# When the Efficient Price Is Observable:

# A Model-Free Decomposition of Price Impact

**Oliver Padmaperuma*** **Andre Guettler†**

September 7, 2026

Permanent price impact estimators infer an unobserved efficient price through structural models. We use novel data from prediction markets that remove this necessity. Its unique feature is observable binary outcomes that reveal the long-term efficient price. We split each trade's corrective value against the final outcome into price-moving (realized price impact) and outcome-predicting (anticipation) components. We find that a 1.9 percent share of informed traders earns $78 million, 94 percent through anticipation. Meanwhile, a 5.2 percent market-maker minority loses $101 million and displays the dealer markout reversal.

**JEL Classification:** C58, D47, D82, G12, G14.

**Keywords:** Prediction markets, Permanent price impact, Market microstructure, Polymarket.

---

*Ulm University, Institute of Strategic Management and Finance, Helmholtzstr. 22, 89081 Ulm, Germany, oliver.padmaperuma@uni-ulm.de (corresponding author)

†Ulm University, Institute of Strategic Management and Finance, Helmholtzstr. 22, 89081 Ulm, Germany, andre.guettler@uni-ulm.de

# 1 Introduction

Market microstructure is built on a price no one can observe. Kyle (1985) defines permanent price impact as the slope of the unobserved efficient price on signed informed order flow, and Hasbrouck (1991) recovers that slope through a structural vector autoregression under a long-run restriction the data cannot directly test. For over three decades, every reported price-impact estimate has been conditional on that latent construct. Prediction markets resolve this identification problem at its source. At resolution, the long-run equilibrium equals an externally observable binary outcome: one dollar if the event happens, zero if it does not. The unobserved efficient price becomes directly observed.

We implement this approach using data from Polymarket, the largest decentralized prediction market to date. Prediction markets aggregate trader beliefs about future events into a single transparent price, and a long literature documents their calibration and information-aggregation properties (Wolfers and Zitzewitz 2004; Manski 2006). Polymarket goes a step further than its predecessors: every trade, wallet, and position is recorded on a public blockchain ledger. We focus on its Politics & Elections[1] category: 22,321 resolved markets between 2023 and 2026, with approximately $5.6 billion of cumulative trade volume across more than a million unique investors. This setting is a clean laboratory for the classical theory of informed trading. Trade is discrete and event-driven rather than continuous. Market making is an emergent role filled by competitive participants rather than designated intermediaries. Order flow is not anonymous, because every order carries an on-chain identifier. Together with the observable resolution value, these features make two classical predictions directly testable. Kyle (1985) predicts that informed order flow earns the permanent price impact it creates. Glosten and Milgrom (1985) predict that the liquidity providers on the other side lose to that flow in expectation, recovering the loss only through the spread they charge uninformed traders.

[1]This category covers elections at every level, primaries, ballot measures, executive-order issuance dates, confirmation votes, and legislative actions.

We measure price discovery in two steps. The corrective impact of a trade is its signed size multiplied by the distance between the pre-trade price and the realized outcome: the dollar value of the price convergence the trade's direction claimed. Because the post-trade price path is also observable, this total splits exactly, with no model and no estimation, into two parts. The realized part is the price movement delivered within a fixed window of the trade. It is a markout, the standard object equity microstructure uses to measure dealer adverse selection. The remainder of this total arrives with later news, and we call it the anticipation. The split separates the two channels that a mark-to-resolution total conflates: earning by moving prices and earning by predicting them. We pair these impact measures with a Jensen (1968)-style bias-adjusted alpha. It nets out the favorite-longshot calibration spread, the mechanically attainable return from the tendency of longshots to be overpriced and favorites underpriced. The residual then measures private-information rents rather than spread harvesting. With these objects we classify the 497,588 active investors into market makers, informed traders, and noise traders, anchoring the three definitions in Glosten and Milgrom (1985), Kyle (1985), and Black (1986).

We find two main results. First, informed traders are 1.9 percent of active investors, supply 24 percent of aggregate corrective impact, achieve 76 percent accuracy, and earn 78 million U.S. dollars in aggregate bias-adjusted alpha. They earn it almost entirely through anticipation: only six percent of their corrective value is realized within a day of their trades. The price system pays informed investors for holding positions that later news vindicates, far more than for moving prices on contact. Second, market makers are 5.2 percent of active investors, supply 31 percent of aggregate corrective impact, achieve 30 percent directional accuracy, and lose 101 million U.S. dollars in aggregate bias-adjusted alpha. Their realized markout displays the textbook dealer reversal, positive at one minute and negative at one day, the Glosten and Milgrom adverse-selection signature. The remaining 93 percent are noise traders, who contribute trade volume but approximately zero net rent.

The corrective-impact measure agrees with the canonical Hasbrouck information share across markets, at a rank correlation of 0.40, and the agreement is strongest for the realized-impact part of our measure, rising to 0.47 at the one-day markout, which is the part a structural impact estimator is built to capture.

Both findings survive a battery of checks. Our three-way classification of investors is not an artifact of testing it on the same data that formed it. Built on the first half of the data and tested only on markets that resolve in the second, it keeps informed investors at 65 percent directional accuracy and market makers at 40 percent, on opposite sides of the 50 percent coin-flip benchmark. A simulation of the alpha cross-section under a no-skill null places both tails beyond what luck produces, separating genuine skill from chance. And the classification, with the same alpha signs, replicates under alternative thresholds and on the Sports and Crypto & Web3 subsamples.

Our paper contributes to three strands of the literature. The first is the microstructure literature on permanent price impact. Kyle (1985) provides the canonical theory: informed order flow moves prices permanently because it carries private information. Glosten and Milgrom (1985) add that the liquidity providers on the other side lose to that flow. Hasbrouck (1991) operationalizes permanent impact through the structural vector autoregression behind the information share, under a long-run restriction the data cannot test. A recurring problem is that impact measures miss much informed trading. Activists quietly building a large stake in a company, disclosed in a Schedule 13D filing, trade in deep markets and leave little immediate price footprint (Collin-Dufresne and Fos 2015), and prosecuted insider trades move prices only partially (Kacperczyk and Pagnotta 2019). The value they miss is realized later, as the price converges to the truth, and that convergence is what we set out to measure. Doing so on a decentralized venue means avoiding two measurement traps. Barardehi and Bernhardt (2025) show that calendar-time aggregation distorts impact estimates, so we work in trade time, and Dubach (2026) finds that off-chain trade signs are wrong roughly 41 percent of the time, so we

take signs from the on-chain record. With those choices in place, we generalize the venue's first price-impact estimate, fit by Tsang and Yang (2026) on a single market, to the full investor cross-section. We replace the structural identification step with the observable terminal value of binary contracts, and we split the mark-to-resolution total into a realized markout and the anticipation that impact estimators miss.

The second is the prediction-markets literature. Forsythe et al. (1992) show on the Iowa Electronic Markets that political prices forecast elections accurately even when individual traders are biased. Wolfers and Zitzewitz (2004) survey the field and its information-aggregation properties, and Manski (2006) shows that a market price need not equal the mean trader belief, making its interpretation an identification question. Recent work takes this to Polymarket, moving from the market price to the individual investors behind it. Reichenbach and Walther (2025) find Polymarket prices close to calibrated at the market level, and the closest antecedents push to the investor level: Gomez-Cram et al. (2026) identify an informed minority of roughly 3 percent whose gains are financed by a losing majority, and Ng et al. (2026) show that order-flow imbalance predicts returns and that Polymarket leads competing venues in price discovery. Our partition corroborates the informed-minority finding in shape. Our contribution is to decompose it: we split how the informed minority earns into moving prices and anticipating news, and find it is almost entirely anticipation, the split the observable terminal value makes possible. We also test the classification more carefully than the prior cross-sectional evidence, replicating it out of sample, clustering our standard errors by market, and comparing it against a no-skill benchmark.

The third is the literature on measuring skill. Jensen (1968) defines alpha as performance above a mechanically attainable benchmark, the principle our calibration-spread adjustment implements. Berk and Green (2004) explain why genuine skill can persist while the rents it earns are competed away. The tools for separating skill from luck follow: Kosowski et al. (2006) introduce the bootstrap that compares the

tails of the estimated alpha distribution with a no-skill null, and Fama and French (2010) find that few mutual funds clear it. Barber et al. (2014) bring the question to the cross-section of individual speculators, where the binding constraint is separating skill from luck and from trading costs over identities that must be inferred, and Berk and Binsbergen (2015) argue that dollar value added is the right measure of skill. Our setting removes those constraints. Every investor's full trade history is directly observed, and the calibration spread supplies the natural benchmark. We regress bias-adjusted alpha on directly measured corrective value across the active cross-section, finding a one-for-one conversion and tail skill that is real and persistent in direction, while the dollar rents rotate across cycles, the Berk and Green pattern at the individual-investor level.

# 2 Setting and Data

## 2.1 Institutional Background

Polymarket is a decentralized prediction-market venue on which binary event contracts can be bought and sold continuously until the underlying event resolves. It runs on Polygon, an Ethereum layer-two public blockchain, so every account and every trade is recorded openly and permanently. Each contract is an *outcome token*, a share that pays one U.S. dollar if the event happens and zero otherwise.[2] Multi-outcome events (for example, "Who wins the next U.S. presidential election?") are decomposed by the venue into one contract per candidate, with a built-in constraint that the implicit probabilities sum to one. The mechanism that enforces this constraint is known on the venue as *NegRisk*.[3]

[2]Prices and payouts are denominated in USDC, a dollar-pegged stablecoin. We report all amounts one-to-one as U.S. dollars, with no exchange-rate conversion.

[3]NegRisk packages a set of mutually exclusive binary contracts so that a position long in all-but-one of the outcomes is equivalent to a position short in the remaining one. This allows the venue to settle every multi-outcome event using the same binary-payoff infrastructure used for two-outcome events. The reader unfamiliar with the technical detail can read NegRisk as the venue's name for the multi-outcome counterpart of its binary settlement mechanism.

Settlement is centrally adjudicated by an off-chain reporter that posts the realized outcome to the venue once the event has resolved, and on-chain prices then snap mechanically to the terminal value. All trade events, including the identifier of every participant on each side of every trade, are publicly observable on the underlying Polygon blockchain, so every trade links back to an account and forward to the eventual resolution. The persistent identifier is the infrastructure feature that distinguishes the venue from centralized prediction and betting markets. Orders are matched and trades settled on-chain by the venue's two exchange contracts, the CTF Exchange for binary markets and the NegRisk CTF Exchange for multi-outcome events.[4] Polymarket therefore functions as a continuous limit order book on binary contracts with no designated market maker, full post-trade transparency, and known terminal value at resolution.

Online prediction-market venues have grown materially in the past decade. Estimated cumulative trade volume on the broader online prediction-market ecosystem, combining decentralized venues (Polymarket, Augur, Manifold) and centrally cleared venues (Kalshi, PredictIt), exceeds $60 billion at trade-price valuation by mid-2026 (Gomez-Cram et al. 2026; Ng et al. 2026). Polymarket is the largest decentralized venue, with approximately $15.8 billion in cumulative trade volume across all resolved markets, an order of magnitude above on-chain competitors such as Augur. Our dataset spans the venue's *Politics & Elections* category, one of its largest, with $5.6 billion of cumulative trade volume, roughly a third of the venue's resolved-market total. These markets also draw the widest public attention, and they resolve on widely reported outcomes that are hard to manipulate at the source, unlike a market settled from a single instrument such as an airport temperature sensor.

[4]CTF abbreviates Conditional Tokens Framework, the on-chain primitive Polymarket uses to mint and settle outcome tokens. These contracts appear in the trade ledger as the recorded counterparty on the duplicate settlement rows we exclude in Section 2.3 and Appendix B.

## 2.2 Data

We use the Polymarket release curated by Becker (2026), who indexed the venue's full on-chain trade ledger directly from the Polygon blockchain and published the dataset. The release covers block 40,000,176 through block 82,120,997, corresponding to the venue's CTF Exchange launch in early 2023 through January 2026. It comprises approximately 404 million `OrderFilled`[5] trade events.

A trade event is recorded each time a buy order on one side is filled against a sell order on the other side. The on-chain log identifies the two counterparties, the contract being traded, the trade size, and the dollar amount transferred. We recover wall-clock timestamps from block numbers by sampling every $100^{\text{th}}$ block from the Polygon RPC and interpolating linearly between samples. Polygon produces a block roughly every two seconds, so consecutive samples are about three minutes apart, and fetching an exact timestamp for each of the release's forty-two million blocks would be impractical. The interpolation error of a few minutes is immaterial, because it enters only the monthly trade-count series and the time-to-resolution filter, not the markout windows, which are counted in blocks. We collect market metadata (contract identifiers, final settled outcome, resolution date, total trade volume) from the venue's Gamma metadata API.

## 2.3 Sample Construction

We build the analysis sample in two stages: a restriction to the resolved Politics & Elections universe (hereafter Politics), and a set of filters applied at the trade level. A market is *resolved* when one outcome's closing price exceeds \$0.99 and the complementary outcome closes below \$0.01.[6] For each resolved market we record a binary indicator $y_m \in \{0, 1\}$ for the winning side. The observability of $y_m$ distin-

[5] `OrderFilled` is the on-chain log entry emitted by Polymarket's CTF Exchange and NegRisk CTF Exchange smart contracts each time a trade is settled on the Polygon blockchain.

[6] A resolved outcome pays exactly one dollar on-chain, but the last recorded trade seldom executes at that value, because a near-certain winner offers little upside net of trading costs. The closing price therefore settles within a cent of the boundary rather than at it.

guishes prediction markets from equity microstructure and motivates our estimator in Section 3.2.

The Politics category contains 22,321 resolved markets. Its scope is not confined to the United States. The events range from U.S. national and state races to elections and political questions across Europe, the Americas, the Middle East, and Asia. The 2024 U.S. presidential cycle is the volume centroid but not the exclusive content. The markets also cover primaries, ballot measures, executive-order issuance dates, confirmation votes, and individual legislative actions. This variety of event types supplies the within-category variation that identifies trader heterogeneity, and the common-knowledge resolution mechanism gives the cleanest data-generating process for the resolution-anchored estimator.

At the trade level we then make two kinds of exclusion. First, we drop rows that carry no economic content. Roughly 41 percent of raw `OrderFilled` rows are settlement-contract summary legs that aggregate the underlying fills, a known feature of Polymarket's on-chain data that we describe and verify in Appendix B, and we drop them to keep one row per fill. We also drop self-crossing trades, where the maker and taker addresses coincide. A trade an investor makes with itself carries no information. It most likely reflects wash trading to inflate a wallet's recorded volume, plausibly to farm airdrop eligibility, or to nudge the displayed price. Second, we apply three pathology filters that remove genuine trades which would contaminate per-trade impact even under a correctly specified estimator: *short-time-to-resolution* trades, within one day of the resolution event, which are contaminated by the terminal price snap, *post-resolution* trades, the mechanical wind-down once on-chain prices snap to $\{0.99, 0.01\}$, and trades in *microscopic markets* below \$1,000 in cumulative volume, where single-direction order flow adds noise without statistical power. Section 5 reports the sample with the three pathology filters relaxed.

These steps reduce the ledger to 34.1 million trades across 16,242 resolved markets and 1.15 million unique addresses, the sample on which the summary statistics

below are computed. Investor-level measures use the *active* subset, the 497,588 addresses with at least ten post-filter trades. The partition is insensitive to this cut (Section 5).

