings. CS (Gonzalo–Granger component shares) are point-identified. VECM estimated with $k = 17$ lags and cointegrating rank $r = 5$.

PolyMarket's IS midpoint of 47.5% is the largest in the system, followed by Betfair at 37.1%. Together these two exchange venues account for the vast majority of the variance of the common efficient price innovation, strongly supporting H1 (exchange architecture dominates sportsbooks). No sportsbook IS midpoint exceeds 11%. The CS broadly confirm the exchange-over-sportsbook hierarchy, though with a different internal ordering: Betfair (0.41) slightly exceeds PolyMarket (0.37) in the permanent-component weight, while William Hill

[3] Everygame, a retail sportsbook that updates in only 0.1% of 5-minute bins, generates an artifactual IS of 21% due to price staleness (see Section 6.6). Its residuals exhibit serial correlation (Ljung-Box $p < 0.01$), confirming model misspecification. The 7-market specification including Everygame is reported in Appendix Table A.1 for completeness; the exchange-dominance finding is unchanged.

receives a negative weight (−0.15), indicating systematic overshooting relative to the efficient price.

*Bootstrap confidence intervals are wide.* In the 7-market specification (Appendix Table A.1), PolyMarket's 90% bootstrap CI spans [1.8%, 84.4%]. This width is typical for Hasbrouck IS estimation, as the Cholesky decomposition is sensitive to small perturbations in the residual covariance matrix $\mathbf{\Sigma}$ (Baillie et al., 2002). The bootstrap CIs should not be confused with the Cholesky ordering bounds, which are narrow for PolyMarket [37.5%, 57.6%] and confirm that its point estimate is not an artifact of variable ordering. The IS *ranking* should be interpreted as a pattern supported by consistent point estimates across multiple robustness checks (Section 6), not as a set of precisely estimated quantities.

Figure 2 displays the Hasbrouck information shares with bounds.

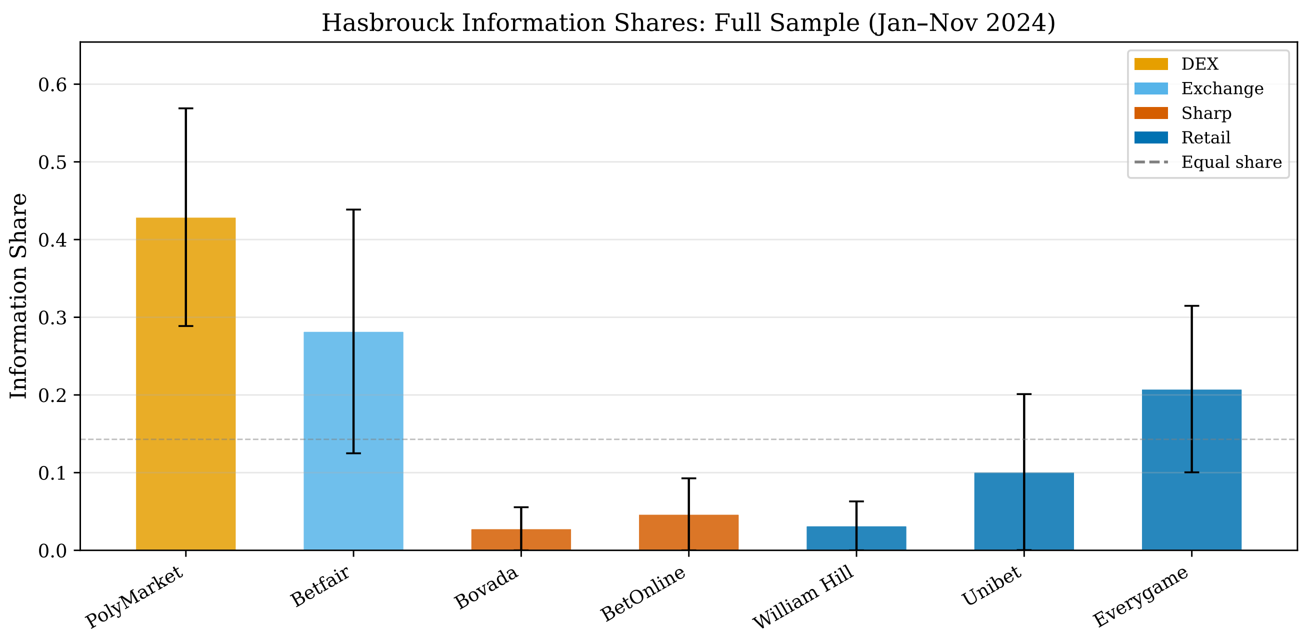


**Figure 2:** Hasbrouck information shares for the seven-market system (full sample, including Everygame for visual completeness; the primary clean-panel specification in Table 4 excludes Everygame). Error bars show upper and lower bounds across all Cholesky orderings. Markets are color-coded by architecture.

### 5.3 Kalshi-Window Information Shares

Table 5 reports results for the nine-market system during October 4–November 5, 2024 (5,658 complete-case observations).

PolyMarket (44%) and Kalshi (41%) jointly account for 85% of price discovery. The Hasbrouck bounds are tight (upper–lower spread of 3–7 percentage points for the leading markets), indicating that this ordering is robust to the Cholesky factorization. All traditional sportsbook IS midpoints fall below 5%. Betfair's IS falls sharply from 37.1% in the full-sample clean panel to 5.7% here. This decline reflects the joint effect of expanding from a 6-market to a 9-market system, a much shorter lag order ($k = 3$ vs. $k = 17$), and the concentration of information shares in PolyMarket and Kalshi during the final month of the campaign.

**Table 5:** Information Shares: Kalshi Window (October 4–November 5, 2024)

| | Hasbrouck IS | | | | |
|---|---|---|---|---|---|
| Market | Lower | Mid | Upper | CS | Arch. |
| PolyMarket | 0.403 | **0.438** | 0.474 | 0.298 | DEX |
| Kalshi | 0.378 | **0.409** | 0.441 | 0.242 | Reg. Exch. |
| Betfair | 0.035 | 0.057 | 0.079 | 0.205 | Exchange |
| Pinnacle | 0.025 | 0.031 | 0.036 | 0.224 | Sharp |
| Bovada | 0.008 | 0.018 | 0.027 | 0.133 | Sharp |
| Everygame | 0.044 | 0.048 | 0.053 | −0.061 | Retail |
| William Hill | 0.001 | 0.002 | 0.003 | 0.032 | Retail |
| BetOnline | 0.000 | 0.002 | 0.003 | −0.058 | Sharp |
| Unibet | 0.000 | 0.000 | 0.001 | −0.016 | Retail |

*Notes:* Hasbrouck bounds approximated using 10,000 random orderings (of $9! = 362{,}880$ total). CS sum to 1. VECM with $k = 3$ lags and rank $r = 8$.

The CS tell a complementary story: during the Kalshi window, the three exchange-based venues (PolyMarket 0.298, Kalshi 0.242, Betfair 0.205) collectively receive 74% of the permanent-component weight, with Pinnacle (0.224) capturing a meaningful share that its low IS does not suggest. This divergence (Pinnacle having a high CS but low IS) indicates that Pinnacle's price is little adjusted toward the other venues—so it carries substantial permanent-component weight—yet its own innovations contribute relatively little variance. The low Putniņš ILS (= IS/CS ≈ 0.14) that follows implies a speed, rather than informational, contribution: Pinnacle's permanent-component weight derives from its slow error correction rather than from generating new information.

That Kalshi attained nontrivial IS point estimates within weeks of launch, consistent with H2 and Proposition 3, carries suggestive regulatory implications, though formal tests (Section 5.14) reject that Kalshi's IS is statistically comparable to PolyMarket's. The conceptual framework identifies a possible mechanism: a regulated exchange that opens access to a previously excluded pool of informed traders can gain meaningful informational presence even without matching the incumbent's volume. Three important caveats apply: (i) this finding rests on a 32-day window during the campaign's most informationally intense period, when all markets should have been most efficient; (ii) block bootstrap confidence intervals (Appendix A) reveal wide 90% bounds for PolyMarket [1.3%, 64.8%] and Kalshi [0.8%, 41.7%], reflecting severe parameter uncertainty over such a short estimation window; and (iii) the short sample window makes the estimates sensitive to idiosyncratic events during October 2024. The event-window analysis in Section 5.6 provides complementary evidence that Kalshi's information share built up gradually over its first month.

### 5.4 Information Leadership Share

To better distinguish *informational* from *speed* advantages, Table 6 reports the Putniņš (2013) Information Leadership Share (ILS = IS/CS) for markets with positive CS, separately for the full sample and the Kalshi window.

**Table 6:** Information Leadership Share (ILS = IS/CS)

| Market | Full Sample IS | Full Sample CS | Full Sample ILS | Kalshi Window IS | Kalshi Window CS | Kalshi Window ILS |
|---|---|---|---|---|---|---|
| PolyMarket | 0.429 | 1.061 | 0.40 | 0.438 | 0.298 | **1.47** |
| Betfair | 0.282 | 1.171 | 0.24 | 0.057 | 0.205 | 0.28 |
| BetOnline | 0.046 | 0.295 | 0.16 | — | — | — |
| Bovada | 0.028 | 0.006 | — | 0.018 | 0.133 | 0.13 |
| Pinnacle | — | — | — | 0.031 | 0.224 | 0.14 |
| Kalshi | — | — | — | 0.409 | 0.242 | **1.69** |

*Notes:* ILS = IS/CS; only defined for markets with CS > 0.01. ILS > 1 indicates an informational advantage (IS exceeds what error correction alone predicts); ILS < 1 indicates a speed advantage. Markets with negative or near-zero CS are omitted. CS values sum to 1 by construction; individual values can exceed 1 when other markets have large negative CS (Unibet −0.84, Everygame −0.69, William Hill −0.01 in the full sample), indicating those venues systematically overshoot relative to the efficient price. Full-sample column uses the 7-market VECM (including Everygame; see Appendix Table A.1). Kalshi window: 9-market VECM.

