# Public Learning from Decentralized Prediction Markets: Evidence from Polymarket Earnings Contracts *

Lin Ma† Wensi Xie‡

**Abstract**

Can small prediction markets inform much larger stock markets? We combine Kyle-style informed trading with cross-market Bayesian learning to study how identity protection and trading transparency shape informed participation and outside learning. Using Polymarket earnings contracts, we document three findings. First, trading advantages persist beyond measured skill and observed public news. Complementary evidence supports private earnings information as a source of these advantages. Second, investors direct attention toward profitable traders, consistent with information extraction. Stock prices also respond to Polymarket signals containing noise, supporting learning beyond responses to common fundamental news. Third, Polymarket coverage is associated with greater earnings incorporation before disclosure, but no significant difference during the announcement window. The pre-announcement ERC is about 28% of its announcement-window counterpart for covered events, with stronger responses when Polymarket signals are more visible and informative. Together, these findings suggest that prediction markets reveal private information that outside investors use to make better-informed investment decisions, informing the debate over their public value.

**Keywords:** prediction markets; informed trading; private information; earnings announcements.

*We thank Ran Duchin, Yang Su, Dan Luo, Xian Wu, Chao Ying, Chen Yao, Jiasun Li, and participants at the CUHK Finance Second-Year Ph.D. Paper Defense (June 2, 2026) for their helpful comments and suggestions. All errors are our own.

†Department of Finance, CUHK Business School, The Chinese University of Hong Kong. Email: linma@link.cuhk.edu.hk.

‡Department of Finance, CUHK Business School, The Chinese University of Hong Kong. Email: wensixie@cuhk.edu.hk.

# 1 Introduction

Prediction markets have grown into major trading venues, with monthly volumes exceeding USD 40 billion (Figure 1). Alongside this growth is an unresolved debate over whether these markets create public value. The U.S. Commodity Futures Trading Commission (CFTC) poses the question directly:

> *"Is there some public interest utility if people with an asymmetric information advantage on a particular event contract are able to trade on prediction markets?"*[1]

Existing research shows that prediction markets aggregate dispersed information and generate informative forecasts (Forsythe et al., 1992; Cowgill and Zitzewitz, 2015; Gomez Cram et al., 2026). Yet forecasting accuracy alone does not establish their public value. Prediction-market prices may echo information already available to decision-makers, especially when traditional financial markets already aggregate news about the same fundamentals. This raises three questions: do prediction markets contain incremental information, do outside investors extract and use it, and does that use improve the information underlying their decisions? These questions closely relate to Arrow et al. (2008)'s argument that prediction markets can provide useful information for decisions beyond the market.

To address these questions, we use 996 Polymarket earnings contracts from September 2025 to June 2026 to study stock-market investors as outside users of prediction-market information. The direct link between each contract and the corresponding stock allows us to trace information across markets exposed to the same earnings news.[2] We first develop a framework combining Kyle-style informed trading with cross-market Bayesian learning. Informed traders enter Polymarket when identity-cost savings compensate for lower trading rents; stock investors acquire its signals when their incremental information is worth the acquisition and interpretation costs; and lower observation noise allows investors to learn more about earnings. Whether these conditions hold is an empirical question.

[1] See the CFTC Prediction Markets ANPRM, March 16, 2026, Section II.E, Question 29.

[2] The setting has four useful features. (1) **Potential private information.** Earnings announcements offer opportunities for private-information trading (Akey et al., 2022), as illustrated by SEC cases involving Dell, Nvidia, Green Mountain Coffee, and Ross Stores. Prediction-market allegations also include the DOJ's May 2026 Google-employee case and a separate U.S.-soldier case (SEC enforcement archive; DOJ Google case; DOJ military-information case). These examples do not establish information sources in our sample. (2) **A direct link to stock prices and realized earnings.** Each contract maps to a stock, avoiding an event-exposure portfolio, and subsequent earnings releases allow us to evaluate signal content. (3) **Access for outside investors.** Polymarket's public interface and Stocktwits integration make signals accessible (Appendix Figure A9). Bloomberg reports that Wolfe Research examined these probabilities for information beyond analyst forecasts (April 16, 2026). (4) **An information-rich benchmark.** Stock investors already have access to analyst forecasts, corporate disclosures, and information reflected in stock prices. This makes the stock market a demanding setting in which to test whether prediction markets provide incremental earnings information.