## 2.4 Summary Statistics

Trading in our sample is event-driven. Figure 1 shows monthly trade counts close to zero until the 2024 U.S. election cycle, spiking around the November election, then settling to a lower but still-elevated baseline through 2025 and 2026. Before that cycle Polymarket was known mainly to crypto-native users. The election brought it broad public attention, which the spike reflects. Table 1 reports the per-trade, per-investor, and per-market distributions of this sample, with variable definitions in Table A.1. Three features of this sample shape how we measure and decompose price discovery.

The first feature is a strong dispersion: the median trade moves 39 outcome tokens and the largest moves 6.2 million. The median active investor places 23 trades and the most active places 730,000. The median market settles $51,198 in volume over a 17-day life, while the 2024 Trump presidential market settles $1.53 billion and the longest-dated election futures run for 935 days. At every level the mean sits far above the median, the signature of a distribution whose mass lives in the tail.

That dispersion, however, is not skill. Directional accuracy, the share of an investor's trades that point toward the realized outcome, averages 0.510 across active investors and has a median of exactly 0.500, the benchmark of a coin flip. Panel A of Figure 2 plots the full distribution: it spikes sharply at 0.50 and falls away in thin, nearly symmetric tails. Therefore, the median investor loses money, with realized profit and loss slightly below zero. The account behind the average trade neither predicts outcomes nor extracts rents.

And yet the aggregate is accurate. Panel B of the same figure plots the empirical resolution rate against the pre-trade price: contracts trading near \$0.20 win about one time in five, contracts near \$0.80 about four times in five, and the curve tracks the 45° identity tightly near the boundaries, where three quarters of trades execute. The mid-range deviates systematically. Contracts priced between 0.33 and 0.49 resolve in favor about five percentage points less often than their price implies, and contracts priced between 0.51 and 0.67 about seven points more often, so mid-range prices are compressed toward one half. This is the favorite-longshot pattern that Snowberg and Wolfers (2010) document in betting markets, visible here exactly where outcome uncertainty is greatest. It is also the systematic price-outcome gap that the calibration benchmark of Section 3.4 absorbs, so the skill measure we build there does not credit investors for mechanically harvesting it.

# 3 Methodology

Three empirical objects do the work in the rest of the paper. The first is the resolution-anchored corrective impact and its exact decomposition into a realized price-path component and a convergence residual. The observable terminal value of a binary prediction contract replaces the latent efficient price that Kyle (1985), Glosten and Milgrom (1985), and Hasbrouck (1991, 1995) must infer. The decomposition then separates the part of an investor's corrective value that the price system delivered around their trades, the Kyle (1985)-comparable impact, from the part delivered by later convergence, the anticipation. The second is the corrective price contribution, an attribution to each executed trade that telescopes exactly within each contract, in the tradition of Barclay and Warner (1993). Address-level contributions therefore exhaust each contract's total price discovery by construction. The third is a bias-adjusted alpha. It pairs the impact measures with a Jensen (1968)-style benchmark that strips out the empirical favorite-longshot calibration spread (Snowberg and Wolfers 2010), isolating skill rents from mechanical spread harvesting.

## 3.1 Setup and Notation

We index trades by $j$ and the market in which they occur by $m$. For each trade $j$ at calendar time $t_j$, $p_j^-$ is the pre-trade price of the traded outcome token, meaning the execution price of the most recent prior trade in the same token (the venue exposes prints, its record of executed trades, not quotes), and $p_j^+(h)$ is the prevailing price $h$ after the trade: the price of the last trade in the window $(t_j, t_j + h]$, or trade $j$'s own execution price when no later trade lands in the window, so that the last print stands. Signed taker-side size is $s_j$, positive for buys and negative for sells. Market $m$ resolves at calendar time $T_m$ to the binary outcome $y_m \in \{0, 1\}$ introduced in Section 2.3, written $y_j$ when indexed by a trade $j$ in that market, and belongs to a category $c_m \in \mathcal{C}$. Investor $i$ trades the set of trades $\mathcal{J}_i$. Table A.1 is the complete variable glossary.

## 3.2 Corrective Impact

Kyle (1985) defines the permanent price-impact coefficient $\lambda$ as the structural parameter that maps signed informed order flow to an unobserved efficient price, and Hasbrouck (1991, 1995) operationalizes it through a structural vector autoregression that splits mid-return innovations into a permanent component, attributed to information, and a transitory component, attributed to inventory and discreteness.

Every estimator in this family rests on the same identifying assumption: a long-run efficient price process $p_t^*$ that is a martingale with respect to public information, with the observed mid equal to $p_t^*$ plus a mean-reverting pricing error. The efficient price is never observed directly. The estimator infers it from the joint dynamics of returns and order flow under a parametric restriction (typically a VAR with a specified lag length and orthogonalization scheme), and the restriction is not directly testable. Where the assumed long-run equilibrium is mis-specified or the empirical horizon does not exhaust the structural impulse response, the decomposition of permanent and transitory components is biased.

Binary prediction contracts on Polymarket, by contrast, admit an externally observable terminal value. Each market $m$ resolves to $y_m \in \{0, 1\}$ at a known calendar time $T_m$, both outcome tokens then converge mechanically to their terminal values, and the post-resolution mid is pinned at $y_m$ by construction. The Kyle long-run efficient price is directly observable: $\lim_{h \to T_m - t_j} p_j^+(h) = y_m$ for every trade $j \in \mathcal{J}_m$. The structural-VAR identification step is unnecessary, because the structural target itself is observed.

We exploit this observability in two steps: a total, and an exact decomposition of the total into a price-path component and a residual. For each trade $j$ in a resolved market $m$, the *corrective impact* at the end-of-market horizon is

$$\mathrm{CI}_j(\mathrm{EOM}) \;\equiv\; s_j \cdot \big(y_m - p_j^-\big). \tag{1}$$

The factor $y_m - p_j^-$ is the remaining distance between the pre-trade price and the eventual truth, so the product is positive whenever the trade direction matched the direction toward truth (a buy when $y_m > p_j^-$, a sell when $y_m < p_j^-$). The unit is U.S. dollars: $\mathrm{CI}_j$ is the dollar value of the price convergence from the pre-trade price to the terminal value, claimed at trade $j$'s size and direction. It is a mark-to-resolution total, and on its own it does not say when the price system delivered that value: a trade followed by an immediate, permanent revision and a trade followed by no movement until resolution day earn the same $\mathrm{CI}_j$.[7]

The decomposition supplies the missing price path. For every horizon $h$, the identity

$$\mathrm{CI}_j(\mathrm{EOM}) \;=\; \underbrace{s_j \cdot \big(p_j^+(h) - p_j^-\big)}_{\mathrm{RCI}_j(h)\text{: realized impact}} \;+\; \underbrace{s_j \cdot \big(y_m - p_j^+(h)\big)}_{\mathrm{CR}_j(h)\text{: convergence residual}} \tag{2}$$

holds exactly, with no estimation and no model. The *realized corrective impact* $\mathrm{RCI}_j(h)$ is a markout: the dollar value of the realized price move from just before

[7] The corrective impact is measured in price (dollar) space rather than the log-odds space of Tsang and Yang (2026), because it is by construction a dollar-valued total, for which a one-cent move is worth one cent at any price level. The bounded-price concern that motivates log-odds bears on slope estimation, and we address it there, in the log-odds Hasbrouck validation of Section 4.5.

the trade to $h$ after it, signed by the trade's direction. It is the price-path object. A trade that moves the price by nothing earns a realized impact of zero, whatever the outcome. As $h$ grows toward resolution, $\mathrm{RCI}_j(h)$ converges to $\mathrm{CI}_j(\mathrm{EOM})$ and the residual vanishes. The *convergence residual* $\mathrm{CR}_j(h)$ is the part of the mark-to-resolution value delivered by price movement after the impact window: anticipation of later news, or luck. The term structure of $\mathrm{RCI}_j(h)$ across horizons traces when the price system paid an investor, and the investor-level split separates the two channels the total conflates: earning by moving prices and earning by predicting them. Both are informed trading. Only the first is price discovery in the Kyle (1985) sense, and the markout form of RCI connects it to the standard toolkit for measuring dealer adverse selection in equity markets.

Three properties follow directly from the definitions, mirroring the linear structure of Kyle (1985)'s permanent impact, in which a trade of signed size $s_j$ moves the price by $\lambda \cdot s_j$. The objects are additive across trades within a market. They aggregate trivially to the investor level: $\mathrm{CI}_i \equiv \sum_{j \in \mathcal{J}_i} \mathrm{CI}_j$ is the dollar value of investor $i$'s corrective positions, and $\mathrm{RCI}_i(h)$ the part the price system realized within $h$ of $i$'s trades. And the identity survives every aggregation: $\mathrm{CI}_i = \mathrm{RCI}_i(h) + \mathrm{CR}_i(h)$ at any horizon, for any investor, with no assumption beyond the observability of $y_m$.

The contrast with the structural-VAR family is the absence of any identifying assumption on the long-run price process: $y_m$ is a known input, not a target to be inferred. Section 4.5 validates the measure against a per-market Hasbrouck information share fit on the same data. Tsang and Yang (2026) fit the first estimate of Kyle's $\lambda$ on Polymarket, a single time series on the 2024 Trump-Harris market, and our estimator generalizes this to the address-level cross-section. Dubach (2026) document that feed-inferred trade direction disagrees with the on-chain truth approximately 41 percent of the time, which motivates our reliance on the on-chain `taker` identifier.

## 3.3 Corrective Price Contribution

Corrective impact and its decomposition are dollar-valued and scale with trade size, which is the right unit for connecting impact to profit, but it leaves one question open: how does a market's total price discovery divide across the investors who traded it? Dollar objects cannot answer this with an adding-up constraint, because overlapping markout windows double-count the same price movement. The per-print attribution can. For each trade $j$, the *corrective price contribution* is

$$\mathrm{CPC}_j \;\equiv\; (2y_m - 1) \cdot \big(p_j - p_j^-\big), \tag{3}$$

where $p_j$ is trade $j$'s own execution price: the reduction in the contract price's distance to its resolution value delivered by that print, positive when the print moved the price toward the outcome. Because consecutive prints chain, with each trade's execution price becoming the next trade's pre-trade price, CPC telescopes exactly within a contract: the sum over a contract's trades equals $(2y_m - 1)(p_{\text{last}} - p_{\text{first}})$, the contract's total realized convergence over its trading history. Address-level sums therefore exhaust each contract's price discovery by construction. The estimand is the prediction-market analog of the weighted price contribution of Barclay and Warner (1993), which Barclay and Hendershott (2003) use to allocate price discovery across trading sessions. The resolution anchor fixes the sign convention that those settings must take from the realized return.

Two accounting conventions complete the definition. First, attribution is to the aggressor: a print is credited to the wallet in the taker field, or to the maker field on the settlement-contract summary prints documented in Appendix B, which name the same aggressor. Second, the attribution window excludes the terminal regime. Any partition of a contract's trades splits its total convergence exactly, and the pathology filters of Section 2.3 partition it into a pre-terminal investor part and the mechanical snap to the terminal value, which is settlement, not discovery.

## 3.4 Bias-Adjusted Alpha

The bias-adjusted alpha defined in this subsection answers a specific counterfactual question: did this investor outperform what someone with no private information but full knowledge of the favorite-longshot calibration curve, the systematic gap between contract prices and empirical resolution rates, could have earned? It is not the investor's total trading profit, which is realized PnL. It is not the dollar value of corrective price discovery the investor produces, which is the corrective impact of Section 3.2. Alpha is the residual after stripping a mechanically attainable favorite-longshot harvest out of realized PnL, and is therefore the right measure of *skill* (return from private information) rather than *trading profit* (which mechanically includes the calibration-spread harvest). On Polymarket the adjustment turns out to be empirically small: the estimated calibration curve stays close to the diagonal over most of the price range, so aggregate alpha tracks realized PnL closely. The value of the construction is not the size of the correction but the discipline it imposes. It makes the mechanical leg of every investor's PnL explicit, and it isolates the investors for whom the two measures diverge because their trades sit where the calibration spread bites.

Jensen (1968) introduced alpha as realized return net of the return attainable by passive exposure to systematic risk, and Sharpe (1992) and Carhart (1997) extend the principle to multi-factor benchmarks. Naive realized return therefore mixes mechanically attainable factor returns with skill, and the residual after stripping them out is the appropriate skill measure.

In prediction markets, the natural "factor" is not an equity index but the empirical *favorite-longshot calibration spread*. In betting markets, empirical resolution probabilities $\mu(p)$ deviate systematically from prices $p$: low-priced longshots are over-priced ($\mu(p) < p$ for small $p$) and high-priced favorites are under-priced ($\mu(p) > p$ for large $p$). Snowberg and Wolfers (2010) document the pattern across the betting-market literature, and Manski (2006) formalizes the broader identification problem

of inferring beliefs from prices. On prediction markets the deviation varies with the venue and the horizon. Page and Clemen (2013) find that the prior generation of venues was well calibrated only close to expiration, with a favorite-longshot bias that grows with time to expiration, while Reichenbach and Walther (2025) find little evidence of a general longshot bias on Polymarket at the market level. The benchmark does not presume a large spread: the empirical $\mu(p, c)$ estimates the calibration curve whatever its shape, and where calibration is exact the spread leg below vanishes and alpha coincides with realized PnL. Wherever the spread $\mu(p) - p$ is nonzero, however, a strategy that buys contracts in cells with $\mu(p) - p > 0$ earns it on average without private information. This calibration spread is the single benchmark factor we strip out, the prediction-market counterpart of the market factor in a one-factor performance model. Conditioning $\mu$ on price and category captures the dominant systematic gap. A finer conditioning set, for example on time-to-resolution, is a refinement we leave to future work.

A naive performance measure conflates skill with the calibration harvest. It decomposes investor $i$'s realized PnL as

$$\mathrm{PnL}_i \;=\; \sum_{j \in \mathcal{J}_i} s_j \cdot (y_j - p_j) \;=\; \underbrace{\sum_{j \in \mathcal{J}_i} s_j \cdot (\mu(p_j, c_j) - p_j)}_{\text{calibration-spread leg}} \;+\; \underbrace{\sum_{j \in \mathcal{J}_i} s_j \cdot (y_j - \mu(p_j, c_j))}_{\text{skill leg}},$$

where $\mu(p, c)$ is the empirical resolution probability at price $p$ and category $c$. The first leg is the PnL an investor would earn by mechanically following the empirical favorite-longshot curve in each $(p, c)$ cell. The second leg is the quantity we identify as skill: the dollar value of the investor's directional forecasts net of what could be earned without information.

We define the *bias-adjusted alpha* of investor $i$ as the skill leg of the decomposition:

$$\alpha_i \;\equiv\; \sum_{j \in \mathcal{J}_i} s_j \cdot \big(y_j - \mu(p_j, c_j)\big). \tag{4}$$

The empirical $\mu(p, c)$ is the resolution rate in each price-bin / category cell, estimated from the post-filter resolved-market sample across all eight categories. Because the realized outcome $y_j$ varies trade by trade, $\alpha_i$ is not a mechanical function of trade size. Within a $(p, c)$ cell, where $\mu$ is constant, each of the investor's trades adds $s_j \cdot (y_j - \mu)$ to $\alpha_i$. The cell's contribution is positive only when the contracts she bought resolve to one more often than the cell rate $\mu$, or the contracts she sold resolve to one less often. An investor whose bought and sold contracts both resolve to one at the cell's own rate earns $\alpha_i \approx 0$ regardless of how much she trades, because each term $s_j \cdot (y_j - \mu)$ is then zero in expectation. The dollar units mirror those of $\mathrm{CI}_i$. In Jensen (1968)'s language, $\alpha_i$ is the prediction-market analogue of CAPM alpha: realized PnL minus the return on the mechanical benchmark portfolio. In Grossman and Stiglitz (1980)'s language, it is the equilibrium information rent that compensates costly private information production. Section 4.4 regresses $\alpha_i$ on $\mathrm{CI}_i$ to identify which class of investor collects these rents.