The ILS reveals a sharp shift between the full sample and the Kalshi window. In the full sample, *all* markets with positive CS exhibit speed advantages (ILS $<$ 1): PolyMarket (0.40), Betfair (0.24), and BetOnline (0.16). Over the full 10 months, these markets' high CS values reflect rapid adjustment to equilibrium rather than the injection of new information. That even exchange-based venues exhibit speed rather than informational advantages over the full sample is consistent with the low information intensity during the early campaign months (January–June), when the election outcome was less salient and informed trading was sparse.

In the Kalshi window, PolyMarket (ILS = 1.47) and Kalshi (ILS = 1.69) flip to *informational* advantages (ILS > 1), while all traditional venues remain below 0.30. During this final month of the campaign, when information intensity and trading volume were at their peak, the two exchange-based venues were not merely adjusting faster but were generating permanent price innovations that other venues subsequently tracked. Betfair, despite retaining a meaningful CS weight (0.205), exhibits only a speed advantage (ILS = 0.28), suggesting that it processes information originating at PolyMarket and Kalshi rather than independently discovering it during this period.

Two implications follow. First, the exchange-versus-sportsbook hierarchy is not merely a mechanical artifact of price-updating frequency. If it were, the ILS would be uniformly below 1 across all venues and all windows, with faster venues leading only because they adjust sooner. The ILS exceeding 1 for PolyMarket and Kalshi during the Kalshi window indicates that these venues attract informed order flow that generates permanent price innovations. Second, Kalshi achieved a higher ILS than PolyMarket (1.69 vs. 1.47) within weeks of launch. This is consistent with Proposition 3: the regulated exchange may have unlocked a pool of U.S.-based informed traders who were previously excluded from all other venues.

### 5.5 Time-Varying Information Shares

Figure 3 displays rolling Hasbrouck information shares (using the default variable ordering from the data panel, 7-day smoothed) across the common sample. The single-ordering IS is used for computational tractability in the rolling analysis; the full-sample results in Table 4 confirm that midpoints across all orderings yield the same ranking. Consistent with H3, the dynamics reveal three distinct regimes:

1. **May–July (Traditional dominance):** Betfair and William Hill each hold 30–45% of price discovery. PolyMarket contributes 5–10%. The traditional, established venues lead during the low-volatility early campaign period.
2. **July–August (Regime shift):** The assassination attempt and Biden withdrawal trigger a sharp reallocation. William Hill's share collapses from 40% to below 10%. The landscape becomes more fragmented.
3. **September–November (Prediction market dominance):** PolyMarket's share rises steadily to 30–45%, establishing clear leadership through Election Day. This coincides with PolyMarket's surge in trading volume and media coverage.

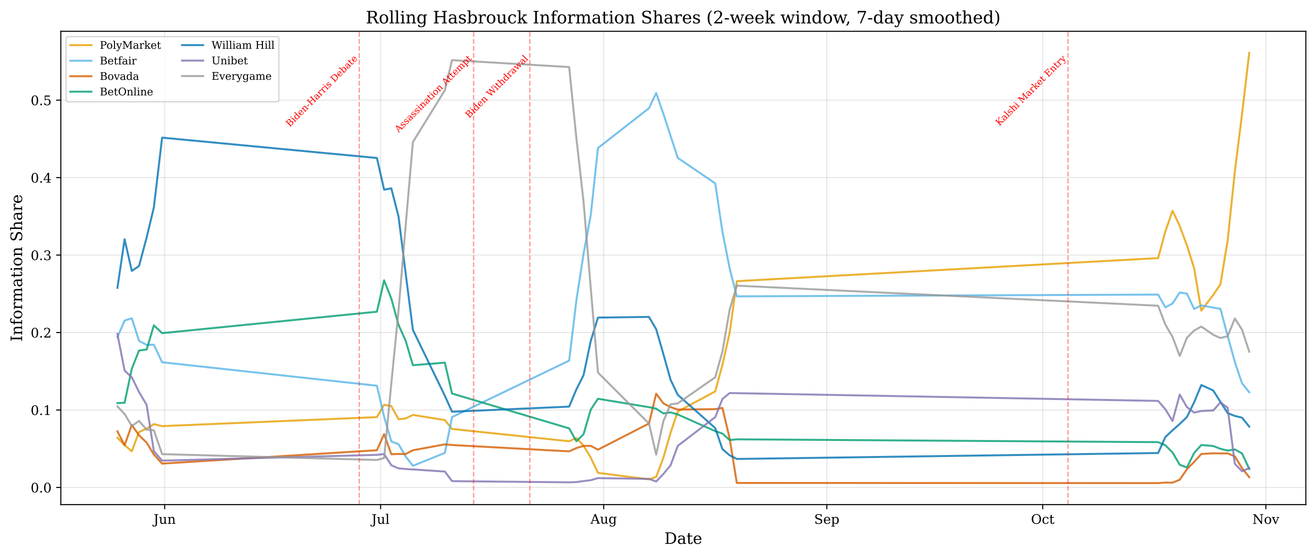


**Figure 3:** Rolling Hasbrouck information shares (2-week window, 1-day step, 7-day moving average). Vertical dashed lines indicate key political events. Note: rolling windows are backward-looking from each date.

Formal structural break tests confirm these visual impressions. Chow-type tests (Table 7) reject the null of parameter stability at all three event dates. The assassination attempt produces the largest F-statistic ($F = 6.59$), consistent with this exogenous weekend shock having the most disruptive effect on the cross-market price discovery structure.

## 5.6 Event-Window Analysis

Table 8 reports Hasbrouck midpoints from VECM estimation in windows around each key event.

The VECM is estimated on short windows (7 or 14 days) with limited degrees of freedom, so the event-window IS should be read as descriptive rather than precisely estimated. The debate window (June 27) shows traditional sportsbooks leading: William Hill (32%) and Betfair (24%) dominate, while PolyMarket contributes just 3%, consistent with the debate occurring during the period of traditional-platform dominance when PolyMarket's volume was still building. The assassination attempt (July 13), an exogenous weekend shock, shows

**Table 7:** Structural Break Tests

| Event | $n$ (markets) | $F$-stat | $p$-value | Verdict |
|---|---|---|---|---|
| Assassination Attempt (Jul 13) | 6 | 6.590 | $< 0.001$ | Reject |
| Biden Withdrawal (Jul 21) | 6 | 3.177 | $< 0.001$ | Reject |
| Kalshi Market Entry (Oct 4) | 7 | 4.105 | $< 0.001$ | Reject |

*Notes:* Chow-type F-tests comparing full-sample VECM residuals to separate pre/post-break VECM residuals. July events use a 6-market panel (excluding Everygame, which has only 4% coverage during July 11–22). The Kalshi entry test uses the full 7-market panel.

**Table 8:** Hasbrouck Information Share Midpoints by Event Window

| Event | PM | BF | BOV | BO | WH | UNI | EG |
|---|---|---|---|---|---|---|---|
| Biden–Harris Debate | 0.031 | 0.239 | 0.142 | 0.200 | **0.315** | 0.162 | 0.177 |
| Assassination Attempt | 0.237 | 0.030 | 0.129 | 0.085 | 0.029 | 0.060 | $0.513^{\dagger}$ |
| Biden Withdrawal | **0.371** | **0.406** | 0.163 | 0.019 | 0.045 | 0.196 | 0.055 |
| Kalshi Entry$^{\ddagger}$ | 0.175 | 0.114 | **0.394** | 0.002 | 0.029 | 0.006 | 0.091 |
| Election Day | 0.145 | 0.203 | 0.044 | 0.021 | 0.207 | 0.052 | $0.330^{\dagger}$ |

*Notes:* PM = PolyMarket, BF = Betfair, BOV = Bovada, BO = BetOnline, WH = William Hill, UNI = Unibet, EG = Everygame. Bold indicates the leading market(s) in each window. Window widths vary: 7 days (debate, election), 14 days (assassination, withdrawal, Kalshi entry). Everygame's high IS in these windows likely reflects the estimation artifact discussed in Section 5.2. Kalshi entry window uses the 9-market panel; Kalshi IS = 0.190, Pinnacle IS = 0.020 (not shown in columns).

PolyMarket surging to 24%. The shooting occurred on a Saturday evening; prediction markets (which trade 24/7) adjusted immediately, while sportsbook lines remained stale until oddsmakers manually updated.[4] The Biden withdrawal (July 21) shows PolyMarket (37%) and Betfair (41%) jointly accounting for 78% of price discovery during the campaign's most consequential single event.

The Kalshi Market Entry event window (centered on October 4) shows a gradual build-up. In the narrow window around its launch, Kalshi's IS is just 19%—lower than Bovada (39%) and comparable to PolyMarket (18%). Yet the full 32-day Kalshi-window VECM (Table 5) shows Kalshi at 41%. This suggests that Kalshi's information share *built up* over the course of October, rather than appearing instantaneously. The gradual ramp-up is consistent with informed traders migrating to the new platform as it established liquidity and demonstrated reliable execution.