First, we find evidence consistent with private earnings information in Polymarket. We begin by examining informed trading at the market and order levels. The median transaction-level price error is 0.18, and prices become more accurate as trading volume increases and resolution approaches, consistent with markets aggregating information. At the order level, larger liquidity-taking orders—aggressive, marketable orders that execute against existing quotes—earn higher unit profits, while larger liquidity-providing orders earn lower unit profits. These patterns are consistent with informed traders acting on their signals and liquidity providers facing adverse selection.

We next examine whether the order-level evidence of informed trading can be explained by public information. First, traders with general skill may profit by processing public information more effectively. We adapt the skill-versus-luck approach of Gomez Cram et al. (2026) to identify wallets whose profits are unlikely to arise from chance. We measure skill using prior trading and non-earnings trading histories. The positive relation between liquidity-taking order size and unit profitability persists after excluding wallets classified as skilled under either measure. Second, we separately exclude days with macroeconomic news and firm-specific I/B/E/S EPS forecasts. The relation remains positive and significant in both tests. Together, these findings suggest that public-information explanations do not fully account for the observed informed-trading patterns.

Three additional patterns support the presence of private earnings information. First, profits are more concentrated across traders than losses, and this concentration gap exceeds the corresponding gap in trading volume, suggesting that trading scale alone does not explain the asymmetry. Second, the positive size–profitability relation is stronger for liquidity-taking trades associated with multiple wallets or repeated trading. These patterns are consistent with informed traders spreading their activity across wallets or over time to conceal their information advantage and limit price impact (Kyle, 1985). Third, some nodes in the auditor, city, and director networks exhibit unusually high prediction accuracy relative to random-assignment benchmarks. This suggests potential information access through professional or geographic connections, although we do not identify individual traders' information sources.

Second, we find evidence consistent with stock investors learning from Polymarket. We begin with profile views, the cumulative visits to public wallet pages displaying trading histories and positions (Appendix Figure A2), as a measure of ex post attention. Regressions show a positive association between profitability and views, especially for liquidity-taking trades. Binned scatterplots further reveal a pronounced monotonic increase in views with profitability. The pattern is consistent with investors rationally directing attention toward potentially informative traders, but profile views do not identify whether those investors participate in the stock market.

We then turn to stock returns to trace the transmission of Polymarket signals to equity prices. We classify traders as skilled, insider candidates, or noise traders. These groups account for 17.7%, 1.3%, and 80.9% of participant token volume and earn aggregate profits of 73,125, 103,272, and −176,397 USDC, respectively. We examine subsequent Polymarket price movements to assess the information content of each group's trades. Skilled flow has a contemporaneous price impact without significant subsequent continuation or reversal, consistent with effective processing of public information that the market quickly incorporates. Continued price movements following insider-candidate flow suggest initial underreaction, whereas the partial reversal of noise-flow price impact and its lack of significant contemporaneous accuracy gains suggest nonfundamental pressure. We then examine stock responses separately during regular trading sessions and overnight closures. Skilled flow is strongly associated with close-to-open stock returns. More strikingly, noise flow also has a positive association with these returns, despite its apparent nonfundamental component. Neither group exhibits a significant contemporaneous association during regular trading. This timing is consistent with greater reliance on Polymarket signals when the more liquid stock market is closed, as suggested by the framework's venue-choice mechanism. By tracing apparent nonfundamental price pressure into stocks, these results strengthen the learning interpretation beyond an explanation based solely on common fundamental news.

Third, we find that stock prices incorporate earnings information before the announcement when Polymarket covers the event. To mitigate selection bias and omitted-variable concerns, we compare covered announcements with propensity-score-matched uncovered events, include firm and year-quarter fixed effects, and control for firms' information environments. Covered events exhibit a significantly higher earnings response coefficient (ERC) in the pre-announcement window $[-5, -1]$, whereas the coverage-related difference in ERCs is insignificant in the announcement window $[0, +2]$. The incremental pre-announcement ERC implies a 2.85-percentage-point larger return spread between the highest and lowest earnings-surprise ranks for covered events. Their total pre-announcement ERC is approximately 28% of the corresponding announcement-window ERC, indicating economically meaningful incorporation of earnings information before disclosure. This finding is robust to alternative event windows, return measures, and matching specifications, and to excluding analyst-forecast or macroeconomic-news days. Permuting coverage within firms or year-quarters rarely yields coefficients as large as the observed estimate.

We then examine whether pre-announcement earnings incorporation varies with signal visibility and information content. We partition covered events into high and low groups at the median of each market characteristic. Pre-announcement ERCs are higher in markets with greater trading volume and wallet profile views, which proxy for signal visibility.