The bias-adjusted alpha and the corrective impact differ by a calibration-spread wedge:

$$\alpha_j \;-\; \mathrm{CI}_j(\mathrm{EOM}) \;=\; s_j \cdot \big(p_j^- - \mu(p_j, c_j)\big). \tag{5}$$

The wedge vanishes when the pre-trade price equals the calibration value $\mu(p_j, c_j)$ and is widest at boundary prices, where the favorite-longshot harvest $\mu - p$ is largest. Combining the wedge with Equation (2) extends the identity through the whole framework: $\alpha_i = \mathrm{RCI}_i(h) + \mathrm{CR}_i(h) + \mathrm{wedge}_i$, so an investor's skill rent splits into an impact-earned part, an anticipation-earned part, and a calibration-spread part. Our paper carries all three measures because they answer different questions: corrective impact is how much price-toward-truth value an investor's positions claim, the decomposition says when the price system delivered it, and bias-adjusted alpha is how much of the resulting profit is skill rather than mechanical spread capture. Section 4.4 tests the relationship between alpha and corrective impact through a Wald test of $\beta_{\mathrm{CI}} = 1$ in the investor-level regression.

## 3.5 Identification Caveats

Three caveats govern interpretation of the estimates. First, the per-trade objects are attributions, not causal effects. $\mathrm{RCI}_j(h)$ books the entire realized price move in the window to trade $j$, without separating the trade's causal contribution from contemporaneous trades by other investors, and a short-window markout can be positive for a trade that pushes the price away from the eventual outcome. The identity in Equation (2) limits how far such pushes can distort the totals. Because the markout and the convergence residual sum to the same fixed corrective impact, whatever markout a trade books by moving the price in its own direction is subtracted one-for-one from its convergence residual, so pushing the price away from the eventual outcome cannot raise a trade's total. And when the push is transient, the reverting price move drains the markout itself out of investor-level RCI at longer horizons. $\mathrm{CPC}_j$ is exhaustive by construction but credits only executed prints, so an investor whose limit orders moved no print earns no contribution. We identify the population-level allocation of information rents through the cross-sectional analyses of Sections 4.1 through 4.4 rather than through any single-trade causal claim. Second, investor identity on Polymarket is the on-chain address and is not in general the same as natural-person identity. A single operator may split activity across multiple addresses. The Sybil issue is acknowledged in the recent on-chain microstructure literature (Sirolly et al. 2025) and is unresolved without an off-platform transfer-graph backfill that is beyond the scope of this paper. All investor-level inferences should therefore be read at the address level. Third, the favorite-longshot calibration $\mu(p, c)$ in Equation (4) is estimated in-sample on the post-filter resolved-market sample across all eight categories. The in-sample benchmark absorbs part of the realized outcomes it was estimated on, which dampens dollar-alpha dispersion. Section 5.6 addresses this from two directions: it re-estimates the cluster aggregates under a hold-out $\mu$ and finds the in-sample benchmark conservative, and it runs the luck-versus-skill inference on the normalized $t(\alpha_i)$, an investor's mean per-market alpha divided by its standard error, rather than on dollar alpha. A related second-

order point is that $\mu(p, c)$ is estimated on discrete price-by-category cells, so price variation within a cell is not reflected in the benchmark and leaves a small mechanical component in the residual. The pathology filters that shape the sample on which Equations (1) and (4) are evaluated are described in Section 2.3.

# 4 Who Drives Price Discovery?

Section 3 built the resolution-anchored measures. We now put them to work on the Politics sample to ask who drives the venue's price discovery. The representative trade carries almost no corrective signal. The median trade moves the price toward the eventual outcome by just +\$0.05 and the mean by −\$0.42, against a per-trade standard deviation of \$913 (Panel A of Table 1), because aggregate order flow averages opposing trades out. Whatever drives price discovery must therefore live in the cross-section of investors, not of trades.

## 4.1 Do the Many or the Few Move Prices?

A small minority does. Aggregating each trade's $\mathrm{CI}_j$ to the taker investor yields $\mathrm{CI}_i \equiv \sum_{j \in \mathcal{J}_i} \mathrm{CI}_j$. We restrict attention to active investors, those with at least ten post-filter trades, $N = 497{,}588$ of the 1,151,494 distinct investors in Politics markets. A signed $\mathrm{CI}_i$ can fall on either side of zero. It is positive when an investor's positions on net moved price toward the eventual outcome and negative when they moved it away, and about half of active investors land on the wrong side. To measure how much of the price-discovery action an investor accounts for, regardless of direction, we rank investors by $|\mathrm{CI}_i|$, the magnitude of their corrective footprint.

That footprint is concentrated in a thin tail. The cumulative distribution of $|\mathrm{CI}_i|$ hugs the horizontal axis until the rightmost few percent of investors, where it turns sharply upward (Figure 3). The top investor alone captures 2.6 percent of total $|\mathrm{CI}|$, the top one hundred capture 27.8 percent, and the top one thousand capture 52.9 percent, while the bottom three quarters of active investors contribute

under one percent between them. Resampling markets with replacement, the natural dependence unit, leaves the top-100 share within roughly $[16\%, 29\%]$ across 500 draws.[8] The platform's Politics calibration is not produced by broad participation.

The tail also rotates. Splitting the sample at the median resolution date and recomputing the top-100 ranking within each half, only 6 of the first-half top-100 investors reappear in the second-half top-100. This is an address-level count, and it overstates true rotation to the extent that one operator spreads activity across wallets (Section 3.5), so we read it as an upper bound on turnover. Even at that bound the informational tail is a rotating cohort of event-specialists rather than a fixed roster, the cross-sectional signature of Grossman and Stiglitz (1980) information rents: the marginal information producer changes as the news cycle changes. An exhaustive per-print attribution that must sum to each contract's total convergence (the corrective price contribution of Equation (3)) tells the same story from the opposite direction. Only 45 percent of active addresses are net-positive contributors, and the top one hundred supply 23 percent of the positive mass, so the concentration is not an artifact of weighting investors by position size.

## 4.2 Market Makers, Informed, or Noise?

Having localized the corrective signal in a thin investor tail, we ask who those investors are. The classical microstructure roles each leave an observable signature in the resolved data. Liquidity suppliers picked off by informed flow sit below 50 percent directional accuracy (Glosten and Milgrom 1985), informed traders sit above it (Kyle 1985), and noise flow sits near it (Black 1986). We therefore partition active investors on three per-investor features, each the observable footprint of one role's defining property: cumulative volume $V_i$, the operational trace of liquidity provision, mean per-trade absolute corrective impact $\overline{|\mathrm{CI}|}_i$, the trace of private-information

[8] We report top-$k$ shares rather than the Gini coefficient. The Gini on $|\mathrm{CI}_i|$ is dominated by trade-volume heterogeneity rather than by informational variation: a sign-permutation placebo that redraws each market's binary outcome with probability one half (200 draws) yields a Gini distribution centered at 0.965, with the actual sample Gini of 0.967 inside the placebo interval.

content, and directional accuracy $r_i$, the direction of that information. The decile-cut classification is

$$\begin{aligned} i \in \text{market maker} &\iff V_i \geq P_{90}(V) \text{ and } r_i < 0.50, \\ i \in \text{informed} &\iff \overline{|\text{CI}|}_i \geq P_{90}(\overline{|\text{CI}|}) \text{ and } r_i > 0.60, \\ i \in \text{noise} &\iff \text{all remaining active investors.} \end{aligned}$$

The market-maker definition encodes Glosten and Milgrom (1985): high volume is the footprint of liquidity provision, and sub-50 percent accuracy is the adverse-selection signature. The informed definition encodes Kyle (1985): large per-trade |CI| marks trades with high private-information content, and above-60 percent accuracy separates information from noise. The noise category is the residual.

Table 2 reports cluster sizes, mean features, and per-cluster aggregates of $\text{CI}_i$ and $\alpha_i$, and Figure 4 plots each active investor in the $(V_i, \overline{|\text{CI}|}_i)$ plane. Across the three clusters, informed traders earn, market makers lose, and noise traders break even. They are 1.9 percent, 5.2 percent, and 92.9 percent of active investors respectively (Table 2).

Informed traders contribute 24 percent of Politics-wide |CI| on a mean directional accuracy of 76 percent. The cluster earns an aggregate bias-adjusted alpha of +$78 million and realized PnL of +$99 million, the slight excess of PnL reflecting that informed investors execute at boundary prices where the favorite-longshot spread marginally favors their bets. Market makers are three times as numerous and carry the opposite sign throughout. The cluster contributes 31 percent of Politics-wide |CI| on a mean accuracy of 30 percent, below the 50 percent Glosten-Milgrom threshold, and generates aggregate alpha of −$101 million and realized PnL of −$132 million. Liquidity provision in Politics is loss-making in aggregate.[9] In Glosten and Milgrom

[9]This loss is measured as mark-to-resolution trading value and bias-adjusted alpha. It excludes the liquidity-provider rewards Polymarket has paid through maker-incentive programs and the expected value of platform token airdrops, both of which accrue disproportionately to high-volume resting liquidity. A market maker that loses at resolution may therefore still break even or profit once these off-measure incentives are counted, so the negative aggregate bounds the trading return, not the all-in return to liquidity provision on the venue.

(1985), competitive market makers lose to informed counterparties but break even overall, because the spread charged to uninformed flow compensates the loss. The negative aggregate here says the compensating revenue falls short on this venue, and the $31 million wedge between PnL and alpha is the calibration-spread loss the cluster absorbs on top of its directional loss. Noise traders supply the remaining 45 percent of corrective mass by sheer number rather than per-trade precision and net to approximately zero alpha and PnL. Market-resampled 95% intervals separate the three aggregates from zero where the text claims a sign: market-maker alpha $[-\$172, -\$47]$ million and informed alpha $[+\$29, +\$142]$ million, while noise alpha $[-\$5, +\$26]$ million straddles zero (Section 5.2), and a no-skill simulation places the informed and market-maker tails beyond what luck produces (Section 5.6). The partition is not an artifact of the classification thresholds, the sample construction, or the Politics universe, as Section 5 establishes through a threshold sweep (Section 5.3), sample-construction perturbations, and a cross-category replication.

## 4.3 Do Traders Move Prices or Predict Them?

The partition so far rests on the mark-to-resolution total, which cannot say when the price system delivered each cluster's corrective value. The decomposition of Equation (2) supplies the answer, and it is the result the observable outcome uniquely enables. Table 3 splits each cluster's aggregate corrective impact into the realized markout $\sum \mathrm{RCI}_i(h)$ at four horizons and the one-day convergence residual, and Figure 5 plots the three term structures. The cluster assignment is the one of Table 2. Only the accounting of each cluster's corrective value changes.

The market-maker cluster displays the textbook dealer markout reversal. Its realized markout is positive at one minute (+$3.9 million, $[+3.2, +4.6]$), shrinks through the five-minute and one-hour horizons, and turns negative at one day ($-$6.7 million, $[-10.1, -3.9]$). The cluster earns the spread on the prints it supplies and is then run over by the information-horizon move. This is the adverse-selection signature that decades of equity microstructure measure with dealer markouts, recovered

here without using the realized outcome's sign, so it is free of any mechanical link between the accuracy-based cluster definition and the outcome-anchored total. The cluster's one-day convergence residual of −$120.1 million carries the rest of its loss: positions held against the eventual outcome.

The informed cluster's markout is positive at every horizon, growing from +$2.3 million [+1.9, +2.8] at one minute to +$6.6 million [+4.3, +10.1] at one day. Yet this one-day markout is only 6.5 percent of the cluster's +$102.3 million corrective total, so informed investors earn overwhelmingly through the convergence residual, by holding positions that later news vindicates, rather than through contemporaneous price impact. The sample as a whole behaves the same way: across all post-filter Politics trades the gross realized share $\sum |\mathrm{RCI}(h)| / \sum |\mathrm{CI}|$ is 1.9 percent at one minute and reaches only 13.5 percent at one day. Price discovery on the venue operates mainly through news arrival rather than order-flow pressure. Noise traders are the mirror image of the informed cluster: 60 percent of their small positive total is realized within a day of their trades, the footprint of momentum-chasing flow that buys into moves already under way. Informed investors are paid for anticipation, market makers for immediacy, and the two are paid on opposite horizons.

## 4.4 Who Gets Paid, and Who Pays?

The typology shows who earns and who loses. We now tie the dollar skill measure to the dollar price-discovery measure, to confirm that the investors paid the information rent are the ones who produce the corrective impact, and to identify which microstructure role collects which rent.

Bias-adjusted alpha and corrective impact should move together, and the regression confirms that they do. Equation (5) shows that $\alpha_j - \mathrm{CI}_j = s_j(p_j^- - \mu(p_j, c_j))$, the per-trade favorite-longshot wedge, which is small in Politics because the empirical $\mu(p, c) \approx p$ across most price bins. Across the $N$ = 421,268 investors with at least ten post-filter trades on both sides, the cluster-bootstrap Spearman correlation

between $\alpha_i$ and $\mathrm{CI}_i$ is +0.78 on signed quantities (95% interval $[+0.78, +0.81]$) and +0.83 on absolute values ($[+0.83, +0.86]$).[10] A cross-sectional OLS regression of $\alpha_i$ on $\mathrm{CI}_i$ (Table 4, column (1)) puts the coefficient at +0.91 ($p < 0.001$, $R^2 = 0.95$), and a Wald test of $\beta_{\mathrm{CI}} = 1$ is not rejected ($p = 0.106$). We read this near-identity, $\alpha \approx \mathrm{CI}$, as the regression result the data support at market-level inference. Adding the $\mathrm{CI}_i \times \log V_i$ interaction lowers the coefficient to +0.64 and points to a volume-dependent conversion, but that interaction does not survive market-level resampling, so we do not rely on it (Section 5.2). The coefficient is not a trading-style artifact either: it is unchanged when the ex-ante behavioral controls introduced later in Section 5.4 (maker share, two-sidedness, and log gross volume) are added (columns (5) and (6)).

The typology and the near-identity together recover all three microstructure predictions. The 1.9 percent informed minority captures the positive aggregate alpha, the informed-flow rent of Kyle (1985). The 5.2 percent market-maker minority absorbs the corresponding loss, the Glosten and Milgrom (1985) adverse-selection prediction. The 92.9 percent noise majority contributes investor count but not net rent, the Black (1986) and Kyle (1985) framing of noise as liquidity flow uncorrelated with fundamentals. The platform-level calibration documented by Wolfers and Zitzewitz (2004) and Reichenbach and Walther (2025) is therefore delivered by a thin specialist cross-section rather than by broad participation, in a way the on-chain data lets us identify directly.