### 5.7 Architecture-Grouped IS Decomposition

To test the headline claim directly with a low-dimensional system, we collapse the seven markets into architecture groups: DEX (PolyMarket), Exchange (Betfair), and Sportsbook (average of five sportsbooks). Next, we estimate a 3-market VECM(5, $r = 2$). With only three series, the system is parsimonious and the IS estimates are transparent. The DEX group captures 64.2% of information shares [52.2%, 76.2%], the Exchange group 33.0% [21.0%, 45.0%], and the Sportsbook composite just 5.1% [0.1%, 10.2%]. Gonzalo–Granger shares confirm

[4] Everygame's 51% IS in this window is the staleness artifact discussed in Section 5.2: it updated only twice during the window.

the pattern: the DEX and Exchange groups are information leaders (66.2% and 58.1%), while the Sportsbook composite is a systematic follower (−24.3%). Excluding Everygame from the Sportsbook composite raises it modestly to 13.3%, but the exchange dominance is unchanged (DEX 62.6%, Exchange 33.0%).

A finer 4-group decomposition, separating Sharp sportsbooks (Bovada, BetOnline) from Retail (William Hill, Unibet, Everygame), reveals that the DEX still dominates (52.5%), followed by the Exchange (22.4%), with the Sharp and Retail composites receiving 17.3% and 16.2% respectively. This confirms that even pooling sportsbooks by information reputation does not dislodge the exchange hierarchy. The architecture-grouped results demonstrate that the exchange-over-sportsbook finding is not an artifact of the 7-market system's dimensionality: it emerges cleanly from the simplest possible specification.

### 5.8 Frequency Sensitivity of Information Shares

A natural concern is whether PolyMarket's high-frequency dominance reflects informational leadership or merely its faster price-updating mechanism. Table 2 documents that update frequency varies by more than an order of magnitude across platforms, from 5.4% (Betfair) to 0.1% (Everygame). When a market carries forward a stale price, the VECM residual for that market is mechanically close to zero, which can inflate its estimated contribution to the permanent component—a well-known concern in the Hasbrouck IS literature (Baillie et al., 2002; De Jong, 2002).

If price staleness were the primary driver of information shares, then re-estimating at lower frequencies, where more markets have time to update, should compress the IS distribution and erode PolyMarket's lead. Table 9 reports the results of this test, which is a direct evaluation of H4.

**Table 9:** Hasbrouck IS Midpoints by Sampling Frequency

| Market | 5-min | 15-min | 30-min | 1-hour |
|---|---|---|---|---|
| PolyMarket | 0.429 | 0.278 | 0.192 | 0.119 |
| Betfair | 0.282 | 0.418 | 0.476 | 0.432 |
| Bovada | 0.028 | 0.054 | 0.046 | 0.061 |
| BetOnline | 0.046 | 0.041 | 0.043 | 0.041 |
| William Hill | 0.032 | 0.039 | 0.038 | 0.051 |
| Unibet | 0.101 | 0.117 | 0.108 | 0.086 |
| Everygame | 0.208 | 0.175 | 0.230 | 0.376 |

*Notes:* Each column re-estimates the 7-market VECM at the indicated sampling frequency. Lag orders adjusted to maintain similar time horizons. Hasbrouck IS midpoints are averages of upper and lower bounds across Cholesky orderings; for a given ordering, shares sum to one, but midpoints (which average across orderings) need not, and in these panels exceed one by 12–16% as a consequence. The ranking across venues, rather than the absolute magnitude, is the informative object.

PolyMarket's IS declines monotonically from 43% at 5-minute to 12% at 1-hour frequency. Betfair's IS *increases* from 28% to 43%. This crossover matches the prediction of Proposition 2: at high frequencies ($\Delta < \bar{\tau}_E$), only the zero-latency AMM can adjust, so it accounts for nearly all efficient-price innovations. At lower frequencies, the deeper LOB accumulates larger informed order flow (because $\kappa_E < \kappa_A$), and its cumulative contribution to innovation variance dominates. The two venues are not simply ranked; they exhibit *complementary* price discovery across horizons. PolyMarket leads at the highest frequencies, consistent with its AMM architecture providing instantaneous price response; while Betfair leads at medium horizons, reflecting the deeper liquidity and perhaps the greater deliberation of its more experienced, higher-stakes participant base.

Sportsbook IS remains low at all frequencies for the sharp books and William Hill (3–6%), confirming that their subordinate role is not merely a staleness artifact but reflects a lack of informational contribution. This is consistent with Proposition 4: bookmakers track the efficient price set elsewhere and contribute little innovation variance at any horizon. The exception is Everygame, whose IS *rises* from 21% at 5-minute to 38% at 1-hour frequency—the opposite of what the staleness explanation predicts. This anomaly likely reflects Everygame's large, discrete catch-up jumps, which at lower frequencies align with the estimation window and are mechanically amplified in the innovation variance. This pattern reinforces the decision to exclude Everygame from the clean panel.

### 5.9 Microstructure Mechanisms

*Cross-market spread compression after Kalshi.* At each 5-minute interval, we compute the spread between the highest and lowest implied probability across all active platforms. The mean cross-market spread fell from 7.2 pp pre-Kalshi to 4.2 pp post-Kalshi (October 4–November 5), a 42% reduction, and the 95th percentile compressed from 10.4 pp to 6.1 pp. To condition out confounds, we regress the daily max-min spread on a Kalshi entry dummy, realized volatility, the absolute distance from $p = 0.5$ (which mechanically compresses spreads), day-of-week and event fixed effects, using Newey–West HAC standard errors. The Kalshi dummy coefficient is −3.2 pp ($t = -7.78$, $p < 0.001$), indicating that cross-market spreads compressed significantly even after controlling for volatility and the approach of Election Day. Realized volatility enters positively ($t = 1.93$, $p = 0.054$), consistent with information-driven widening. Note, however, that the regression identifies conditional correlation rather than causation, as Kalshi's entry is confounded with election proximity.

*Cross-correlogram evidence.* Cross-correlograms between PolyMarket and the six other common-panel markets at 5-minute frequency (Figure 4) are flat, with all pairwise correlations between 0.87 and 0.95 across lags. Contemporaneous correlations of 5-minute probability *changes* (Figure 5) are near zero (0.002–0.051), despite the very high correlation in levels. This contrast confirms that price discovery operates through the error correction channel rather than through simultaneous innovation, motivating the VECM approach.

### 5.10 Ordering Invariance and Weak Exogeneity

The ordering-invariant Lien and Shrestha (2009) Modified Information Shares (MIS) closely track the Hasbrouck midpoints, confirming that the IS ranking is not driven by Cholesky ordering. In the Kalshi window, MIS similarly confirms PolyMarket's leadership (47.4%) and Kalshi's nontrivial share (38.1%). Using MIS in the Putniņš (2013) ILS yields consistent conclusions: exchange venues exhibit informational advantage ($\text{ILS}_{\text{MIS}} > 1$) while sportsbooks show speed advantage ($\text{ILS}_{\text{MIS}} < 1$). Weak exogeneity Wald tests provide a complementary perspective: all seven markets reject the null $H_0\colon \alpha_{j,\cdot} = 0$ at the 1% level, meaning no venue is a pure price leader in the formal sense. However, the test statistics differ by an order of magnitude—PolyMarket ($\chi^2 = 28.0$) and Betfair ($\chi^2 = 36.2$) versus BetOnline ($\chi^2 = 570$) and William Hill ($\chi^2 = 620$)—indicating far stronger statistical evidence of error correction for sportsbooks, consistent with their role as followers.

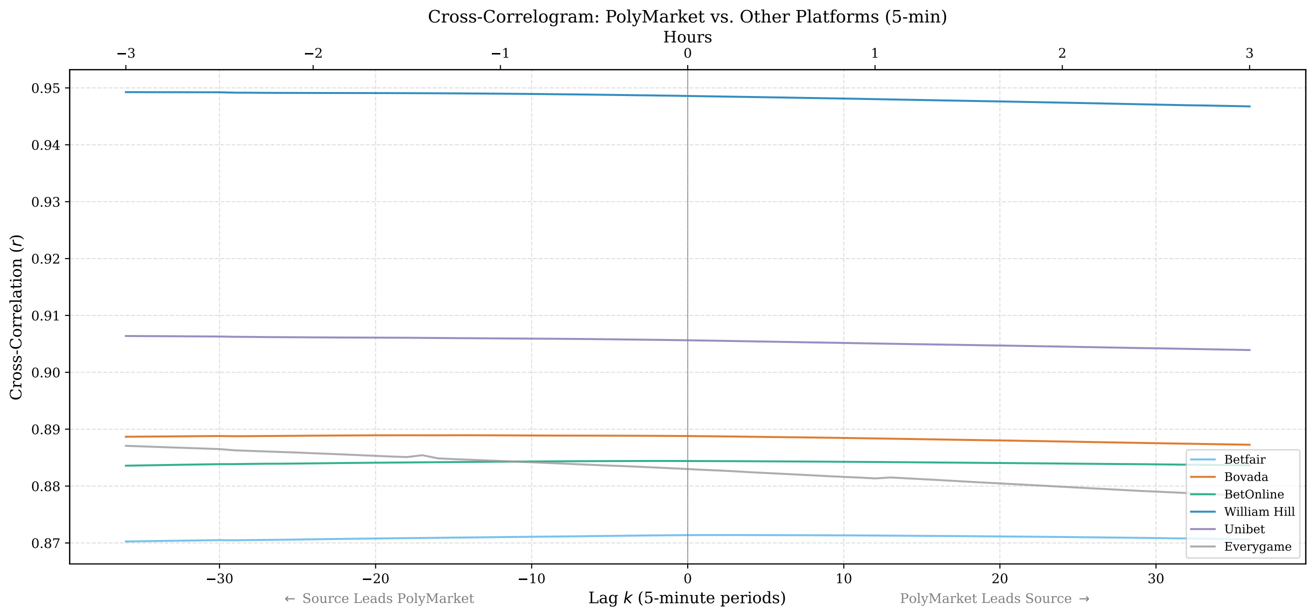


**Figure 4:** Cross-correlogram between PolyMarket and six other platforms at 5-minute frequency. Positive lags indicate PolyMarket leads.