## 4.5 Does the Measure Recover Known Price Impact?

The findings rest on the resolution-anchored measure, so we close by checking that it recovers what the established estimator recovers. A measure that dispenses with the structural identification step owes that demonstration. We fit a per-market Hasbrouck structural VAR on the top resolved markets by volume and compare its

[10] The regression sample of Table 4, $N = 488{,}083$, requires ten trades only on the corrective-impact side. The alpha leg's interior-price restriction removes the difference.

information share $\mathrm{IS}_m$, the share of long-run return variance attributable to order-flow innovations, with our market-level mean |CI|. Across the $N = 422$ markets that admit a clean fit, the Spearman rank correlation is $\rho = +0.40$ ($p < 0.001$), rising to $+0.44$ on the Politics-only subsample (Appendix C details the construction).

The decomposition sharpens the test. The information share attributes variance to order-flow innovations, so if the realized markout $\mathrm{RCI}(h)$ isolates the order-flow-driven leg of corrective impact, its ranking should agree with $\mathrm{IS}_m$ more closely than the mark-to-resolution total does. It does, and monotonically in the horizon: replacing mean |CI(EOM)| with mean $|\mathrm{RCI}(h)|$ moves the Spearman from $+0.40$ at one minute to $+0.41$, $+0.43$, and $+0.47$ at five minutes, one hour, and one day. The agreement is strongest exactly where our estimator claims to measure realized price impact, which is the behavior Equation (2) predicts. The realized leg recovers the impact a structural VAR is built to capture. The anticipation leg, which carries the bulk of informed value, is the part the latent-price estimator cannot see, because no contemporaneous order flow accompanies it. The absolute level of the information share is itself low, far below what order-flow decompositions recover in equity markets, because Polymarket prices move predominantly on exogenous news. That is the regime in which the observable terminal value helps most, because the news-driven component of a price move does not have to be backed out of order-flow innovations.

# 5 Robustness

The empirical findings rest on a partition of investors and a set of dollar magnitudes, and this section subjects both to the perturbations a skeptic would apply. We show that the partition is not an artifact of how the sample is built, that the dollar claims survive market-level dependence and the removal of the 2024 cycle, that the partition is insensitive to the classification thresholds, that the trader types persist

out of sample and replicate outside Politics, and that the skill in the alpha tails exceeds what luck produces.

## 5.1 Robustness to Sample Construction

The trader partition of Section 4.2 could in principle be an artifact of how the sample is built. Four perturbations say it is not.

First, the pathology filters can each be toggled. Disabling the short-time-to-resolution, post-resolution, and minimum-volume filters raises the Politics trade count by about a fifth and lifts the aggregate mean per-trade CI at the end-of-market horizon from \$0.40 to \$0.44, because the very-near-resolution trades the filters remove point disproportionately toward the already-known outcome. Re-running the classification on the unfiltered sample shifts the cluster shares by less than 0.3 percentage points (market maker 5.2 to 5.5, informed 1.85 to 1.86, noise 92.9 to 92.6 percent) and preserves the alpha signs (market maker −\$101 to −\$111 million, informed +\$78 to +\$95 million, noise near zero at +\$10 to +\$3 million). Self-trades and platform settlement contracts are excluded throughout (Appendix B).

Second, the concentration and the per-cluster aggregates weight each investor by cumulative |CI|, which scales with trade count, so a handful of high-volume investors could drive them mechanically. Winsorizing the active sample at the 99th percentile of trade count (about 470 post-filter trades) leaves the cluster shares essentially unchanged at 5.2 / 1.8 / 92.9 percent. The alpha magnitudes shrink, because some of the largest informed and market-maker representatives sit above the cutoff and are dropped (market-maker aggregate from −\$101 to −\$76 million, informed from +\$78 to +\$29 million, noise from +\$10 to −\$2 million), but the signs hold. The concentration is not the work of a few high-volume accounts.

Third, a wash-trading screen in the spirit of Sirolly et al. (2025), who detect tight wallet clusters that churn volume among one another without taking net positions, removes every active address whose two-sidedness exceeds 0.9, the behavioral

signature of round-trip churn.[11] The screen drops 22 percent of active addresses and leaves the partition essentially unchanged: cluster shares 6.1 / 1.9 / 92.0 percent, market-maker accuracy 29.5 percent, informed accuracy 76.6 percent, and aggregate alphas −$98 million and +$75 million.

Fourth, the ten-trade cut that defines an active investor. Lowering it to five or raising it to twenty leaves the partition intact. The cluster shares move only slightly (market maker 5.1 to 5.6 percent, informed 2.0 to 1.7 percent, noise 92.9 to 92.7 percent), and the alpha signs hold (market-maker aggregate −$107 to −$93 million, informed +$82 to +$72 million, noise near zero at +$8 to +$13 million). Across all four perturbations the cluster shares and the alpha signs hold. Only the dollar magnitudes move, and only when the largest positions are removed, the same thin-tail dependence the main results document.

## 5.2 Market-Level Inference

Thousands of investors share exposure to the same realized outcome, so the market, not the investor, is the natural dependence unit for every investor-level statistic, and the $1.53 billion 2024 presidential market is a single cluster. We re-derive the investor-level inference by resampling markets with replacement (500 draws) on the per-investor, per-market panel, re-aggregating every variable on each draw.

The cluster dollar aggregates keep their signs. The market-maker alpha interval is $[-\$172, -\$47]$ million and the informed interval $[+\$29, +\$142]$ million, while the noise interval $[-\$5, +\$26]$ million straddles zero, consistent with the approximately-zero reading in the text. The intervals are wide, as they must be when 38 percent of sample volume resolves in a single quarter.

The same resampling overturns one main-text regression result, which is why Section 4.4 does not lean on it. The volume interaction in the alpha regression does

[11] Two-sidedness is the gross-volume-weighted average across an investor's tokens of one minus net over gross signed flow: one when buys and sells offset within a token, zero when the investor takes a single side. Variable definitions are collected in Table A.1.

not survive. Its 95% interval is $[-0.09, +0.06]$ and it is positive in only 36 percent of draws, while the main coefficient's interval $[0.44, 1.67]$ comfortably includes one. Modal-month clustering with 34 clusters understates this uncertainty, so we read the volume-dependent conversion rate as suggestive and the near-identity $\alpha \approx \text{CI}$ as the result the data support.

Finally, the dollar magnitudes are event-driven but the partition is not. Excluding the 2,136 markets that resolve between October 2024 and January 2025, which carry 38.1 percent of post-filter Politics volume, moves the cluster shares to 5.8 / 2.4 / 91.8 percent, lowers market-maker accuracy to 27.2 percent and raises informed accuracy to 77.0 percent, and leaves the aggregate alphas sign-stable at $-\$34$ and $+\$23$ million. Concentration attenuates sharply without the cycle, the top-100 share of $|\text{CI}_i|$ falling from 27.8 to 16.2 percent. The rents are a property of the cycle being resolved. The partition is a property of the venue.

## 5.3 Typology Sensitivity to Thresholds

The decile cuts that define the three types are deliberate but not unique, so the partition should not turn on the exact cutoffs. It does not. The pattern survives a threshold sweep (Table 5) over three decile-cutoff percentiles $\{0.75, 0.90, 0.95\}$, each crossed with three paired settings of the market-maker accuracy ceiling and the informed accuracy floor, $(0.45, 0.55)$, $(0.50, 0.60)$, and $(0.55, 0.65)$. Across all nine specifications the market-maker cluster delivers negative aggregate alpha (between $-\$89$ and $-\$109$ million) and the informed cluster delivers positive aggregate alpha (between $+\$65$ and $+\$98$ million) in every cell. Cluster sizes shift mechanically with the cutoffs. The sign and order of magnitude of the aggregates do not.

## 5.4 Out-of-Sample Validation

Two objections apply to any typology built on realized accuracy. The classification uses the realized outcome, through directional accuracy $r_i$, and so does the alpha by

which Table 2 judges it, so the in-sample finding that informed investors earn positive alpha is partly built in. And with a median of 23 trades per active investor, roughly a fifth of pure coin-flippers would clear the $r_i > 0.60$ gate by chance. We answer both with a split-sample design. Markets are divided at the median resolution block, so each outcome belongs to exactly one window. Classifications are computed on window-1 information and evaluated on window-2 markets only, for addresses with at least ten trades in each window (Table 6).

The accuracy dimension survives almost undiminished. Investors classified informed on window-1 accuracy and impact have an average window-2 accuracy of 65.1 percent (Panel B of Table 6), and investors classified market maker average 40.3 percent, far below the coin-flip rate their window-2 sample would produce under luck.[12] Directional accuracy is a persistent investor trait, not a selection artifact. Dollar magnitudes are a different matter. Window-2 aggregate alpha is small in every cluster, because window 2 excludes the 2024 election cycle that carries most of the sample's volume and because the heavy dollar tail rotates across windows, as the 6-of-100 persistence count of Section 4.1 already indicated. Skill persists in direction. The dollar value of that skill is event-driven.

The persistence is sharp enough to predict. For the 37,018 investors active in both halves, window-1 characteristics jointly predict window-2 directional accuracy with an $R^2$ of 0.26 on markets the investor has never traded (Table 7), where window-1 accuracy carries a coefficient of $+0.56$ ($p < 0.001$) and the behavioral features add little beyond it. The same characteristics predict essentially none of the variation in window-2 dollar alpha ($R^2 = 0.002$). Skill, the propensity to trade on the side the market eventually resolves to, is a persistent attribute. The dollar reward to it is not.

[12] These are unweighted investor means. Pooling fills instead, the trade-weighted window-2 accuracies are 60.9 percent for the informed cluster and 41.4 percent for the market-maker cluster, so the persistence conclusion is unchanged under either weighting.

A final check removes the outcome from the classification entirely. Panel A of Table 6 assigns types on window-1 behavioral footprints alone: maker share, two-sidedness, and gross volume.[13] The behaviorally defined aggressive directional traders (taker-dominant, one-sided, large) average below-coin-flip window-2 accuracy (47.1 percent across investors), so aggression and size do not mark information for the typical address. Their trade-weighted accuracy of 86.8 percent shows that only their few largest members are informed, the same thin-tail structure the rest of the paper documents. The behavior alone does not identify the information type. The outcome does, and the type it identifies persists. The aggregate liquidity-provision loss of Section 4.2 is not uniform across suppliers. The behaviorally defined suppliers approximately break even at resolution on their resting fills (+$3.8 million on 44 percent fill accuracy), while the outcome-based market-maker cluster loses on its resting fills too (−$36.8 million), Figure 6.[14]

## 5.5 Cross-Category Replication

The typology is identified within Politics, for the reasons of Section 2.3. If it captures a generic feature of decentralized prediction markets rather than a Politics-specific artifact, it should replicate where the information environment is different. We rerun the decile-cut classification on the two largest non-Politics categories: Sports, where information arrives continuously through a game, and Crypto & Web3, where prices update against external blockchain-data oracles.

Table 8 reports the three categories side by side, and the partition is what replicates. The market-maker cluster has sub-50 percent accuracy and negative aggregate alpha in every category (−$101 million in Politics, −$35 million in Sports, −$12 million in Crypto), and the informed cluster has above-70 percent accuracy and positive aggregate alpha in every category (+$78, +$46, and +$11 million).

[13] Maker share is the fraction of an investor's fills made as the resting rather than the aggressing side. Two-sidedness is defined in Section 5.1.

[14] Maker-side aggregates apply the trade-level objects to the resting side of each user-versus-user fill, where the maker's direction is the negative of the taker-signed size. Taker-order summary rows are excluded as duplicates whose maker field holds the aggressor (Appendix B).

The noise cluster holds about 92 to 93 percent of investors throughout, with a small net alpha (+$10 million in Politics, +$4 million in Crypto, +$27 million in Sports). Cluster sizes and dollar magnitudes shift across categories, the magnitudes smaller where volume is lower, consistent with the cycle-driven nature of the rents. The signs of the market-maker and informed aggregates do not shift. The partition is a generic feature of these venues, not an artifact of the Politics subsample.

## 5.6 Luck versus Skill

A cross-section of 497,588 investors with a median of 23 trades each throws off extreme alpha realizations by chance alone, so the question is not whether large alphas exist but whether they exceed what luck produces. Following Fama and French (2010), we simulate the alpha cross-section under a no-skill null. Each market's binary outcome is redrawn as an independent Bernoulli whose success probability is the winner's last pre-terminal price, and the full cross-section of per-investor alpha and $t(\alpha_i)$ is recomputed on each of 1,000 draws, holding positions and the calibration benchmark fixed. Here $t(\alpha_i)$ is investor $i$'s mean per-market alpha divided by its standard error across the markets she traded, so an investor with many small consistent wins scores above one with a single lucky large position. The design follows Kosowski et al. (2006) and Fama and French (2010), and Harvey and Liu (2022) show it is conservative about detecting skill, which works against the tail rejections we report.

The actual $t(\alpha_i)$ distribution has fatter tails than the null on both sides (Figure 7). At the 99th percentile the actual $t$ is 4.33 against a simulated mean of 4.15 (95% of draws within $[4.01, 4.28]$, luck-adjusted $p = 0.002$), and at the 99.9th percentile 13.27 against 12.71 ($p = 0.004$). At the other extreme the actual 1st percentile is $-4.74$ against $-4.39$ ($p = 0.001$): the worst investors lose more reliably than luck allows, the mirror image of the persistent 40 percent market-maker accuracy of Section 5.4. Between the tails the distribution is indistinguishable from the null. Skill lives at both extremes of the cross-section and nowhere in its middle.

Dollar alpha percentiles, by contrast, are not more extreme than the null anywhere in the right tail, partly because the in-sample calibration benchmark absorbs a share of the outcomes it was estimated on. The normalized $t(\alpha_i)$ is the appropriate skill metric, and headline dollar magnitudes alone are not evidence of skill. The simulation covers the 289,852 active investors with at least five markets (421,353 for the dollar percentiles), and the 3 percent with degenerate within-investor variance drop from the $t$ percentiles in both the actual and the simulated cross-sections.

The in-sample calibration benchmark is the last concern, and a hold-out version closes it. Splitting markets in half by a deterministic hash, estimating $\mu(p, c)$ on each half, and recomputing every Politics trade's alpha against the benchmark from the other half moves the cluster aggregates away from zero, not toward it. The market-maker aggregate goes from $-\$97$ million in-sample to $-\$120$ million out, and the informed aggregate from $+\$72$ to $+\$93$ million, with an address-level Spearman of $+0.72$ between the two. The in-sample benchmark absorbs part of the realized outcomes and shrinks measured skill, so the headline estimates are conservative on this dimension.

# 6 Conclusion

Who drives price discovery in prediction markets? A thin and rotating minority. On Polymarket, 1.9 percent of active investors supply a quarter of aggregate corrective impact and earn the platform's information rents, almost entirely by anticipating news rather than by moving prices. Their accuracy persists on markets that resolve after the sample on which they were identified. The measurement that supports this answer starts from one observation. The price-impact estimators in three decades of microstructure research all share a structural identification of an unobserved efficient price. Binary prediction contracts make that price observable. We use this to measure each trade's corrective value directly, in dollars, to decompose it exactly into realized price impact and anticipation, and to test the predictions of Kyle (1985),

Glosten and Milgrom (1985), and Hasbrouck (1991, 1995) at scale in a venue where the standard microstructure assumptions are relaxed or testable.