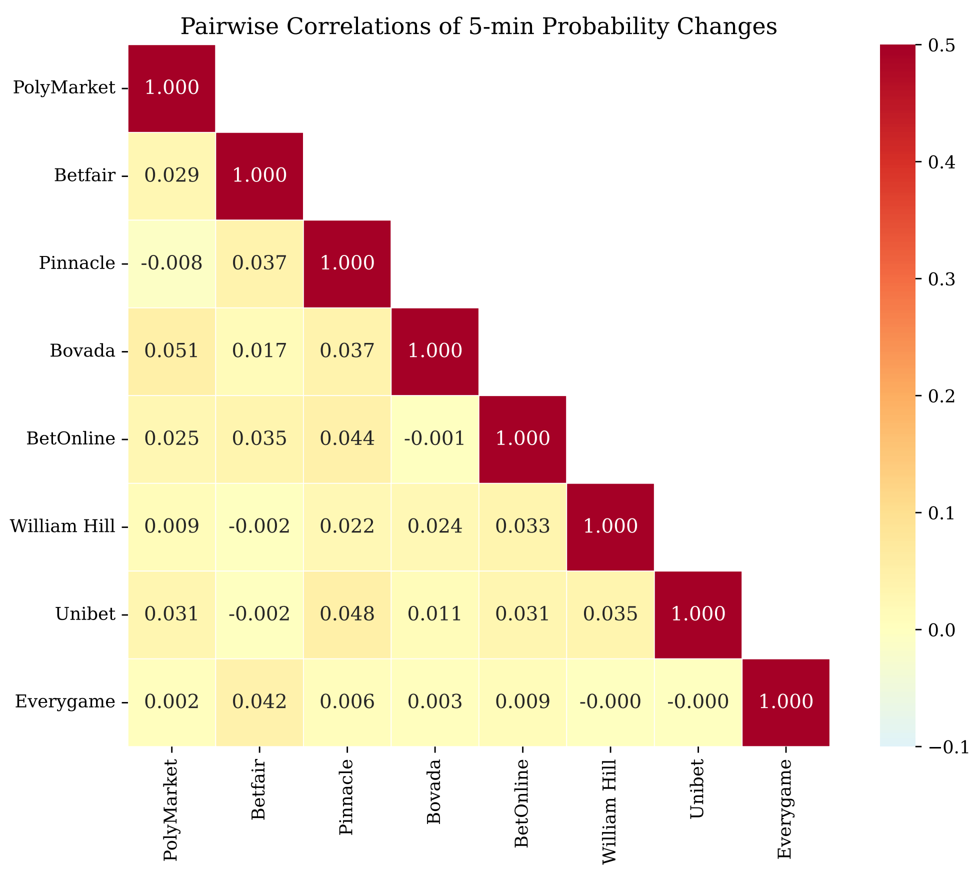


**Figure 5:** Pairwise correlations of 5-minute probability changes across all eight markets in the full panel (including Pinnacle from July onward; Kalshi excluded due to its shorter sample).

### 5.11 Connectedness Network

#### 5.11.1 Full-Sample Connectedness

The Diebold and Yılmaz (2014) connectedness analysis reveals a total connectedness index (TCI) of 55.9% for the common panel, indicating that over half of each market's forecast error variance is attributable to shocks from other markets. The directional spillover analysis identifies Betfair as the strongest net transmitter (NET = +0.18), followed by Unibet (+0.16) and William Hill (+0.13). PolyMarket is a net *receiver* (NET = −0.21) in the full-sample connectedness, which at first appears to contradict its high IS. This is consistent with the frequency-dependent leadership hypothesis: the time-domain GFEVD aggregates across all horizons, and at the 1-hour forecast horizon used here, Betfair's lower-frequency leadership dominates.

#### 5.11.2 Frequency Decomposition

The Baruník and Křehlík (2018) frequency decomposition directly resolves this apparent tension. The TCI decomposes into 2.5% at short-run (<30 minutes), 8.2% at medium-run (30 minutes to 2 hours), and 45.2% at long-run (>2 hours). Betfair's net transmitter status increases monotonically across bands, confirming its dominance at lower frequencies. In the Kalshi window, both PolyMarket and Kalshi are net transmitters at all frequency bands, while Betfair's net transmission increases with the horizon—a within-model confirmation of H4 that does not require re-estimation at different sampling frequencies.

#### 5.11.3 Rolling Connectedness

The rolling TCI ranges from 4.1% to 75.2%, with the lowest values in the early sample (May–June 2024, before major political events) and peaks around the Biden withdrawal and the election itself.

### 5.12 Jump Detection

The Lee and Mykland (2008) test identifies 761 statistically significant jumps across the seven common-panel markets. PolyMarket (140 jumps) and Betfair (556 jumps) account for 91% of all detected jumps, consistent with their role as the primary venues where new information is incorporated. Cross-market jump clustering analysis reveals 45 episodes in which multiple markets jump within a 30-minute window. PolyMarket leads 49% of these clusters, Betfair 33%, and BetOnline 18%. No retail sportsbook ever leads a multi-market jump cluster; BetOnline, a sharp book, is the only non-exchange venue to do so, providing independent microstructure evidence for the exchange dominance documented through the VECM framework.

### 5.13 Endogenous Structural Breaks

To validate the event-window analysis, we apply the Bai and Perron (1998) endogenous break detection algorithm to the rolling IS time series without specifying break dates *ex ante*. The algorithm identifies structural breaks within 3 days of the assassination attempt (July 13) and within 8 days of Biden's withdrawal (July 21) across multiple markets' IS series. Of the 29 detected breaks across all markets, 15 (52%) fall within 14 days of a known political event. The overlap between data-driven and event-driven break dates supports the interpretation that political shocks drive regime changes in the information structure.

### 5.14 Model Confidence Set

Following Hansen et al. (2011), we construct Model Confidence Sets (MCS) to formally identify the group of markets that cannot be statistically distinguished as the "best" price discoverers. Using the max-$t$ statistic with 5,000 bootstrap replications, the 90% MCS contains *only* PolyMarket—all other venues, including Betfair, are sequentially eliminated with $p < 0.02$. The result is equally stark for the Kalshi-inclusive panel, where PolyMarket again forms a singleton MCS. This means that while Kalshi's point-estimate IS is substantial, the MCS formally rejects the hypothesis that Kalshi's true IS is comparable to PolyMarket's. The Kalshi results should therefore be interpreted as showing that a newly launched exchange can attain nontrivial IS—not that it achieved statistical parity with the established leader.

## 6 Robustness

We examine the sensitivity of our findings across several dimensions. Detailed results are provided in the Appendix.

### 6.1 Sensitivity to Everygame Inclusion

Appendix Table A.1 reports the 7-market VECM including Everygame. Its artifactual IS of 21% is absorbed almost entirely by the two exchange venues: in the 7-market specification, PolyMarket's IS falls from 47.5% to 42.9% and Betfair's from 37.1% to 28.2%. The exchange-over-sportsbook hierarchy is identical in both specifications; the clean panel simply provides a sharper estimate by removing the contaminated series.

### 6.2 Specification Sensitivity

We verify that the IS ranking is robust to specification choices. Varying the lag order across $k \in \{13, 15, 17, 19, 21\}$ (BIC $\pm$ 4) yields PolyMarket IS midpoints in the range 38.0–42.9%, with the venue ranking unchanged. Reducing the cointegrating rank from $r = 5$ to $r = 4$ preserves the exchange-over-sportsbook hierarchy. The BIC strongly favors $r = 1$ over $r = 5$, reflecting its heavy penalty for parameter-rich specifications; since the economics dictate a single common stochastic trend, we maintain $r = n - 1$ but note that the IS ranking is insensitive to this choice. Replacing the logit transformation with raw probabilities produces near-identical results (PolyMarket IS: 41.6% vs. 42.9% in the 7-market specification). Excluding overnight periods (8 PM to 8 AM ET) slightly increases PolyMarket's IS, consistent with its 24/7 architecture. Pairwise Granger causality tests (Appendix Table A.8) confirm that Betfair Granger-causes all other markets ($p < 0.01$), with all 36 significant pairs surviving Holm-Bonferroni and Benjamini-Hochberg corrections at the 5% level.

### 6.3 Bootstrap Robustness

The stationary bootstrap with BCa correction yields confidence intervals that are generally comparable to the fixed-block bootstrap, while providing two methodological improvements: (i) the optimal block length (22 periods ≈ 110 minutes) is data-driven rather than arbitrarily fixed at one day, and (ii) the BCa correction accounts for the skewness inherent in bounded $[0, 1]$ IS estimates. The recursive-design wild bootstrap, which preserves the conditional heteroskedasticity structure of the original VECM residuals, produces similar confidence intervals, confirming that the GARCH effects present around political events do not materially distort the standard bootstrap inference.

### 6.4 DCC-GARCH Time-Varying Information Shares

As an alternative to rolling-window VECMs with their arbitrary window-size choice, we fit a DCC-GARCH(1,1) model (Engle, 2002) to the VECM residuals and compute period-by-period IS using the time-varying covariance matrix $\mathbf{\Sigma}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$. The DCC parameters ($\hat{a} = 0.005$, $\hat{b} = 0.93$) indicate highly persistent but slowly evolving correlations. PolyMarket's time-varying IS averages 31.6% with a 5th–95th percentile range of [20.4%, 71.2%], confirming the qualitative pattern from the rolling analysis while revealing finer-grained dynamics.

### 6.5 System Sensitivity: Leave-One-Out Analysis

To assess whether individual series drive the IS ranking, we re-estimate the VECM after dropping each of the seven markets in turn, computing 6-market IS for the remaining series. PolyMarket's IS midpoint averages 41.2% across the six configurations in which it appears, ranging from 7.8% (when William Hill is dropped) to 59.2% (when Betfair is dropped). Betfair averages 47.5% across its six configurations. The exchange-over-sportsbook hierarchy is preserved in all seven leave-one-out specifications: the top two IS markets are always drawn from {PolyMarket, Betfair}. The outlier (PolyMarket's low IS when William Hill is excluded) reflects the disruption to the error correction structure when a key slow-adjusting market is removed, not a decline in PolyMarket's informational role.

### 6.6 Staleness and Freshness Robustness

The Everygame staleness artifact (Section 5.2) raises the question of whether forward-fill staleness inflates IS for other low-frequency markets. We address this with a fresh-quote filter: we identify 5-minute timestamps where the two most active markets (PolyMarket and Betfair) have each shown a non-zero return within the preceding 24 periods (2 hours). This retains 3,410 observations (20.7% of the sample)—a meaningful subset that excludes stale-price periods. On this fresh-quote subsample, PolyMarket's IS is 40.1% [27.7%, 52.5%], consistent with its full-sample estimate (42.9%), confirming that exchange dominance is not an artifact of staleness.

As a complementary exercise, we simulate 1,000 synthetic cointegrated systems (common random walk + i.i.d. venue noise with equal true IS) and apply staleness at the empirically-calibrated update rates from the data. The simulation confirms that staleness systematically redistributes IS away from actively-updating markets and toward stale markets: Betfair receives the largest positive reallocation (+48.2 pp), followed by Everygame (+15.4 pp), the most stale venue. This directional pattern matches the empirical anomaly (Everygame's inflated IS) and reinforces the interpretation that the clean-panel and fresh-quote results provide the more reliable IS estimates.

### 6.7 Kalshi Window Sensitivity

The 32-day Kalshi window during peak election intensity warrants additional scrutiny. We draw 200 random 32-day windows from the pre-Kalshi period (January–September 2024) and estimate 7-market VECM(3, $r = 6$) for each, building a placebo distribution of IS midpoints. PolyMarket's Kalshi-window IS in the 7-market specification (70.4%) falls at the 100th percentile of the placebo distribution (placebo mean: 27.7%, SD: 10.8%), far above the 95th percentile (44.3%). Sportsbooks' Kalshi-window IS values fall at or near the 0th percentile, suggesting the concentration of price discovery in exchanges is unusually high during this period even by the standards of similarly-sized windows. We also extend the frequency-sensitivity analysis to the 9-market Kalshi panel: Kalshi's IS persists across frequencies (40.9% at 5-minute, 31.2% at 15-minute, 33.9% at 30-minute, 25.8% at 1-hour), declining at lower frequencies but remaining among the top two venues at all horizons.

### 6.8 IS–Volatility Interaction

To test H3 formally—whether exchange IS increases disproportionately during high-information periods—we regress rolling IS on standardized realized volatility interacted with architecture dummies (DEX, Exchange, Sharp; Retail omitted), with market fixed effects and cluster-robust standard errors. None of the interaction coefficients are statistically significant (Vol×DEX: $t = -0.07$, $p = 0.95$; Vol×Exchange: $t = 0.30$, $p = 0.76$; Vol×Sharp: $t = 0.54$, $p = 0.59$). Market-by-market Newey–West regressions confirm the null: no venue shows a significant positive relationship between IS and volatility. This null result is informative: the exchange dominance in price discovery is not concentrated in high-volatility episodes but is instead a persistent structural feature of the market architecture.

## 7 Conclusion

This paper brings the standard price discovery toolkit to political prediction markets for the first time, applying it to nine venues during the 2024 U.S. presidential election. The results point to three broad conclusions with implications for market design and regulation.

First, exchange-based architectures dominate price formation. PolyMarket and Betfair jointly account for the vast majority of information shares; all sportsbooks are excluded from the Hansen et al. (2011) Model Confidence Set. Weak exogeneity tests, connectedness analysis, and jump detection all reinforce this hierarchy, which is robust to lag choice, cointegrating rank, transformation, sample period, and bootstrap method. For regulators weighing whether prediction markets serve a legitimate economic function, this finding suggests that exchange-organized venues contribute substantively to price discovery in a way that traditional bookmakers do not.

Second, information leadership is frequency-dependent. PolyMarket's AMM leads at high frequencies while Betfair's LOB leads at lower frequencies, a crossover confirmed both by re-estimation at different sampling frequencies and by the Baruník and Křehlík (2018) spectral decomposition within a single model. This complementarity implies that different exchange architectures serve distinct informational roles. Market designers may thus benefit from fostering architectural diversity rather than converging on a single trading protocol.

Third, Kalshi's rapid attainment of nontrivial information shares following its October 2024 launch illustrates that a newly opened exchange can quickly gain informational presence. The Model Confidence Set rejects that Kalshi achieved statistical parity with PolyMarket, and the 32-day window during peak election salience limits causal inference. Nevertheless, the episode is consistent with the hypothesis that exchange architecture, rather than incumbency or accumulated liquidity, facilitates informed trading.

Several limitations constrain the conclusions that can be drawn. The absence of contract-level volume data precludes volume-weighted IS or direct tests of liquidity channels. Our analysis is descriptive, and we cannot rule out confounding factors such as differences in trader composition or liquidity. Nevertheless, four independent empirical patterns collectively point to architecture rather than liquidity or user composition as the primary driver of the IS hierarchy. First, the frequency-dependent crossover between PolyMarket and Betfair (Section 5.8) is inconsistent with a pure liquidity or information-advantage explanation: if volume or user sophistication alone determined IS, the same venue would dominate at all horizons, whereas the observed crossover maps directly to the architectural trade-off between AMM latency and LOB depth. Second, Kalshi attained a 41% IS within 32 days despite launching with a fraction of PolyMarket's volume, suggesting that exchange architecture facilitates informed trading even before deep liquidity has accumulated. Third, the IS–volatility interaction is null (Section 6): exchange dominance is a persistent structural feature, not one that emerges only during high-information episodes when liquidity differences might be most binding. Fourth, the architecture-grouped VECM (Section 5.7) confirms the hierarchy in the simplest possible specification, ruling out artifacts of high-dimensional estimation. While none of these tests constitutes a clean causal identification, their convergence substantially narrows the set of plausible alternative explanations. Future research with on-chain transaction data from PolyMarket and order-book data from Kalshi and Betfair could close the volume gap and enable more direct tests of the liquidity channel. The Kalshi launch also constitutes a natural experiment that, with appropriate controls, could support causal identification of the effect of adding a regulated exchange on system-wide pricing efficiency.

## References


Aktuğ, E. and Torul, O. (2026). Political capital: Real-time electoral probabilities and the value of corporate political alignment. SSRN Working Paper No: 5234903.

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., Neumann, G. R., Ottaviani, M., Schelling, T. C., Shiller, R. J., Smith, V. L., Snowberg, E., Sunstein, C. R., Tetlock, P. C., Tetlock, P. E., Varian, H. R., Wolfers, J., and Zitzewitz, E. (2008). The promise of prediction markets. *Science*, 320(5878):877–878.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Baillie, R. T., Booth, G. G., Tse, Y., and Zabotina, T. (2002). Price discovery and common factor models. *Journal of Financial Markets*, 5(3):309–321.

Baruník, J. and Křehlík, T. (2018). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics*, 16(2):271–296.

Berg, J. E., Forsythe, R., Nelson, F. D., and Rietz, T. A. (2008). Results from a dozen years of election futures markets research. In *Handbook of Experimental Economics Results*, volume 1, pages 742–751. Elsevier.

Cavaliere, G. and Taylor, A. M. R. (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory*, 24(1):43–71.

Chakravarty, S., Gulen, H., and Mayhew, S. (2004). Informed trading in stock and option markets. *The Journal of Finance*, 59(3):1235–1257.

Croxson, K. and Reade, J. J. (2014). Information and efficiency: Goal arrival in soccer betting. *Journal of Financial Economics*, 117(2):295–312.

De Jong, F. (2002). Measures of contributions to price discovery: A comparison. *Journal of Financial Markets*, 5(3):323–327.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Diebold, F. X. and Yılmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Business & Economic Statistics*, 32(1):128–146.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Flynn, M. and Tarkom, A. (2025). How do financial markets price political uncertainty? evidence from the 2024 united states presidential election. *Finance Research Letters*, 75:106879.

Foucault, T. and Menkveld, A. J. (2008). Competition for order flow and smart order routing systems. *The Journal of Finance*, 63(1):119–158.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Business & Economic Statistics*, 22(1):89–101.

Gonzalo, J. and Granger, C. W. J. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13(1):27–35.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Hanson, R., Oprea, R., and Porter, D. (2006). Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization*, 60(4):449–461.

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

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563.

Lehar, A. and Parlour, C. A. (2021). Decentralized exchanges. *The Review of Financial Studies*, 34(11):5275–5313.

Lehmann, B. N. (2002). Some desiderata for the measurement of price discovery across markets. *Journal of Financial Markets*, 5(3):259–276.

Lien, D. and Shrestha, K. (2009). A new information share measure. *Journal of Futures Markets*, 29(4):377–395.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258.

Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *The Journal of Finance*, 75(2):585–617.

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

O'Hara, M. and Ye, M. (2011). Is market fragmentation harming market quality? *The Review of Financial Studies*, 24(11):3437–3483.