Our methodological contribution is the resolution-anchored measurement framework of Sections 3.2 and 3.3: a mark-to-resolution corrective impact that splits, by identity, into a realized markout and a convergence residual, and a per-print attribution that sums exactly to each contract's total realized convergence. All three objects are model-free, additive, and aggregable to the investor level. Validated against a Hasbrouck structural-VAR information share on 422 markets, the framework agrees at a Spearman rank correlation of +0.40, rising monotonically to +0.47 on the one-day realized markout ($p < 0.001$), and the agreement is strongest exactly where the decomposition claims to isolate realized impact. The bias-adjusted alpha (Section 3.4) pairs the impact measures with a Jensen (1968)-style benchmark that nets out the favorite-longshot calibration spread (Snowberg and Wolfers 2010).

The setting also disciplines what the classical predictions can mean. Glosten-Milgrom losses arise here without designated intermediaries or posted spreads: adverse selection finds the liquidity supplier even when supply is an emergent role rather than an assigned one. Kyle-style information rents arrive almost entirely through anticipation rather than contemporaneous impact, because on a venue where prices move mainly on news, the information channel is position-taking ahead of resolution rather than order-flow pressure. The impact channel is real but small, and the structural VAR sees the same thing from the other direction. Market makers are paid on the opposite horizon. They earn the spread on contact and surrender it as the same news arrives, the dealer markout reversal. The informed are paid late and the makers early, for the two halves of the same convergence. The dollar value of both rents is event-driven, concentrated in whichever cycle is being resolved, while the partition itself, and the accuracy that defines it, persists across windows and categories.

Several limitations bound the results and point to future work. The investor identifier we observe is an on-chain address, not a natural-person identifier. A single operator may control multiple addresses, and ruling that out requires an off-platform transfer-graph backfill we leave to future work. The empirical analysis is restricted to a single category. The cross-category replication in Section 5.5 supports external validity on two other large categories, but a full cross-category re-estimation is the natural next step. The market-maker loss we report is a mark-to-resolution trading return. It excludes the liquidity-provider rewards and token-airdrop expectations the venue extends to high-volume resting liquidity, so liquidity provision on the platform may break even or profit once those off-measure incentives are counted. The resolution value we treat as ground truth is itself a potential target. A trader who can corrupt the real-world resolution source, by tampering with the weather station behind a temperature market or by gaming a disputed oracle, breaks the outcome anchor for the affected markets, and screening such cases at scale is beyond our data. The Hasbrouck VAR validation is run on the cross-category top-500 markets sample, although a Politics-only re-estimation strengthens the rank agreement to $+0.44$ on 398 markets.

The findings carry two implications beyond the academic literature. The first concerns platform design. Polymarket and its competitors run liquidity-incentive programs that reward resting depth, which treats liquidity supply as the scarce input. Our results locate the scarce input elsewhere. Liquidity supply clears competitively, so the binding constraint is the accuracy that makes prices converge. That accuracy lives in a small informed class whose membership rotates with the news cycle: only 6 of the top-100 impact contributors in the first half of the sample reappear in the second. A venue that wants calibrated prices in a new event domain cannot rely on a standing population of generalist forecasters to show up, because the specialists it needs are domain-specific and largely not yet on the platform. That shifts the design margin from rewarding tenure and depth toward lowering the entry cost of informed capital in each new domain.

The second concerns the growing class of consumers who read prediction-market prices as probability forecasts. The informed minority earns its rents through anticipation, by positioning ahead of news that later proves it right, rather than through price impact on contact. A Polymarket price therefore embeds positioned beliefs about news that has not yet arrived, which is precisely what makes it valuable as a forecast and what the venue's calibration record reflects. The same evidence counsels against reading short price movements as information: most contemporaneous order-flow pressure carries little corrective content, and the convergence that validates the informed positions arrives with the news itself. Forecast consumers should weight the level of the price and discount its high-frequency movements.

# Tables

## Table 1: Summary statistics

Panel A summarizes the per-trade distribution across the 34.1 million Politics post-filter trades that survive the trade-level filters of Section 2.3: the exclusion of settlement-contract and self-trade rows, and the short-time-to-resolution, post-resolution, and $1,000 minimum-volume filters. Panel B summarizes the investor-level aggregates across active investors with at least ten post-filter trades. Panel C summarizes per-market aggregates over the markets with at least one post-filter trade. The duration row covers the subset of those markets that carry a recorded start block. Token-quantity units are $10^{-6}$. Price units are USD per outcome token. Volume and corrective-impact aggregates are in USD. The table summarizes the core sample-level quantities. Derived decomposition and inference objects (markouts, contributions, behavioral features, test statistics) are defined in Table A.1 and summarized where they are used. Variable definitions: see Table A.1.

| **Variable** | $N$ | Mean | SD | Min | p25 | p50 | p75 | Max |
|---|---|---|---|---|---|---|---|---|
| *Panel A: Per trade ($N$ = 34,134,106)* | | | | | | | | |
| $\lvert s_j \rvert$ (tokens) | 34.1 M | 394.5 | 4,906 | 0.00 | 10.9 | 39.4 | 139.5 | 6.22 M |
| $p_j$ (execution price) | 34.1 M | 0.40 | 0.37 | 0.00 | 0.02 | 0.33 | 0.74 | 1.00 |
| $p_j^-$ (pre-trade mid) | 34.1 M | 0.40 | 0.37 | 0.00 | 0.02 | 0.33 | 0.74 | 1.00 |
| $\mathrm{CI}_j$(EOM) (USD) | 34.1 M | −0.42 | 913 | −0.62 M | −2.25 | 0.05 | 2.46 | 2.48 M |
| $\lvert \mathrm{CI}_j \rvert$(EOM)$\rvert$ (USD) | 34.1 M | 58.2 | 911 | 0.00 | 0.37 | 2.38 | 14.20 | 2.48 M |
| *Panel B: Per active investor ($N$ = 497,588)* | | | | | | | | |
| $n_i$ (post-filter trades) | 497.6 k | 64.6 | 1,800 | 10 | 14 | 23 | 42 | 0.73 M |
| $V_i$ (cumulative volume, USD) | 497.6 k | 8,127 | 148,893 | 0.00 | 126 | 536 | 2,806 | 39.5 M |
| $\mathrm{CI}_i$ (USD) | 497.6 k | −12.8 | 27,752 | −3.2 M | −13.3 | −0.15 | 5.1 | 11.3 M |
| $\lvert \mathrm{CI}_i \rvert$ (USD) | 497.6 k | 3,871 | 100,268 | 0.00 | 21.6 | 129.0 | 695.8 | 27.0 M |
| $\alpha_i$ (USD) | 488.1 k | −27.4 | 26,185 | −3.4 M | −13.4 | −0.02 | 7.9 | 11.6 M |
| Realized $\mathrm{PnL}_i$ (USD) | 488.1 k | −44.4 | 27,976 | −3.2 M | −15.3 | −0.98 | 2.6 | 11.3 M |
| $r_i$ (directional accuracy) | 497.6 k | 0.51 | 0.21 | 0.00 | 0.43 | 0.50 | 0.59 | 1.00 |
| *Panel C: Per resolved Politics market* | | | | | | | | |
| $\mathrm{Volume}_m$ (USD) | 16.2 k | 1,075,056 | 16.37 M | 1,002 | 14,318 | 51,198 | 228,567 | 1.53 B |
| $\mathrm{Duration}_m$ (days) | 14.7 k | 54.8 | 82.6 | 1.0 | 6.3 | 17.3 | 61.5 | 935.3 |
| Trades per market | 16.2 k | 2,102 | 25,687 | 1 | 84 | 297 | 1,171 | 2.49 M |

# Table 2: Trader typology

Each active investor with at least ten post-filter trades in the Politics & Elections subsample is assigned to one of three types via the decile cuts $\big(V_i \geq P_{90}(V)$ and $r_i < 0.50\big)$ for Market Maker, $\big(\overline{|\mathrm{CI}|}_i \geq P_{90}$ and $r_i > 0.60\big)$ for Informed, and the residual for Noise. Columns: $N$ is the cluster size. Share is the cluster share of the $N = 497{,}588$ active investors. Mean $r_i$ is the cluster-mean directional accuracy. $\sum |\mathrm{CI}_i|$ is cluster-aggregate absolute corrective impact in million USD. $\sum \alpha_i$ is cluster-aggregate bias-adjusted alpha in million USD. $\sum \mathrm{PnL}_i$ is cluster-aggregate realized PnL in million USD. Mean $\alpha_i$ reports the cluster-mean per-investor alpha in USD. The threshold-sensitivity sweep is reported in Table 5. Variable definitions: see Table A.1.

| **Type** | $N$ investors | Share (%) | Mean $r_i$ (%) | $\sum \lvert\mathrm{CI}_i\rvert$ ($M) | $\sum \alpha_i$ ($M) | $\sum \mathrm{PnL}_i$ ($M) | Mean $\alpha_i$ ($) |
|---|---|---|---|---|---|---|---|
| Market Maker (Glosten–Milgrom) | 26,122 | 5.2 | 30.4 | 603 | −101 | −132 | −3,856 |
| Informed (Kyle) | 9,208 | 1.9 | 76.3 | 454 | +78 | +99 | +8,433 |
| Noise (Kyle–Black) | 462,258 | 92.9 | 51.6 | 870 | +10 | +11 | +21 |
| Total | 497,588 | 100 | | 1,926 | −13 | −22 | |

## Table 3: Markout decomposition by trader type

Each cluster's aggregate corrective impact $\sum \mathrm{CI}_i$ on the post-filter Politics & Elections sample is decomposed at four markout horizons via the exact identity $\mathrm{CI} = \mathrm{RCI}(h) + \mathrm{CR}(h)$ of Equation (2). Clusters are the decile-cut types of Section 4.2 (cluster assignment is unchanged). Columns report the cluster-aggregate realized markout $\sum \mathrm{RCI}_i(h)$ at each horizon and the one-day convergence residual $\sum \mathrm{CR}_i(1\mathrm{d})$, in million USD. Square brackets under the one-minute and one-day markouts are 95% cluster-bootstrap intervals from resampling markets with replacement (500 iterations) on the per-(address, market) aggregates. Variable definitions: see Table A.1.

| **Type** | $\sum$ CI ($M) | $\sum$ RCI(1m) ($M) | $\sum$ RCI(5m) ($M) | $\sum$ RCI(1h) ($M) | $\sum$ RCI(1d) ($M) | $\sum$ CR(1d) ($M) |
|---|---|---|---|---|---|---|
| Market Maker | −126.8 | +3.9 | +3.1 | −0.6 | −6.7 | −120.1 |
| | | [+3.2, +4.6] | | | [−10.1, −3.9] | |
| Informed | +102.3 | +2.3 | +2.5 | +2.4 | +6.6 | +95.6 |
| | | [+1.9, +2.8] | | | [+4.3, +10.1] | |
| Noise | +18.1 | +7.4 | +8.1 | +8.2 | +10.8 | +7.3 |
| | | [+6.9, +8.0] | | | [+8.4, +13.5] | |

# Table 4: Investor-level regression of alpha on corrective impact

Cross-sectional OLS over the $N = 488{,}083$ active investors in the Politics & Elections subsample. Dependent variable is per-investor bias-adjusted alpha $\alpha_i$ (USD). Regressors are per-investor corrective impact $\mathrm{CI}_i$ (USD), log cumulative trade volume $\log V_i$ centered to its sample mean, the interaction $\mathrm{CI}_i \times \log V_i$, and directional accuracy $r_i$. The $\log V_i$ row is reported scaled by $10^5$ and the $r_i$ and behavioral-control rows by $10^3$ for legibility. Columns (1)–(4) report HC0-robust standard errors. Column (5) adds modal-resolution-month fixed effects (each investor's modal resolution-time bucket, in 1.3-million-block bins of approximately one calendar month, across the 2023–2026 sample, 34 buckets) with modal-month-clustered standard errors. Column (6) adds the ex-ante behavioral controls of Section 5.4: maker share, two-sidedness, and log gross volume centered to its sample mean. The Wald row reports the $p$-value for the test of $H_0 : \beta_{\mathrm{CI}} = 1$ against the two-sided alternative. Variable definitions: see Table A.1. Significance: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

| | (1) $\alpha$ on CI | (2) +log $V$ | (3) +CI×log $V$ | (4) +accuracy $r_i$ | (5) +month FE | (6) +behavioral |
|---|---|---|---|---|---|---|
| $\mathrm{CI}_i$ | $+0.913^{***}$ | $+0.913^{***}$ | $+0.640^{***}$ | $+0.641^{***}$ | $+0.641^{***}$ | $+0.641^{***}$ |
| | (0.054) | (0.054) | (0.187) | (0.188) | (0.109) | (0.109) |
| $\log V_i$ (centered, $\times 10^5$) | | $-2.634^{**}$ | $-3.681^{***}$ | $-4.005^{***}$ | $-4.326^{***}$ | $+7.773^{**}$ |
| | | (1.254) | (1.210) | (1.145) | (0.933) | (3.103) |
| $\mathrm{CI}_i \times \log V_i$ | | | $+0.030$ | $+0.029$ | $+0.029^{***}$ | $+0.030^{***}$ |
| | | | (0.024) | (0.024) | (0.011) | (0.011) |
| $r_i$ (directional accuracy, $\times 10^3$) | | | | $-0.269$ | $-0.271^{***}$ | $-0.273^{***}$ |
| | | | | (0.214) | (0.076) | (0.069) |
| Maker share ($\times 10^3$) | | | | | | $+0.167^{***}$ |
| | | | | | | (0.062) |
| Two-sidedness ($\times 10^3$) | | | | | | $+0.107^{*}$ |
| | | | | | | (0.056) |
| Log gross volume (centered, $\times 10^3$) | | | | | | $-0.124^{***}$ |
| | | | | | | (0.038) |
| Wald $p$-value ($H_0 : \beta_{\mathrm{CI}} = 1$) | 0.106 | 0.106 | 0.053 | 0.057 | 0.001 | 0.001 |
| Modal-month FE | no | no | no | no | yes | yes |
| Cluster SEs (modal month) | no | no | no | no | yes | yes |
| $N$ | 488,083 | 488,083 | 488,083 | 488,083 | 488,083 | 488,083 |
| $R^2$ | 0.955 | 0.955 | 0.956 | 0.956 | 0.956 | 0.956 |

# Table 5: Typology sensitivity to thresholds

Each row reports cluster size (Share) and aggregate bias-adjusted alpha ($\sum \alpha$, million USD) for one combination of decile-cutoff percentile $q$, market-maker accuracy ceiling $r_{\text{MM}}$, and informed accuracy floor $r_{\text{Inf}}$. The default specification used throughout the body (Panel B, row in bold) is $q = 0.90$, $r_{\text{MM}} < 0.50$, $r_{\text{Inf}} > 0.60$. Variable definitions: see Table A.1.