Patton, A., Politis, D. N., and White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap". *Econometric Reviews*, 28(4):372–375.

Pesaran, M. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17–29.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70.

Putniņš, T. J. (2013). What do measures of price discovery tell us? an empirical investigation. *Journal of Financial Markets*, 16(4):619–648.

Rothschild, D. (2009). Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly*, 73(5):895–916.

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

Snowberg, E., Wolfers, J., and Zitzewitz, E. (2007). Partisan impacts on the economy: evidence from prediction markets and close elections. *The Quarterly Journal of Economics*, 122(2):807–829.

Wolfers, J. and Zitzewitz, E. (2008). Prediction markets in theory and practice. In Blume, L. E. and Durlauf, S. N., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London, 2nd edition.

# Appendices

## A Robustness Tables

**Table A.1:** 7-Market VECM Including Everygame (for Comparison)

| | Hasbrouck IS | | | Bootstrap 90% CI | | | |
|---|---|---|---|---|---|---|---|
| Market | Lower | Mid | Upper | $5^{th}$ | $95^{th}$ | CS | Arch. |
| PolyMarket | 0.289 | 0.429 | 0.569 | 0.018 | 0.844 | 1.061 | DEX |
| Betfair | 0.125 | 0.282 | 0.439 | 0.000 | 0.303 | 1.171 | Exchange |
| Bovada | 0.000 | 0.028 | 0.056 | 0.000 | 0.295 | 0.006 | Sharp |
| BetOnline | 0.000 | 0.046 | 0.093 | 0.001 | 0.398 | 0.295 | Sharp |
| William Hill | 0.000 | 0.032 | 0.063 | 0.000 | 0.229 | −0.007 | Retail |
| Unibet | 0.000 | 0.101 | 0.201 | 0.000 | 0.169 | −0.841 | Retail |
| Everygame | 0.100 | 0.208 | 0.315 | 0.001 | 0.435 | −0.686 | Retail |

*Notes:* 7-market VECM including Everygame. Everygame's 21% IS is an artifact of staleness (0.1% update rate); its residuals exhibit serial correlation (Ljung-Box $p < 0.01$). The clean 6-market panel (Table 4) is the primary specification. Bootstrap CIs from 500 block bootstrap replications. $k = 17$ lags, $r = 6$.

**Table A.2:** Block Bootstrap Confidence Intervals: Kalshi Window

| | Hasbrouck IS (Bootstrap Distribution) | | |
|---|---|---|---|
| Market | Mean | 5% Lower | 95% Upper |
| PolyMarket | 0.259 | 0.013 | 0.648 |
| Kalshi | 0.157 | 0.008 | 0.417 |
| Betfair | 0.112 | 0.001 | 0.411 |
| Everygame | 0.103 | 0.000 | 0.420 |
| Unibet | 0.085 | 0.000 | 0.422 |
| Pinnacle | 0.075 | 0.000 | 0.367 |
| Bovada | 0.074 | 0.000 | 0.361 |
| William Hill | 0.072 | 0.000 | 0.274 |
| BetOnline | 0.063 | 0.000 | 0.259 |

*Notes:* Results from 500 block bootstrap replications (1-day block size) on the 9-market panel during the 32-day Kalshi window. The wide and overlapping 90% confidence intervals for PolyMarket and Kalshi reflect the short estimation window, confirming that the precise degree of informational parity is sensitive to sampling variation.

**Table A.3:** Lag Sensitivity: Hasbrouck IS Midpoints and GG Component Shares

| | $k = 13$ | $k = 15$ | $k = 17$ (BIC) | $k = 19$ | $k = 21$ |
|---|---|---|---|---|---|
| *Hasbrouck IS Midpoint* | | | | | |
| PolyMarket | 0.401 | 0.380 | 0.429 | 0.427 | 0.428 |
| Betfair | 0.320 | 0.343 | 0.282 | 0.289 | 0.260 |
| Bovada | 0.026 | 0.027 | 0.028 | 0.026 | 0.033 |
| BetOnline | 0.064 | 0.052 | 0.046 | 0.064 | 0.081 |
| William Hill | 0.032 | 0.031 | 0.032 | 0.028 | 0.042 |
| Unibet | 0.102 | 0.100 | 0.101 | 0.090 | 0.094 |
| Everygame | 0.169 | 0.177 | 0.208 | 0.195 | 0.193 |
| *GG Component Share* | | | | | |
| PolyMarket | 0.842 | 0.842 | 1.061 | 0.931 | 0.947 |
| Betfair | 1.022 | 1.095 | 1.171 | 1.032 | 0.989 |
| Bovada | −0.057 | −0.057 | 0.006 | −0.030 | −0.083 |
| BetOnline | 0.369 | 0.274 | 0.295 | 0.404 | 0.555 |
| William Hill | −0.094 | −0.014 | −0.007 | −0.054 | −0.166 |
| Unibet | −0.727 | −0.767 | −0.841 | −0.683 | −0.652 |
| Everygame | −0.354 | −0.373 | −0.686 | −0.600 | −0.590 |

*Notes:* BIC selects $k = 17$ for the 7-market common panel. Cointegrating rank $r = 6$ in all specifications.

**Table A.4:** Cointegrating Rank Sensitivity: Hasbrouck IS Midpoints

| | $r = 6$ $(n-1)$ | $r = 5$ $(n-2)$ | $r = 4$ $(n-3)$ |
|---|---|---|---|
| PolyMarket | 0.429 | 0.382 | 0.611 |
| Betfair | 0.282 | 0.409 | 0.164 |
| Bovada | 0.028 | 0.050 | 0.038 |
| BetOnline | 0.046 | 0.037 | 0.030 |
| William Hill | 0.032 | 0.039 | 0.126 |
| Unibet | 0.101 | 0.077 | 0.086 |
| Everygame | 0.208 | 0.104 | 0.109 |

*Notes:* 7-market common panel with $k = 17$ lags. The theoretical rank is $n - 1 = 6$ (one common stochastic trend).

**Table A.5:** Logit vs. Raw Probability: Hasbrouck IS Midpoints

| | Common Panel | | Kalshi Window | |
|---|---|---|---|---|
| Market | Logit | Raw | Logit | Raw |
| PolyMarket | 0.429 | 0.416 | 0.438 | 0.424 |
| Betfair | 0.282 | 0.286 | 0.057 | 0.062 |
| Bovada | 0.028 | 0.030 | 0.018 | 0.019 |
| BetOnline | 0.046 | 0.056 | 0.002 | 0.001 |
| William Hill | 0.032 | 0.035 | 0.002 | 0.002 |
| Unibet | 0.101 | 0.111 | 0.000 | 0.000 |
| Everygame | 0.208 | 0.202 | 0.048 | 0.049 |
| Pinnacle | — | — | 0.031 | 0.032 |
| Kalshi | — | — | 0.409 | 0.417 |

*Notes:* "Logit" columns use the baseline logit transformation $\log(p/(1-p))$. "Raw" columns use untransformed implied probabilities. VECM specification otherwise identical.

**Table A.6:** Overnight Exclusion: Hasbrouck IS and GG Component Shares

| | Full 24-hour | | Excl. 8PM–8AM ET | |
|---|---|---|---|---|
| Market | IS Mid | CS | IS Mid | CS |
| PolyMarket | 0.429 | 1.061 | 0.439 | 1.504 |
| Betfair | 0.282 | 1.171 | 0.173 | 1.280 |
| Bovada | 0.028 | 0.006 | 0.051 | 0.480 |
| BetOnline | 0.046 | 0.295 | 0.055 | 0.482 |
| William Hill | 0.032 | −0.007 | 0.062 | −0.435 |
| Unibet | 0.101 | −0.841 | 0.109 | −1.134 |
| Everygame | 0.208 | −0.686 | 0.297 | −1.178 |

*Notes:* Excluding overnight hours (8PM–8AM Eastern) removes periods when U.S. sportsbooks are least active.

**Table A.7:** Sub-Sample Stability: Pre-July vs. Post-July

| | Jan–Jul 22 | | Jul 22–Nov | |
|---|---|---|---|---|
| Market | IS Mid | CS | IS Mid | CS |
| PolyMarket | 0.267 | 1.540 | 0.362 | 1.177 |
| Betfair | 0.116 | 0.275 | 0.268 | 1.083 |
| Bovada | 0.106 | −0.429 | 0.047 | 0.116 |
| BetOnline | 0.254 | −0.777 | 0.127 | 0.702 |
| William Hill | 0.266 | 1.106 | 0.069 | −0.264 |
| Unibet | 0.068 | −0.024 | 0.093 | −0.514 |
| Everygame | 0.157 | −0.691 | 0.265 | −1.298 |

*Notes:* Sample split at July 22, 2024 (after both the assassination attempt and Biden's withdrawal).

**Table A.8:** Pairwise Granger Causality $p$-Values (5-Minute First Differences)

| From ↓ / To → | PM | BF | BOV | BO | WH | UNI | EG |
|---|---|---|---|---|---|---|---|
| PM | — | 0.000 | 0.000 | 0.455 | 0.510 | 0.600 | 0.000 |
| BF | 0.000 | — | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| BOV | 0.000 | 0.000 | — | 0.000 | 0.000 | 0.000 | 0.000 |
| BO | 0.000 | 0.000 | 0.000 | — | 0.000 | 0.000 | 0.000 |
| WH | 0.000 | 0.000 | 0.000 | 0.720 | — | 0.000 | 0.000 |
| UNI | 0.000 | 0.000 | 0.000 | 0.005 | 0.354 | — | 0.000 |
| EG | 0.000 | 0.000 | 0.103 | 0.000 | 0.000 | 0.000 | — |

*Notes:* Entry $(i, j)$ is the $p$-value for the null hypothesis that market $i$ does not Granger-cause market $j$. Tests conducted on 5-minute first differences of logit-transformed probabilities.

**Table A.9:** Architecture-Grouped IS Decomposition

| Specification | Group | IS Lower | IS Mid | IS Upper | GG Share |
|---|---|---|---|---|---|
| 3-Group | DEX | 0.522 | **0.642** | 0.762 | 0.662 |
| | Exchange | 0.210 | 0.330 | 0.450 | 0.581 |
| | Sportsbook | 0.001 | 0.051 | 0.102 | −0.243 |
| 3-Group Clean | DEX | 0.479 | **0.626** | 0.773 | 0.541 |
| | Exchange | 0.185 | 0.330 | 0.475 | 0.451 |
| | Sportsbook | 0.000 | 0.133 | 0.266 | 0.009 |
| 4-Group | DEX | 0.369 | **0.525** | 0.681 | 0.677 |
| | Exchange | 0.081 | 0.224 | 0.368 | 0.530 |
| | Sharp | 0.003 | 0.173 | 0.343 | 0.574 |
| | Retail | 0.013 | 0.162 | 0.310 | −0.781 |

*Notes:* Markets averaged within architecture groups. 3-Group: DEX = PolyMarket, Exchange = Betfair, Sportsbook = 5 sportsbooks. 3-Group Clean excludes Everygame. 4-Group: Sharp = Bovada + BetOnline, Retail = William Hill + Unibet + Everygame. VECM(5) with $r = n - 1$.

**Table A.10:** Leave-One-Out IS Sensitivity: PolyMarket IS Midpoint When Each Market Is Dropped

| Dropped Market | PM IS Mid | PM IS Lower | PM IS Upper | BF IS Mid | $n$ |
|---|---|---|---|---|---|
| Betfair | 0.592 | 0.470 | 0.714 | — | 6 |
| Bovada | 0.423 | 0.282 | 0.565 | 0.302 | 6 |
| BetOnline | 0.423 | 0.310 | 0.536 | 0.246 | 6 |
| William Hill | 0.078 | 0.000 | 0.157 | 0.324 | 6 |
| Unibet | 0.405 | 0.285 | 0.525 | 0.316 | 6 |
| Everygame | 0.503 | 0.396 | 0.610 | 0.290 | 6 |

*Notes:* Each row drops the indicated market and re-estimates a 6-market VECM(17, $r = 5$). PolyMarket's IS averages 41.2% across configurations. The outlier when William Hill is dropped (7.8%) reflects disruption to the error correction structure.

**Table A.11:** Kalshi Placebo Windows: Actual vs. Pre-Kalshi Null Distribution

| | Placebo Distribution ($N = 200$) | | | | | |
|---|---|---|---|---|---|---|
| Market | Mean | SD | 5th pct | 95th pct | Actual | Pct. rank |
| PolyMarket | 0.277 | 0.108 | 0.098 | 0.443 | **0.704** | 100.0% |
| Betfair | 0.153 | 0.155 | 0.024 | 0.539 | 0.155 | 59.5% |
| Bovada | 0.108 | 0.045 | 0.054 | 0.192 | 0.017 | 0.0% |
| BetOnline | 0.132 | 0.068 | 0.024 | 0.203 | 0.061 | 24.5% |
| William Hill | 0.204 | 0.152 | 0.048 | 0.402 | 0.001 | 0.0% |
| Unibet | 0.106 | 0.043 | 0.049 | 0.165 | 0.006 | 0.0% |
| Everygame | 0.218 | 0.105 | 0.092 | 0.405 | 0.062 | 0.5% |

*Notes:* 200 random 32-day windows drawn from the pre-Kalshi period (Jan–Sep 2024). Each window estimated with 7-market VECM(3, $r = 6$). Percentile rank indicates the fraction of placebo IS values below the actual Kalshi-window IS.

**Table A.12:** Kalshi-Window Frequency Sensitivity

| Market | 5min | 15min | 30min | 1h |
|---|---|---|---|---|
| PolyMarket | 0.438 | 0.538 | 0.455 | 0.018 |
| Kalshi | 0.409 | 0.312 | 0.339 | 0.258 |
| Betfair | 0.057 | 0.052 | 0.119 | 0.216 |
| Pinnacle | 0.031 | 0.061 | 0.015 | 0.054 |
| Bovada | 0.018 | 0.001 | 0.003 | 0.008 |
| BetOnline | 0.002 | 0.015 | 0.040 | 0.052 |
| William Hill | 0.002 | 0.010 | 0.074 | 0.003 |
| Unibet | 0.000 | 0.003 | 0.010 | 0.030 |
| Everygame | 0.048 | 0.022 | 0.004 | 0.414 |

*Notes:* IS midpoints for the 9-market Kalshi panel at different sampling frequencies. Kalshi's IS persists across frequencies, declining from 40.9% (5-minute) to 25.8% (1-hour) but remaining among the top two venues at all horizons. At the 1-hour frequency, PolyMarket's IS drops to 1.8% while Everygame's rises to 41.4%, likely reflecting staleness-driven redistribution in this short, 32-day sample.

**Table A.13:** Cross-Market Spread Regression

| Variable | Coefficient | NW SE | $t$-stat | $p$-value |
|---|---|---|---|---|
| Constant | 0.066 | 0.006 | 10.40 | $< 0.001$ |
| Kalshi Dummy | −0.032 | 0.004 | −7.78 | $< 0.001$ |
| Realized Vol | 5.778 | 2.996 | 1.93 | 0.054 |
| $\lvert p - 0.5 \rvert$ | 0.060 | 0.063 | 0.95 | 0.342 |
| Monday | −0.011 | 0.005 | −2.25 | 0.025 |
| Election Day | −0.018 | 0.007 | −2.43 | 0.015 |

*Notes:* OLS regression of daily max-min cross-market probability spread on covariates. Newey–West HAC standard errors. $N = 310$ trading days. Kalshi dummy = 1 after October 4, 2024.

# B Conceptual Framework: Venue Choice and Price Discovery

This appendix presents a partial-equilibrium framework that rationalizes our main empirical findings and maps them to hypotheses H1–H4. The framework is deliberately simple—it features no market clearing, no strategic interaction, and no welfare analysis—and is intended to provide economic intuition for the empirical patterns, not to generate structural restrictions beyond those captured by our reduced-form hypotheses.

## B.1 Environment

Time is continuous on $[0, T]$. A single binary event $Y \in \{0, 1\}$ is realized at time $T$ with prior probability $p_0 = \mathbb{P}(Y = 1)$.

There is a continuum of risk-neutral traders. A measure $\mu_I$ are *informed* and receive private information about $Y$; the remaining traders generate uninformed (noise) order flow. Conditional on $Y$, each informed trader $i$ receives a private signal $s_i$ at random times according to a Poisson process with intensity $\lambda > 0$. Signals are independent across traders and satisfy the Monotone Likelihood Ratio Property so that each signal induces a posterior belief $p_s \in (0, 1)$ with $p_s \neq p_0$ almost surely.

There are three trading venues indexed by $v \in \{A, E, B\}$:

1. An automated market maker (AMM) prediction market $A$ (PolyMarket-type).
2. An order-driven exchange $E$ (Betfair/Kalshi-type).
3. A bookmaker $B$ (traditional sportsbook).

Each venue quotes a price $P_v(t) \in (0, 1)$ for an Arrow–Debreu-style contract paying 1 if $Y = 1$ and 0 otherwise. We interpret $P_v(t)$ as the venue's implied probability of $Y = 1$.

## B.2 Venue Technologies

Venues differ along three dimensions: latency, depth (price impact), and participation cost.

### B.2.1 Latency.

After a signal arrives, there is a minimal time $\tau_v \geq 0$ before venue $v$ can adjust its price:

$$\tau_A = 0, \qquad \tau_E = \bar{\tau}_E > 0, \qquad \tau_B = \bar{\tau}_B \gg \bar{\tau}_E.$$

Thus, the AMM can update its price instantaneously when it receives an order, the exchange has positive latency due to order submission and matching frictions, and the bookmaker reviews odds only infrequently.

### B.2.2 Depth and price impact.

Trading $q_v$ contracts at venue $v$ when the current price is $P_v$ moves the price by

$$\Delta P_v = \kappa_v q_v, \qquad \kappa_A > \kappa_E > \kappa_B \approx 0. \tag{6}$$

The AMM has high price impact (thin depth), the exchange has deeper books (lower impact), and the bookmaker absorbs order flow at (essentially) fixed odds except at review times.

#### B.2.3 Participation costs.

A trader who trades $q_v$ contracts at venue $v$ incurs a per-unit cost $c_v \geq 0$, capturing access frictions, regulatory constraints, and perceived counterparty risk. We allow $c_v$ to differ across trader types. In particular, a mass $\mu_R \leq \mu_I$ of "regulated" informed traders face $c_A = c_E^{\text{off}} = +\infty$ (they cannot use offshore or on-chain venues) but finite $c_E^{\text{reg}}$ at a regulated exchange $E$; the remaining informed traders can use all venues.

### B.3 Informed Trading at a Given Venue

Consider a trader with posterior belief $p_s$ who arrives at time $t$ when venue $v$ quotes price $P_v(t)$. She chooses $q_v$ to maximize expected profit net of price impact and participation cost:

$$\Pi_v(q_v) = q_v(p_s - P_v(t)) - \frac{1}{2}\kappa_v q_v^2 - c_v q_v. \tag{7}$$

The first term is expected payoff from the contract, the second term is the cost induced by price impact (6), and the third term is the participation cost.