| **Cutoffs** | | | **Market Maker** | | **Informed** | | **Noise** | |
|---|---|---|---|---|---|---|---|---|
| $q$ | $r_{\text{MM}} <$ | $r_{\text{Inf}} >$ | Share (%) | $\sum \alpha$ ($M) | Share (%) | $\sum \alpha$ ($M) | Share (%) | $\sum \alpha$ ($M) |
| *Panel A: Looser percentile cutoff* $q = 0.75$ | | | | | | | | |
| 0.75 | 0.45 | 0.55 | 7.8 | −103.8 | 6.2 | +97.7 | 85.9 | −7.3 |
| 0.75 | 0.50 | 0.60 | 11.7 | −109.1 | 4.2 | +84.2 | 84.1 | +11.5 |
| 0.75 | 0.55 | 0.65 | 17.5 | −104.6 | 3.1 | +74.8 | 79.4 | +16.4 |
| *Panel B: Default percentile cutoff* $q = 0.90$ *(main specification)* | | | | | | | | |
| 0.90 | 0.45 | 0.55 | 3.8 | −95.5 | 2.7 | +89.7 | 93.4 | −7.7 |
| **0.90** | **0.50** | **0.60** | **5.2** | **−100.7** | **1.9** | **+77.7** | **92.9** | **+9.7** |
| 0.90 | 0.55 | 0.65 | 7.3 | −97.2 | 1.4 | +69.4 | 91.3 | +14.3 |
| *Panel C: Tighter percentile cutoff* $q = 0.95$ | | | | | | | | |
| 0.95 | 0.45 | 0.55 | 2.3 | −88.6 | 1.6 | +82.3 | 96.1 | −7.1 |
| 0.95 | 0.50 | 0.60 | 2.8 | −93.8 | 1.2 | +71.8 | 96.0 | +8.5 |
| 0.95 | 0.55 | 0.65 | 3.5 | −90.5 | 0.9 | +65.1 | 95.6 | +12.0 |

# Table 6: Out-of-sample typology evaluation

Politics & Elections markets are split at the median resolution block, so each market's outcome belongs to exactly one window. Panel A classifies addresses on window-1 *behavioral* features that never use the realized outcome (maker share $\geq 0.5$, two-sidedness $\geq 0.5$, and gross volume $\geq P_{90}$ for Market Maker. Maker share $< 0.5$, two-sidedness $\leq 0.25$, and gross volume $\geq P_{90}$ for Informed. The residual is Noise.) Panel B classifies addresses on window-1 *realized* features using the decile cuts of Section 4.2. Both classifications are then evaluated on window-2 markets only, for addresses with at least ten post-filter window-2 trades. Columns report the cluster size in the evaluation sample, mean window-2 directional accuracy (the unweighted mean across investors, with trade-weighted accuracies reported in the text), and window-2 aggregates of bias-adjusted alpha, corrective impact, and the one-day realized markout, in million USD. The evaluation panel is taker-side, so Panel A's Market Maker row covers those addresses' aggressive trades only. Variable definitions: see Table A.1.

| **Type** | $N$ | Mean $r_i$ (w2) (%) | $\sum \alpha_i$ (w2) ($M) | $\sum \mathrm{CI}_i$ (w2) ($M) | $\sum \mathrm{RCI}_i$(1d) (w2) ($M) |
|---|---|---|---|---|---|
| *Panel A: classified on window-1 behavior (outcome-free)* | | | | | |
| Market Maker | 1,591 | 55.6 | +1.3 | +1.8 | +2.1 |
| Informed | 1,424 | 47.1 | +0.9 | +2.0 | +0.2 |
| Noise | 38,085 | 52.7 | +1.8 | +3.5 | +2.7 |
| *Panel B: classified on window-1 realized accuracy and impact* | | | | | |
| Market Maker | 3,371 | 40.3 | +0.4 | −0.1 | +0.8 |
| Informed | 921 | 65.1 | +0.5 | +2.1 | +0.6 |
| Noise | 33,301 | 53.6 | +2.6 | +4.8 | +3.4 |

# Table 7: Predicting out-of-sample alpha and accuracy

Predictive OLS over the $N = 37{,}018$ investors active in both halves of the split sample of Section 5.4, with at least ten post-filter trades in each window. Window-1 characteristics predict window-2 bias-adjusted alpha (left column, USD thousands) and window-2 directional accuracy (right column) on disjoint market populations. Regressors are window-1 corrective impact $\mathrm{CI}_i$ (USD millions), log cumulative trade volume $\log V_i$ centered to its sample mean, directional accuracy $r_i$, maker share, and two-sidedness. Standard errors are HC0 robust. Variable definitions: see Table A.1. Significance: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

| | DV: $\alpha_i$ (window 2, $\times 10^3$) | DV: $r_i$ (window 2) |
|---|---|---|
| $\mathrm{CI}_i$ (window 1) | $+15.335$ | $-0.237^{***}$ |
| | (20.395) | (0.085) |
| $\log V_i$ (window 1, centered) | $+0.082$ | $-0.0032^{***}$ |
| | (0.068) | (0.0004) |
| $r_i$ (window 1) | $-0.072$ | $+0.559^{***}$ |
| | (0.538) | (0.006) |
| Maker share (window 1) | $+0.055$ | $+0.010^{**}$ |
| | (0.228) | (0.005) |
| Two-sidedness (window 1) | $+0.007$ | $+0.016^{***}$ |
| | (0.201) | (0.004) |
| $N$ | 37,018 | 37,018 |
| $R^2$ | 0.002 | 0.262 |

# Table 8: Typology replication across categories

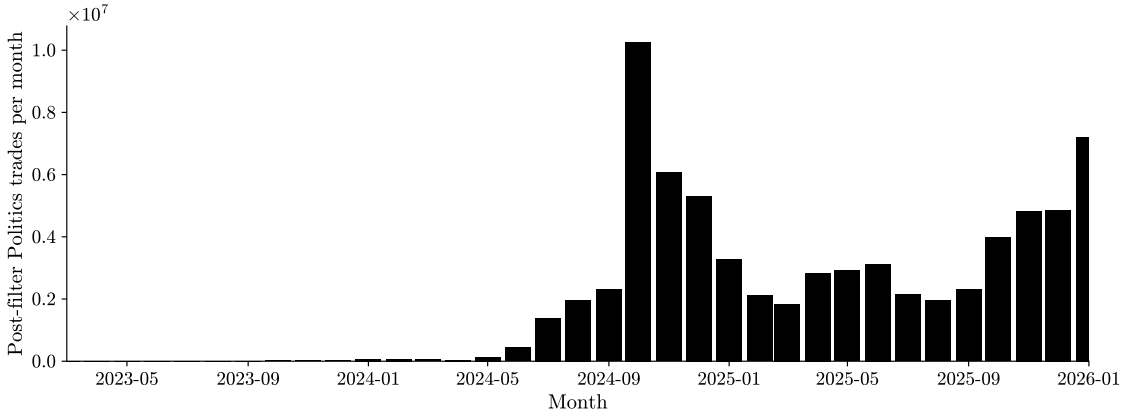
Each row reports the decile-cut classification of Section 4.2 ($V_i \geq P_{90}$ and $r_i < 0.50$ for Market Maker, $\overline{|\mathrm{CI}|}_i \geq P_{90}$ and $r_i > 0.60$ for Informed, and the residual for Noise) on the indicated category subsample, with active investors defined as those with at least ten post-filter trades in that category. Columns: $N$ is the cluster size. Share is the cluster share of active investors. Mean $r_i$ is the cluster-mean directional accuracy (percent). $\sum |\mathrm{CI}_i|$ is cluster-aggregate absolute corrective impact (million USD). $\sum \alpha_i$ is cluster-aggregate bias-adjusted alpha (million USD). Variable definitions: see Table A.1.

| **Type** | $N$ | Share (%) | Mean $r_i$ (%) | $\sum \lvert\mathrm{CI}_i\rvert$ ($M) | $\sum \alpha_i$ ($M) |
|---|---|---|---|---|---|
| *Panel A: Politics & Elections (main sample, $N = 497{,}588$ active investors)* | | | | | |
| Market Maker | 26,122 | 5.2 | 30.4 | 603 | −101 |
| Informed | 9,208 | 1.9 | 76.3 | 454 | +78 |
| Noise | 462,258 | 92.9 | 51.6 | 870 | +10 |
| *Panel B: Sports ($N = 540{,}679$ active investors)* | | | | | |
| Market Maker | 25,111 | 4.6 | 31.2 | 576 | −35 |
| Informed | 11,472 | 2.1 | 73.6 | 224 | +46 |
| Noise | 504,096 | 93.2 | 50.2 | 412 | +27 |
| *Panel C: Crypto & Web3 ($N = 162{,}488$ active investors)* | | | | | |
| Market Maker | 7,927 | 4.9 | 27.1 | 152 | −12 |
| Informed | 5,072 | 3.1 | 77.6 | 61 | +11 |
| Noise | 149,489 | 92.0 | 73.0 | 160 | +4 |

# Figures

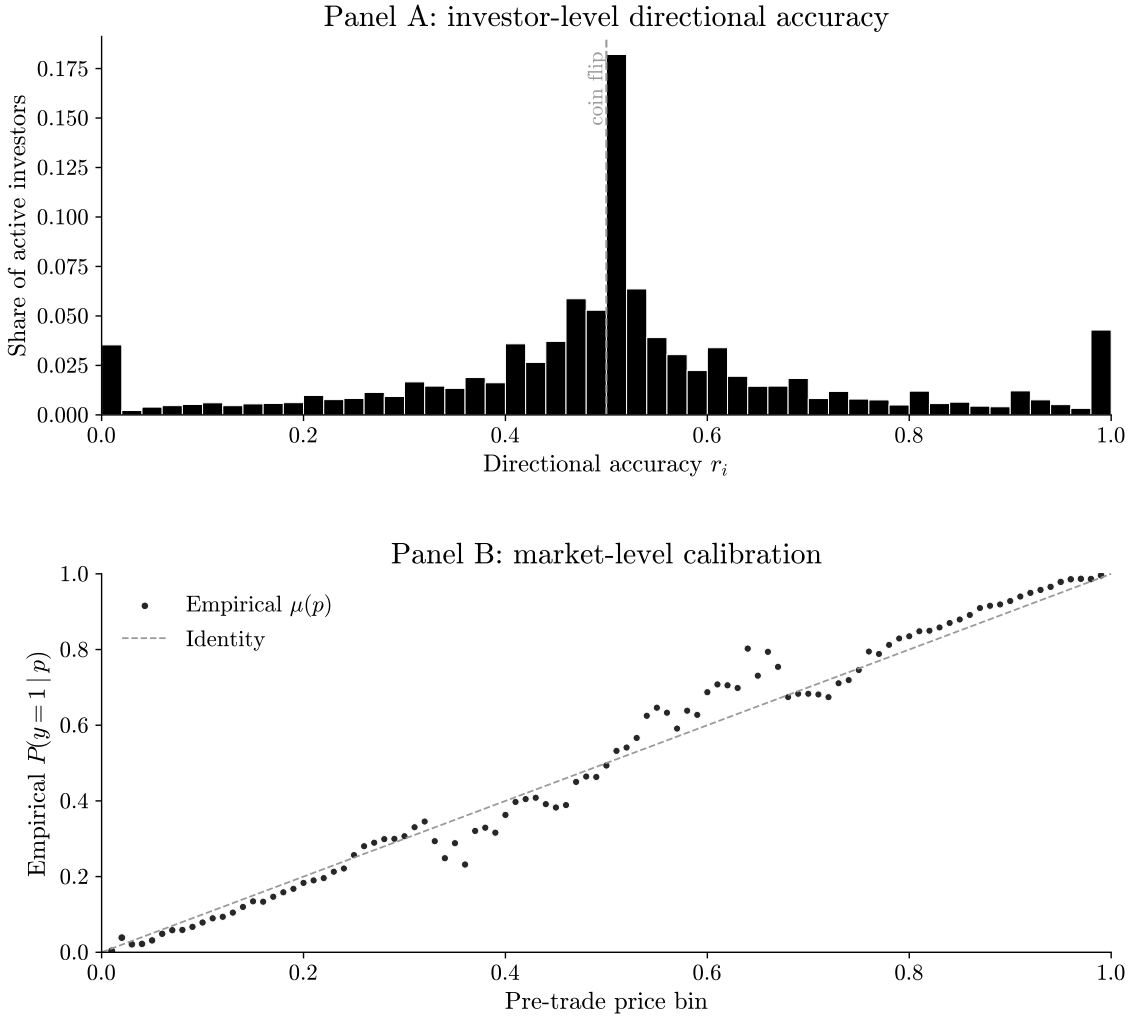
## Figure 1: Monthly trade counts

Monthly post-filter trade count in the Politics & Elections category, March 2023 through January 2026. The vertical axis is trades per month. Variable definitions: see Table A.1.



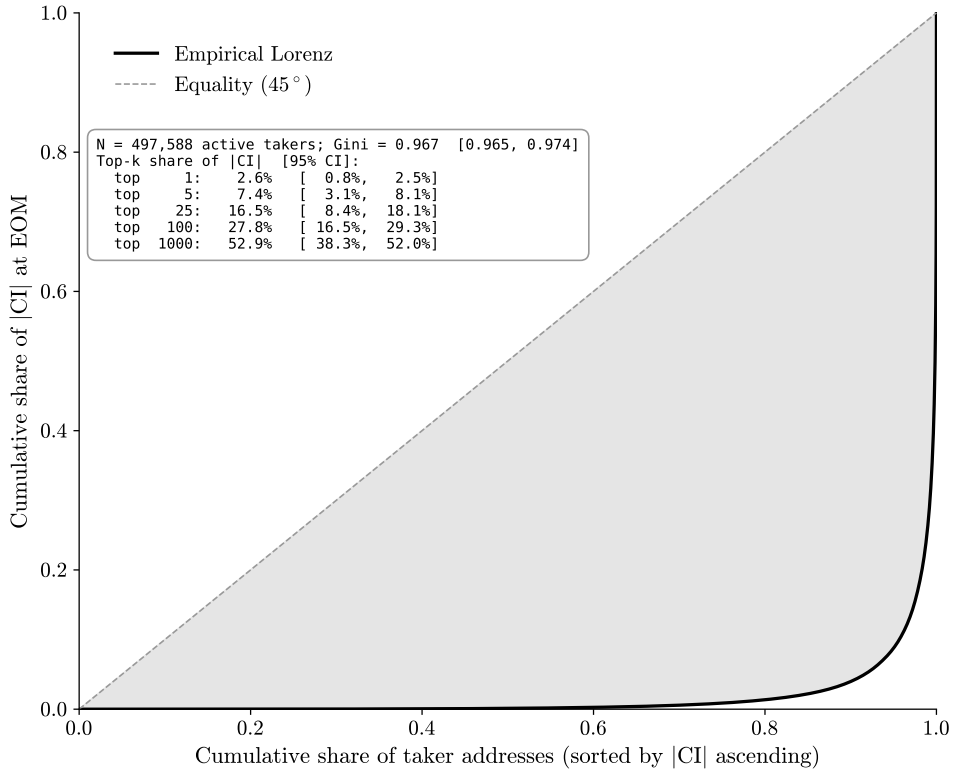
# Figure 2: Investor accuracy and market calibration

Panel A: distribution of per-investor directional accuracy $r_i$, the share of an investor's post-filter trades whose signed direction matched the realized outcome, across the $N = 497{,}588$ active Politics investors with at least ten post-filter trades. The dashed line marks the 0.50 coin-flip benchmark, and the bars at 0 and 1 are low-trade accounts whose accuracy is mechanically extreme. Panel B: market-level calibration curve, the empirical resolution rate against the pre-trade price bin, overlaid with the 45° identity and sized by per-bin trade count. Variable definitions: see Table A.1.



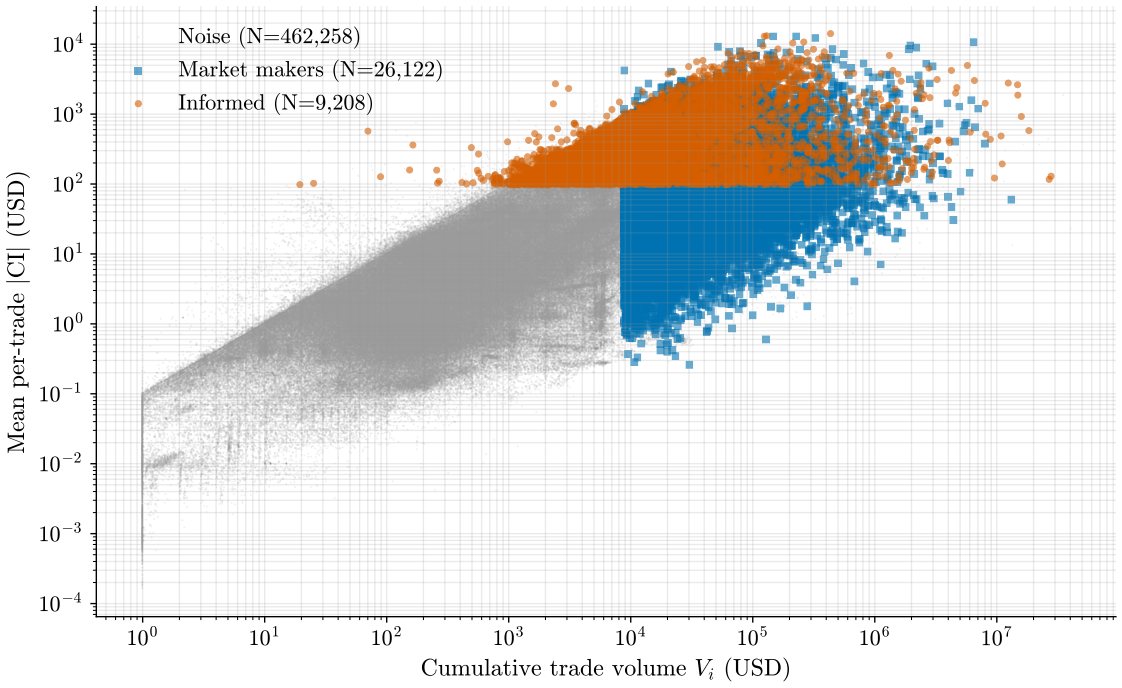
# Figure 3: Concentration of corrective impact across investors

Active investors trade at least ten Politics trades after the pathology filters of Section 2.3 and the contract exclusion of Appendix B. The $N$ = 497,588 active investors are sorted by $|\mathrm{CI}_i|$ on the horizontal axis. The dashed line is the 45° equality benchmark. The inset reports the Gini coefficient and cluster-bootstrap top-$k$ shares with 95% confidence intervals. Variable definitions: see Table A.1.



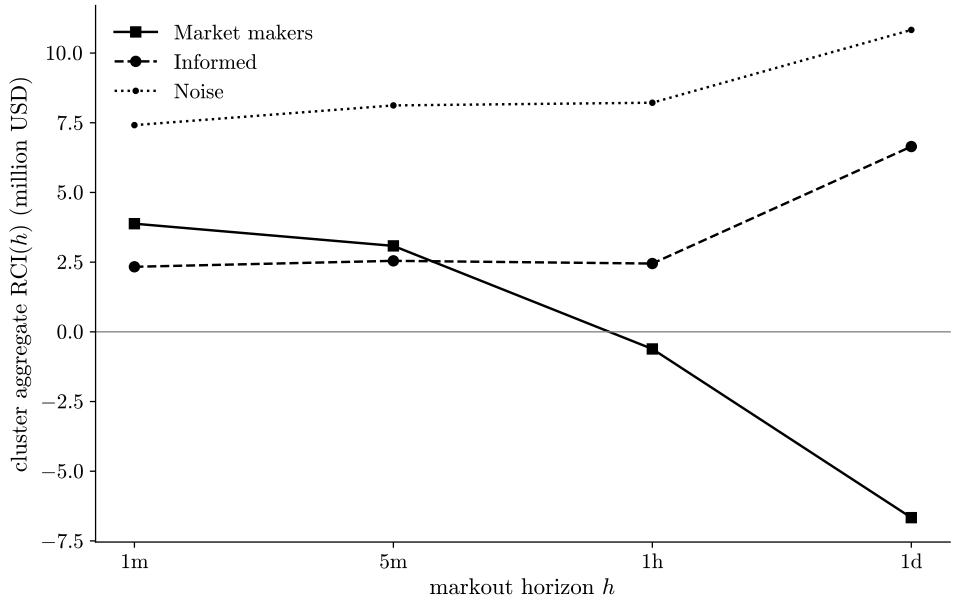
## Figure 4: Trader typology in the volume-impact plane

Each point is one active investor in the Politics subsample plotted on log-log axes of cumulative trade volume $V_i$ versus mean per-trade absolute corrective impact $\overline{|\mathrm{CI}|}_i$. Color and marker shape encode the decile-cut classification of Section 4.2: market makers (blue squares, $V_i \geq P_{90}$ and $r_i < 0.5$), informed traders (vermilion circles, $\overline{|\mathrm{CI}|}_i \geq P_{90}$ and $r_i > 0.6$), and noise traders (gray dots, all remaining active investors). The cluster-level aggregates are reported in Table 2. Variable definitions: see Table A.1.



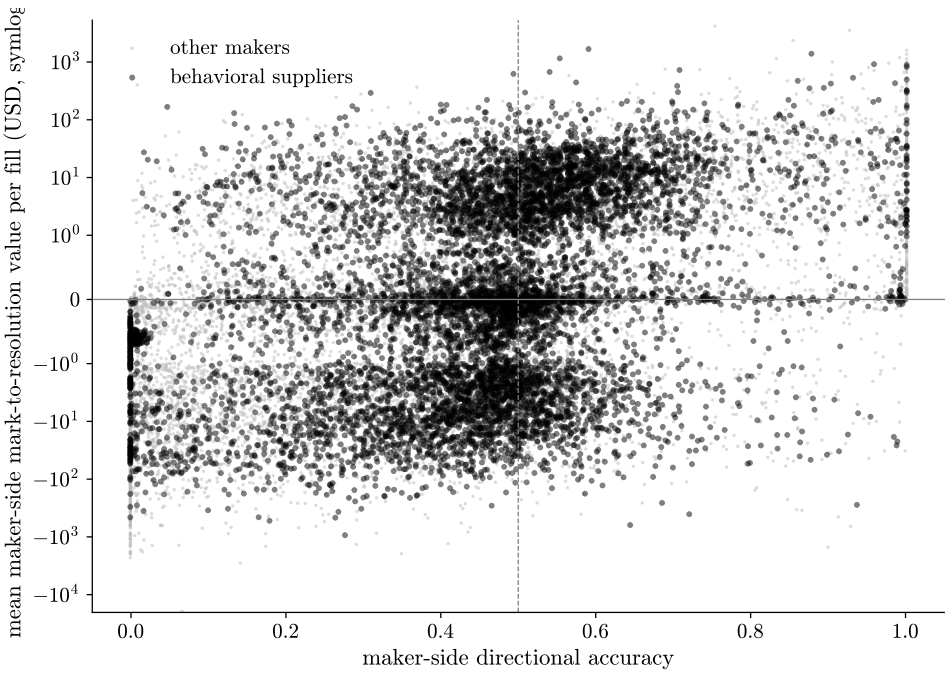
## Figure 5: Markout term structure by trader type

Cluster-aggregate realized corrective impact $\sum \mathrm{RCI}_i(h)$, in million USD, at the one-minute, five-minute, one-hour, and one-day horizons, for the market-maker, informed, and noise clusters of Table 2 on the post-filter Politics & Elections sample. The horizontal line marks zero. The underlying values and their market-bootstrap confidence intervals are reported in Table 3. Variable definitions: see Table A.1.



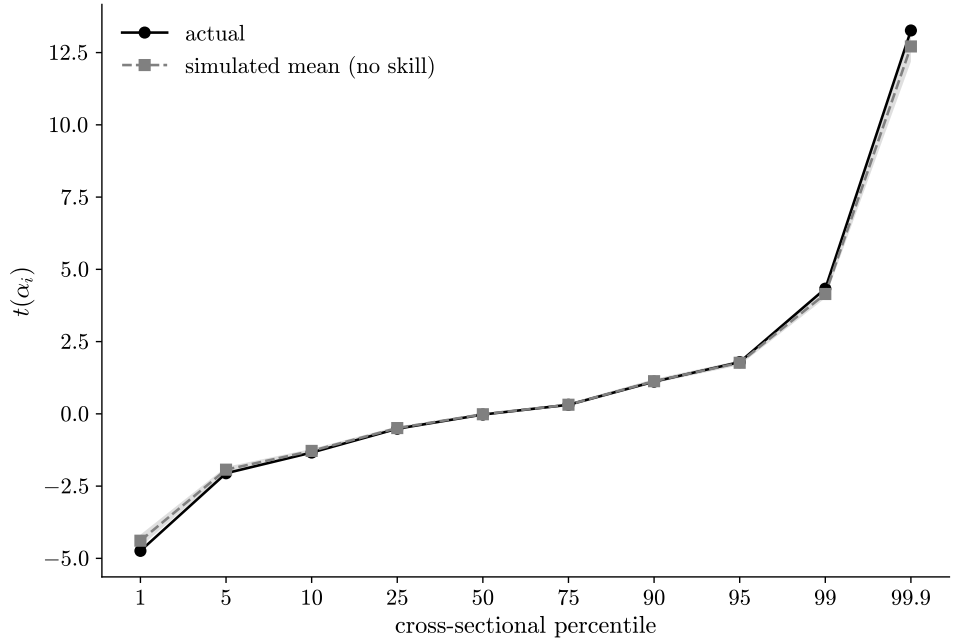
## Figure 6: Maker-side accuracy and value per fill

Each point is an address with at least 100 resting fills on the post-filter Politics & Elections sample. The horizontal axis is the address's maker-side directional accuracy. The vertical axis is its mean maker-side mark-to-resolution value per fill in USD on a symmetric log scale. Black points are the addresses classified market maker by the outcome-free behavioral rule of Section 5.4 (passive, two-sided, large). Gray points are all other makers. The horizontal line marks zero value. The dashed vertical line marks coin-flip accuracy. Variable definitions: see Table A.1.



## Figure 7: Alpha cross-section against the no-skill null

Cross-sectional percentiles of the per-investor $t(\alpha_i)$ on the post-filter Politics & Elections sample (black, circles) against the mean of the same percentiles across 1,000 no-skill simulations (gray, squares), in which each market's outcome is redrawn as an independent Bernoulli with success probability equal to the actual winner's last pre-terminal price. The shaded band is the central 95% of the simulated percentiles. Variable definitions: see Table A.1.



# Appendix

# A Variable Definitions

Master glossary for the variables used in Sections 3 through 5. Sources: "`OrderFilled`" is the on-chain trade event, "Gamma API" the Polymarket metadata service, "Derived" a construction documented in Section 3, and "Aggregated" a sum over the indicated subset of trades.

| Variable | Definition | Source |
|---|---|---|
| *Trade-level (per trade $j$ in market $m$)* | | |
| $s_j$ | Signed taker size in token units of $10^{-6}$. Positive for taker buys, negative for taker sells. | `OrderFilled` |
| $p_j$ | Outcome-token execution price in U.S. dollars per token, $p_j \in (0,1)$. | `OrderFilled` |
| $p_j^-$ | Pre-trade price: the execution price of the most recent prior trade in the same token. | Derived |
| $p_j^+(h)$ | Prevailing price $h$ after trade $j$: the last execution price in $(t_j, t_j + h]$, or $p_j$ when no trade lands in the window. | Derived |
| $y_m$ | Realized binary outcome of trade $j$'s contract at resolution, $y_m \in \{0,1\}$. | Gamma API |
| $\mathrm{CI}_j(\mathrm{EOM})$ | Corrective impact at the end-of-market horizon: $s_j \cdot (y_m - p_j^-)$. Units USD. | Eq. (1) |
| $\mathrm{RCI}_j(h)$ | Realized corrective impact (markout) at horizon $h$: $s_j \cdot (p_j^+(h) - p_j^-)$. Units USD. | Eq. (2) |
| $\mathrm{CR}_j(h)$ | Convergence residual at horizon $h$: $s_j \cdot (y_m - p_j^+(h)) = \mathrm{CI}_j(\mathrm{EOM}) - \mathrm{RCI}_j(h)$. | Eq. (2) |
| $\mathrm{CPC}_j$ | Corrective price contribution of trade $j$'s own print: $(2y_m - 1)(p_j - p_j^-)$, aggressor-attributed. | Eq. (3) |
| $\alpha_j$ | Per-trade bias-adjusted alpha: $s_j \cdot (y_m - \mu(p_j, c_j))$. Units USD. | Eq. (4) |
| $\mu(p, c)$ | Empirical resolution probability for price-bin $p$ in category $c$, estimated on the post-filter resolved markets. | Derived |
| *Investor-level (per active investor $i$ with $n_i \geq 10$ trades)* | | |
| $n_i$ | Number of post-filter trades attributed to investor $i$. | Aggregated |
| $V_i$ | Cumulative trade volume of investor $i$ in USD: $\sum_{j \in \mathcal{J}_i} \lvert s_j \rvert \cdot p_j$. | Aggregated |
| $\mathrm{CI}_i$ | Investor-level signed corrective impact: $\sum_{j \in \mathcal{J}_i} \mathrm{CI}_j(\mathrm{EOM})$. | Aggregated |
| $\lvert\mathrm{CI}_i\rvert$ | Magnitude of $\mathrm{CI}_i$, the headline concentration measure. | Aggregated |
| $\overline{\lvert\mathrm{CI}\rvert}_i$ | Mean per-trade absolute corrective impact: $\sum_{j \in \mathcal{J}_i} \lvert\mathrm{CI}_j(\mathrm{EOM})\rvert / n_i$. | Aggregated |
| $\alpha_i$ | Investor-level bias-adjusted alpha: $\sum_{j \in \mathcal{J}_i} \alpha_j$. | Aggregated |
| $\mathrm{PnL}_i$ | Realized profit and loss: $\sum_{j \in \mathcal{J}_i} s_j \cdot (y_j - p_j)$. Units USD. | Aggregated |
| $r_i$ | Directional accuracy: share of investor $i$'s trades with $\mathrm{CI}_j(\mathrm{EOM}) > 0$. | Aggregated |
| Maker share | Maker fills over total fills of investor $i$. | Aggregated |
| Two-sidedness | One minus net over gross signed token flow per (investor, token), gross-weighted mean. One for round-trip flow, zero for one-directional flow. | Aggregated |
| Gross volume | Sum of $\lvert s_j \rvert \cdot p_j$ over both maker and taker fills of investor $i$, in USD. | Aggregated |
| *Market-level (per resolved Politics market $m$)* | | |
| $T_m$ | Resolution block (end of market). | Gamma API |
| $\tau_j$ | Time-to-resolution at trade $j$ in Polygon blocks: $\tau_j \equiv T_m - t_j$. | Derived |
| $c_m$ | Category label for market $m$. Analyses restrict to $c_m$ = Politics & Elections. | Classifier (Section 2.2) |
| $\mathrm{Vol}_m$ | Cumulative trade volume of market $m$ in USD. | Aggregated |
| *Robustness* | | |
| $\mathrm{IS}_m$ | Hasbrouck (Hasbrouck 1995) information share from a bivariate VAR(1) on (returns, signed flow) at 30-minute buckets. | Appendix C |
| $t(\alpha_i)$ | Per-investor $t$-statistic of alpha: mean per-market alpha over its standard error across investor $i$'s markets (at least five). | Derived |

# B Settlement Contracts

The Polymarket CTF Exchange (`0x4bFb41d5B3570DeFd03C39a9A4D8dE6Bd8B8982E`) and Neg-Risk CTF Exchange (`0xC5d563A36AE78145C45a50134d48A1215220f80a`) are the two on-chain contracts that emit the platform's `OrderFilled` events. Their matching function settles one aggressive taker order against one or more resting maker orders in a single atomic call, and that call emits one event per maker order plus one summary event for the taker order. The deployed contract source fixes the field semantics. Each maker-leg event carries the resting maker's wallet in the `maker` field, the aggressor's wallet in the `taker` field, and amounts from the maker order's perspective. The taker summary event carries the aggressor's wallet in the `maker` field, the exchange contract itself in the `taker` field, and amounts from the taker order's perspective. The summary event fires on every multilateral match, on both exchange contracts and for every match type. In our raw event stream of approximately 404 million `OrderFilled` rows from Polygon inception through the sample cut-off, approximately 165 million rows (41 percent) carry one of the two contracts in the taker field, consistent with an average of 1.3 maker legs per match. The contracts never appear as `maker`, because that field always holds the funding wallet of a signed order, which the exchange cannot produce. We verified both facts in the deployed contract source, against the live Polygon event log for sampled transactions,[15] and exhaustively in the event stream, including that within matched transactions the summary event's `maker` equals the maker legs' `taker` without exception.