**Proposition 1** (Optimal order size and venue comparison)**.** *For given* $(p_s, P_v, c_v, \kappa_v)$*, the unique optimal order at venue* $v$ *is*

$$q_v^* = \frac{p_s - P_v - c_v}{\kappa_v} \quad \textit{if } p_s - P_v > c_v, \qquad q_v^* = 0 \textit{ otherwise.}$$

*The associated maximal profit is*

$$\Pi_v^* = \frac{(p_s - P_v - c_v)^2}{2\kappa_v} \cdot \mathbb{1}\{p_s - P_v > c_v\}.$$

*For a given signal realization* $p_s$*, an informed trader will submit her marginal order to the venue* $v$ *that maximizes* $\Pi_v^*$ *among those with finite* $c_v$*.*

*Proof.* The objective $\Pi_v(q_v)$ is strictly concave in $q_v$ for any $\kappa_v > 0$, with derivative

$$\frac{\partial \Pi_v}{\partial q_v} = p_s - P_v - \kappa_v q_v - c_v.$$

The first-order condition $\frac{\partial \Pi_v}{\partial q_v} = 0$ yields $q_v^* = (p_s - P_v - c_v)/\kappa_v$. If $p_s - P_v \leq c_v$, the derivative at $q_v = 0$ is weakly negative and the optimal order is $q_v^* = 0$. Substituting $q_v^*$ back into $\Pi_v(q_v)$ gives the expression for $\Pi_v^*$. Since informed traders are risk-neutral and trading opportunities across venues are independent, a trader allocates her marginal order to the venue with highest $\Pi_v^*$ subject to feasibility (finite $c_v$). □

### B.4 High- vs Low-Frequency Leadership

We now show how latency and depth generate different leaders at different sampling frequencies. Consider a small time window $[t, t + \Delta]$ following an information arrival, and suppose that before the shock all venues quote the same price $P_A(t^-) = P_E(t^-) = P_B(t^-) = P_0$.

**Proposition 2** (Frequency-dependent price discovery)**.** *Suppose (i) $\tau_A = 0 < \bar{\tau}_E \le \bar{\tau}_B$ and (ii) $\kappa_A > \kappa_E$. Then:*

1. *For sufficiently small windows $\Delta < \bar{\tau}_E$, only the AMM can adjust its price in $[t, t+\Delta]$, so the AMM accounts for (almost) all efficient-price innovations at this horizon.*
2. *For sufficiently large windows $\Delta \gg \bar{\tau}_E$, a larger share of the cumulative efficient-price innovation variance is generated by the exchange than by the AMM, provided that informed signal arrivals are sufficiently frequent ($\lambda$ and $\mu_I$ large) and participation costs at A and E are comparable.*

*Sketch of argument.* For part (i), if $\Delta < \bar{\tau}_E$, then by definition of latency, $P_E$ and $P_B$ cannot adjust in $[t, t+\Delta]$ following a new signal. Only the AMM $A$ can update its price through orders placed by informed traders. Under cointegration, the common efficient price must therefore move in lockstep with $P_A$ over this horizon, so any econometric decomposition attributes (almost) all high-frequency innovations to the AMM.

For part (ii), fix a longer horizon $\Delta \gg \bar{\tau}_E$. Within $[t, t+\Delta]$, both $A$ and $E$ receive many informed orders. By Proposition 1, for a given signal size $|p_s - P_v|$ and similar $c_v$, the optimal order size is inversely proportional to $\kappa_v$, so the cumulative quantity traded by informed traders at the exchange is larger than at the AMM because $\kappa_E < \kappa_A$. Aggregating over many signals and traders, the cumulative efficient-price innovation—a weighted sum of order flows across venues—will thus load more on the deeper exchange than on the thinner AMM. In the limit as the number of signals in $[t, t+\Delta]$ grows, the law of large numbers implies that the variance contribution of exchange-driven innovations dominates that of AMM-driven innovations. This implies a higher information share for $E$ than for $A$ at low sampling frequencies. □

Proposition 2 maps directly into our empirical finding that PolyMarket leads price discovery at 5-minute frequency while Betfair leads at 1-hour frequency (Table 9), and provides the theoretical foundation for H4.

### B.5 Regulated Entry and Information Leadership

We now extend the model to two exchange-type venues: an incumbent offshore exchange $E_o$ and a newly opened regulated exchange $E_r$ (Kalshi-type). Both share similar depth but serve different clienteles.

Before $E_r$ opens, a mass $\mu_R$ of "regulated" informed traders face $c_A = c_{E_o} = +\infty$ and cannot trade; only the mass $\mu_U$ of unconstrained informed traders participate on $A$ and $E_o$. At time $t = t^*$, $E_r$ opens with participation cost $c_{E_r} < +\infty$ for the regulated traders and depth parameter $\kappa_{E_r} \approx \kappa_{E_o}$.

**Proposition 3** (Rapid catch-up of a regulated exchange)**.** *Suppose that (i) $\mu_R > 0$ informed traders can only trade at the regulated exchange $E_r$ after entry, (ii) $\kappa_{E_r}$ is comparable to $\kappa_{E_o}$, and (iii) the arrival rate and informativeness of signals for regulated traders are similar to those of unconstrained traders. Then, over a sufficiently short post-entry window $[t^*, t^* + \Delta]$, the regulated exchange $E_r$ can contribute a share of efficient-price innovations comparable to that of the incumbent AMM A, even if its total trading volume is smaller.*

*Sketch of argument.* After $t^*$, each regulated trader with signal $p_s$ solves the same problem as in Proposition 1 but can only choose $E_r$. Given $\kappa_{E_r}$ and $c_{E_r}$, each signal generates a strictly positive optimal order $q^*_{E_r}$ whenever $|p_s - P_{E_r}| > c_{E_r}$. Since $\mu_R > 0$ and signals arrive at rate $\lambda$, the aggregate informed order flow to $E_r$ over

$[t^*, t^* + \Delta]$ is of order $\mu_R \lambda \Delta$, which can be large even if $\Delta$ is modest. Under the maintained assumption that the distribution of signals for regulated traders is as informative as for unconstrained traders, these orders are *highly selected* on information: they come from traders who were previously unable to trade anywhere.

By contrast, the AMM $A$ receives order flow only from unconstrained traders, of measure $\mu_U$, with similar signal structure. Over a short horizon, the variance in the efficient price innovation attributed to each venue is proportional to the variance in its informed order flow multiplied by its price impact. Because $\kappa_{E_r}$ and $\kappa_A$ are of comparable magnitude and the signal variances are similar by assumption, the relative contribution of $E_r$ and $A$ to total innovation variance is approximately proportional to their respective informed order-flow variances. For $\mu_R$ not too small relative to $\mu_U$, this implies that $E_r$'s information share can quickly approach that of $A$ even if the *level* of volume at $E_r$ remains below that of $A$. □

Proposition 3 provides a simple mechanism through which a newly launched regulated exchange can reach information-share parity with an established prediction market within a few weeks, as we document for Kalshi in Table 5. This is the theoretical foundation for H2.

### B.6 Bookmakers as Informational Followers

We finally sketch a simple model in which the bookmaker acts as an informational follower. Unlike $A$ and $E$, the bookmaker does not adjust prices mechanically with order flow. Instead, it periodically sets odds to balance expected profit against the reputational or regulatory cost of visibly diverging from the "efficient" probability.

Let $\tilde{P}(t)$ denote the efficient price implied by the AMM and exchange. At discrete review times $\{t_k\}_{k \geq 1}$, the bookmaker chooses a price $P_B(t_k)$ to solve

$$\max_{P_B} \mathbb{E}\big[\text{betting profit}(P_B) \,|\, \tilde{P}\big] - \gamma\big(P_B - \tilde{P}\big)^2, \qquad \gamma > 0. \tag{8}$$

The second term captures the cost of quoting odds that deviate from the consensus efficient price $\tilde{P}$.

**Proposition 4** (Bookmakers track the efficient price)**.** *Suppose that (i) bettor demand is symmetric around the perceived efficient price $\tilde{P}$ so that expected betting profit is locally maximized at $P_B = \tilde{P}$ in the absence of the penalty term, and (ii) the penalty parameter $\gamma > 0$. Then, at each review time $t_k$, the bookmaker's optimal odds satisfy*

$$P_B^*(t_k) = \tilde{P}(t_k).$$

*If reviews occur only when $|P_B - \tilde{P}|$ exceeds a threshold $\theta > 0$, then between reviews $P_B$ is a piecewise-constant, lagged, and thresholded version of $\tilde{P}$.*

*Proof.* By assumption (i), without the penalty term, the bookmaker's expected profit is locally maximized at $P_B = \tilde{P}$. A second-order Taylor expansion around $\tilde{P}$ implies that, to a first approximation, the expected profit term is quadratic and symmetric around $\tilde{P}$. Adding the quadratic penalty in (8), the overall objective is strictly concave and uniquely maximized at $P_B^* = \tilde{P}$. If the bookmaker only reviews odds when $|P_B - \tilde{P}| > \theta$, then $P_B$

remains fixed between review times and jumps toward $\tilde{P}$ when the deviation hits the threshold. Thus, the bookmaker's odds process is a lagged and thresholded transformation of the efficient price. $\square$

Proposition 4 implies that the bookmaker rarely initiates price changes and mostly adjusts odds in response to movements in the more informative venues. In a cointegrated system, this generates:

- small or zero information shares (the bookmaker contributes little to the variance of the efficient price innovation), and
- negative or near-zero component shares (bookmaker odds overshoot and then mean-revert toward the efficient price).

These predictions match our empirical findings (Table 4) and provide the theoretical foundation for H1.