Contract-as-taker rows are therefore per-call aggregates of the maker legs, not additional trades, and treating each row as a trade would count every match once per maker leg and once more in aggregate. Two further hazards compound the duplication. First, because the summary row's amounts are encoded from the taker order's perspective, any pipeline that signs trades by the taker-aggressor convention assigns these rows

[15] For example, Polygon transaction `0x134b11de04fa3649900376b90d7b947aecca54de3728d48260340f771bc279ab` settles one taker order against five maker orders and emits six `OrderFilled` events: five maker legs naming the aggressor in the `taker` field, and one summary leg naming that aggressor in the `maker` field and the CTF Exchange in the `taker` field. The decoded chain logs match the dataset rows field for field.

the side opposite to what the named investor actually did, so per-trade corrective impact $\mathrm{CI}_j = s_j \cdot (y_m - p_j^-)$ on these rows is sign-inverted. Second, the wallet in the summary row's `maker` field is the aggressor, so treating these rows as resting liquidity supply misclassifies aggressive fills as maker fills.

Throughout the body of this paper, every address-level aggregation therefore excludes rows with a contract in the taker field, which deduplicates the stream to one row per maker-order fill: each surviving row names the resting maker, the aggressor, and amounts from the maker order's perspective. No investor activity is lost. When two buy orders on complementary outcomes cross, the exchange mints the complete set, the match has no opposing seller, and the aggressor's fill is recorded on the maker leg of the complement token, where buying one outcome equals selling the other. The excluded summary rows remain in the price chain as prints, because on such mint and merge matches they are the only print in the taker-side token, and the attribution of Section 3.3 credits them to the aggressor recorded in their `maker` field. Across all eight categories, the deduplicated post-filter sample retains approximately 87 million trades after the additional pathology filters (Section 2.3), of which the 34.1 million Politics & Elections trades form the main analysis sample.

This event structure is a known feature of Polymarket's data rather than a defect of our indexing. The resulting volume double-counting was documented publicly by Slivkoff (2025), who flags the same two contract addresses, and concurrent academic work decomposes platform volume to separate exchange-equivalent turnover from share minting (Tsang and Yang 2026). Those treatments target aggregate volume. Because our analysis runs at the address level, we additionally re-attribute each summary print to the aggressor named in its `maker` field and exclude it from the resting-liquidity panel. Any investor-level inference on Polymarket from raw `OrderFilled` events must deduplicate the summary legs to avoid double counting, inverted signs, and misclassified liquidity supply.

# C Hasbrouck Information Share

This appendix records the construction and the formula for the market-level information share whose cross-market agreement with |CI| is reported in Section 4.5.

The share is fit per market on a volume-restricted sample. A structural-VAR identification needs enough within-market observations to support a stable fit, and the microstructure literature restricts such fits accordingly: Hasbrouck (1995) uses the 30 Dow Jones stocks, Brogaard et al. (2014) a panel of 120, and Alexander et al. (2025) liquid Uniswap pools. We take the top-500 resolved markets by cumulative trade volume, of which 422 admit at least 30 thirty-minute buckets and yield a clean bivariate VAR(1) fit, a retention rate of 84 percent. For each market we bin trades into 900-block intervals (about 30 minutes on Polygon) and form the series $y_t = (r_t, \mathrm{OF}_t)^\top$, where the bucket return $r_t$ is measured in log-odds following Tsang and Yang (2026) and the signed order flow is $\mathrm{OF}_t = \sum_{j \in t} s_j$. The log-odds transform makes a one-cent move carry comparable information across the bounded price range, where a move near a boundary reflects more evidence than a move near 0.5.

The reduced-form VAR(1) is

$$y_t \;=\; A\, y_{t-1} + u_t, \qquad \Sigma \;=\; \mathrm{Var}(u_t),$$

with innovation covariance $\Sigma$. When the VAR is stable (all eigenvalues of $A$ inside the unit circle), inverting the autoregressive polynomial gives the moving-average representation

$$y_t \;=\; \sum_{k=0}^{\infty} A^k\, u_{t-k},$$

so the cumulative long-run effect of a one-time innovation $u_t$ on the level of $y$ is $\sum_{k=0}^{\infty} A^k = (I - A)^{-1} \equiv \Psi(1)$.

The cumulative return $\sum_k r_{t+k}$ is the price-level analogue of the efficient-price innovation in Hasbrouck (1995). Its long-run response to the innovation vector is the first row of

$\Psi(1)$: a return innovation $u_{r,t}$ moves the long-run price by $\Psi(1)_{r,r}\, u_{r,t}$, and an order-flow innovation $u_{\mathrm{OF},t}$ moves it by $\Psi(1)_{r,\mathrm{OF}}\, u_{\mathrm{OF},t}$. If the two innovations were orthogonal, the long-run return variance would decompose as

$$\mathrm{Var}_\infty \;=\; \Psi(1)^2_{r,r}\,\Sigma_{r,r} \;+\; \Psi(1)^2_{r,\mathrm{OF}}\,\Sigma_{\mathrm{OF},\mathrm{OF}},$$

and the share attributable to order flow is

$$\mathrm{IS}_m \;=\; \frac{\Psi(1)^2_{r,\mathrm{OF}}\,\Sigma_{\mathrm{OF},\mathrm{OF}}}{\Psi(1)^2_{r,r}\,\Sigma_{r,r} + \Psi(1)^2_{r,\mathrm{OF}}\,\Sigma_{\mathrm{OF},\mathrm{OF}}},$$

which is the $\mathrm{IS}_m$ used in the cross-method validation of Section 4.5. In practice the reduced-form innovations are correlated, and Hasbrouck (1995) brackets the share by a Cholesky rotation under both orderings. We report the direct variance-ratio form without the rotation, which treats the off-diagonal covariance as belonging to neither source. The cross-sectional validation uses only the ranking of $\mathrm{IS}_m$ across markets, not its level, so the bracketing convention is second-order for our purpose.

Two implementation notes. First, both series are standardized to unit variance before the OLS fit. The information share is invariant to diagonal rescaling of $y_t$, because both the numerator and the denominator scale by the same factor, but the raw series differ by several orders of magnitude and the rescaling conditions the regression numerically. Second, the formula uses the reduced-form innovation variances directly, so no structural identification of the order-flow shock is required. That agnosticism is what makes the share a useful external validator for the resolution-anchored estimator: the two approaches share no identifying assumption.

The cross-market agreement is robust to these construction choices. Measuring the bucket return in raw price differences rather than log-odds leaves the Spearman correlation between $\mathrm{IS}_m$ and the market-level mean |CI| at +0.31, against the +0.40 in log-odds, so the result does not turn on the return definition. The Pearson correlation is +0.27, below the Spearman because the information-share distribution is heavy-tailed (median

$\mathrm{IS}_m \approx 8 \times 10^{-4}$, 97.5th percentile $\approx 0.05$), the case Spearman is built for. Restricting the fit to the top-500 Politics markets, 398 of which fit, strengthens the agreement to +0.44.

# Statements and Declarations

*Funding.* The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

*Competing interests.* The authors declare no competing interests.

*Data and code availability.* The raw Polymarket `OrderFilled` events used in this paper are public on the Polygon blockchain and can be reconstructed from on-chain logs. The curated trade-level parquet (`trade_features.parquet`) and the analysis code that produces every table and figure in this paper will be made available at a public GitHub repository upon acceptance, in compliance with the journal's code and data sharing policy.

# References


Alexander, C., X. Chen, J. Deng, and Q. Fu. 2025. "Price discovery and efficiency in Uniswap liquidity pools." *J. Futures Mark.* 45 (8): 1023–1048. https://doi.org/10.1002/fut.22593.

Barardehi, Y. H., and D. Bernhardt. 2025. "Revisiting the U-shaped patterns in volatility and price impacts: Novel results using trade-time estimates." *J. Financ. Mark.* 74:100971. https://doi.org/10.1016/j.finmar.2025.100971.

Barber, B. M., Y.-T. Lee, Y.-J. Liu, and T. Odean. 2014. "The cross-section of speculator skill: Evidence from day trading." *J. Financ. Mark.* 18:1–24. https://doi.org/10.1016/j.finmar.2013.05.006.

Barclay, M. J., and T. Hendershott. 2003. "Price discovery and trading after hours." *Rev. Financ. Stud.* 16 (4): 1041–1073. https://doi.org/10.1093/rfs/hhg030.

Barclay, M. J., and J. B. Warner. 1993. "Stealth trading and volatility: Which trades move prices?" *J. Financ. Econ.* 34:281–305. https://doi.org/10.1016/0304-405X(93)90029-B.

Becker, J. 2026. *Prediction market analysis: Trade-level datasets for Polymarket and Kalshi.* Dataset release, GitHub repository.

Berk, J. B., and J. H. van Binsbergen. 2015. "Measuring skill in the mutual fund industry." *J. Financ. Econ.* 118 (1): 1–20. https://doi.org/10.1016/j.jfineco.2015.05.002.

Berk, J. B., and R. C. Green. 2004. "Mutual fund flows and performance in rational markets." *J. Polit. Econ.* 112 (6): 1269–1295. https://doi.org/10.1086/424739.

Black, F. 1986. "Noise." *J. Finance* 41:529–543. https://doi.org/10.1111/j.1540-6261.1986.tb04513.x.

Brogaard, J., T. Hendershott, and R. Riordan. 2014. "High-frequency trading and price discovery." *Rev. Financ. Stud.* 27:2267–2306. https://doi.org/10.1093/rfs/hhu032.

Carhart, M. M. 1997. "On persistence in mutual fund performance." *J. Finance* 52:57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Collin-Dufresne, P., and V. Fos. 2015. "Do prices reveal the presence of informed trading?" *J. Finance* 70 (4): 1555–1582. https://doi.org/10.1111/jofi.12260.

Dubach, P. D. 2026. "The anatomy of a decentralized prediction market: Microstructure evidence from the Polymarket order book." ArXiv:2604.24366.

Fama, E. F., and K. R. French. 2010. "Luck versus skill in the cross-section of mutual fund returns." *J. Finance* 65 (5): 1915–1947. https://doi.org/10.1111/j.1540-6261.2010.01598.x.

Forsythe, R., F. Nelson, G. R. Neumann, and J. Wright. 1992. "Anatomy of an experimental political stock market." *Am. Econ. Rev.* 82:1142–1161.

Glosten, L. R., and P. R. Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *J. Financ. Econ.* 14:71–100. https://doi.org/10.1016/0304-405X(85)90044-3.

Gomez-Cram, R., Y. Guo, T. I. Jensen, and H. Kung. 2026. "Prediction market accuracy: Crowd wisdom or informed minority?" SSRN working paper 6617059.

Grossman, S. J., and J. E. Stiglitz. 1980. "On the impossibility of informationally efficient markets." *Am. Econ. Rev.* 70:393–408.

Harvey, C. R., and Y. Liu. 2022. "Luck versus skill in the cross section of mutual fund returns: Reexamining the evidence." *J. Finance* 77 (3): 1921–1966. https://doi.org/10.1111/jofi.13123.

Hasbrouck, J. 1991. "Measuring the information content of stock trades." *J. Finance* 46:179–207. https://doi.org/10.1111/j.1540-6261.1991.tb03749.x.

———. 1995. "One security, many markets: Determining the contributions to price discovery." *J. Finance* 50:1175–1199. https://doi.org/10.1111/j.1540-6261.1995.tb04054.x.

Jensen, M. C. 1968. "The performance of mutual funds in the period 1945–1964." *J. Finance* 23:389–416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x.

Kacperczyk, M., and E. S. Pagnotta. 2019. "Chasing private information." *Rev. Financ. Stud.* 32 (12): 4997–5047. https://doi.org/10.1093/rfs/hhz029.

Kosowski, R., A. Timmermann, R. Wermers, and H. White. 2006. "Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis." *J. Finance* 61 (6): 2551–2595. https://doi.org/10.1111/j.1540-6261.2006.01015.x.

Kyle, A. S. 1985. "Continuous auctions and insider trading." *Econometrica* 53:1315–1335. https://doi.org/10.2307/1913210.

Manski, C. F. 2006. "Interpreting the predictions of prediction markets." *Econ. Lett.* 91:425–429. https://doi.org/10.1016/j.econlet.2006.01.004.

Ng, H., L. Peng, Y. Tao, and D. Zhou. 2026. "Price discovery and trading in modern prediction markets." SSRN working paper 5331995.

Page, L., and R. T. Clemen. 2013. "Do prediction markets produce well-calibrated probability forecasts?" *Econ. J.* 123:491–513. https://doi.org/10.1111/j.1468-0297.2012.02561.x.

Reichenbach, F., and M. Walther. 2025. "Exploring decentralized prediction markets: Accuracy, skill, and bias on Polymarket." SSRN working paper 5910522.

Sharpe, W. F. 1992. "Asset allocation: Management style and performance measurement." *J. Portfolio Manag.* 18:7–19. https://doi.org/10.3905/jpm.1992.409394.

Sirolly, A., H. Ma, Y. Kanoria, and R. Sethi. 2025. "Network-based detection of wash trading." SSRN working paper 5714122.

Slivkoff, S. 2025. *Polymarket volume is being double-counted.* Paradigm research note, December 2025.

Snowberg, E., and J. Wolfers. 2010. "Explaining the favorite-long shot bias: Is it risk-love or misperceptions?" *J. Polit. Econ.* 118:723–746. https://doi.org/10.1086/655844.

Tsang, K. P., and Z. Yang. 2026. "The anatomy of a blockchain prediction market: Polymarket in the 2024 U.S. presidential election." ArXiv:2603.03136.

Wolfers, J., and E. Zitzewitz. 2004. "Prediction markets." *J. Econ. Perspect.* 18:107–126. https://doi.org/10.1257/0895330041371321.