ERCs are also higher in markets with greater liquidity-taking, repeated-trading, or multi-wallet activity, which proxy for information content. These patterns are consistent with the framework's Bayesian-learning prediction: stock prices incorporate more earnings information before announcements when prediction-market signals are easier to observe and more informative.

Our paper contributes to three related literatures.

First, we contribute to the prediction-market literature by developing a framework for understanding when small prediction markets can inform much larger stock markets and providing systematic evidence from earnings markets. Classic work establishes their capacity to aggregate dispersed information and generate informative forecasts (Forsythe et al., 1992). Recent studies examine macroeconomic and earnings forecasts (Diercks et al., 2026; Gomez Cram et al., 2025; Rabetti et al., 2026; Feng, 2026), and the roles of trader skill and adverse selection in price formation (Gomez Cram et al., 2026; Bartlett and O'Hara, 2026). Related research examines information sharing within firms and the use of prediction-market signals in political and financial settings (Feigenbaum and Shelton, 2013; Cowgill and Zitzewitz, 2015; Goldstein et al., 2026; Aktuğ and Torul, 2026). We bring an institutional perspective to this literature: wallet pseudonymity can attract informed trading by lowering identity costs, while transparent blockchain records facilitate outside learning. Our framework formalizes these mechanisms by extending the Kyle model to venue choice and incorporating cross-market Bayesian learning.

Our evidence clarifies both the source and the outside value of prediction-market information. First, trading patterns support private earnings signals beyond measured general skill and observed public news. Second, profile views document attention to profitable trading, while stock-price responses trace the transmission of prediction-market signals. These tests provide direct evidence of attention to potential information sources and complementary evidence of information extraction and learning. Third, stock prices incorporate more earnings information before announcements when signals are more visible and informative, suggesting that stock-market investors collectively learn economically meaningful information. Together, the framework and evidence explain how a small prediction market can supply incremental information to a larger, information-rich stock market. This supports Arrow et al. (2008)'s view that prediction markets can create public value by informing decisions beyond their participants.

Second, we contribute to the literature on informed trading venue choice and cross-market price discovery. Leverage, downside protection, and lower trading costs can attract informed traders to options markets (Easley et al., 1998). Option-market measures predict stock returns and contribute to price discovery (Pan and Poteshman, 2006; Cao et al.,

2005; Chakravarty et al., 2004). Related evidence documents information flows from credit derivatives (Acharya and Johnson, 2007), corporate bonds (Abdi and Wu, 2018), and stock-index futures around macroeconomic announcements (Ying, 2020). We extend this line of research to decentralized prediction markets, an emerging source of information for financial markets. Our theoretical and empirical analyses explain how signals emerge from trading, become available for stock investors to learn from, and contribute to earnings information in stock prices. We also highlight a mechanism linking venue choice to outside learning: wallet pseudonymity can attract informed traders by reducing identity-exposure costs, while public post-trade records allow outside investors to learn from their trading histories. This combination of identity protection and trading transparency provides a channel for cross-market information transmission.

Third, we contribute to accounting research on how earnings information enters stock prices. Since Ball and Brown (1968), this literature has studied the timing of stock-price responses to earnings news and the roles of investor attention and the information environment (Bernard and Thomas, 1989; Hirshleifer et al., 2009; Hirshleifer and Sheng, 2022). Technological changes can alter investors' incentives and ability to acquire and process accounting information (Blankespoor et al., 2020). Recent studies examine crowdsourced forecasts (Schafhäutle and Veenman, 2024), integrated trading platforms (Chau, 2025), and generative AI (Bertomeu et al., 2025; Chang et al., 2026) as channels affecting investors' access to and processing of earnings information. Related work examines the forecasting value of earnings prediction markets (Gomez Cram et al., 2025; Rabetti et al., 2026; Feng, 2026). We extend this research by examining whether stock investors learn from prediction-market signals and how much earnings information enters stock prices before announcements. Polymarket coverage is associated with a higher pre-announcement earnings response coefficient, particularly when signals are more visible and informative. For covered events, this coefficient is about 28% of its announcement-window counterpart, indicating economically meaningful incorporation before disclosure. These findings identify an information-intermediation role for prediction markets: they supplement the information investors use to anticipate earnings, while earnings announcements remain an important source of information.

The rest of the paper proceeds as follows. Section 2 presents stylized facts about earnings prediction markets and a framework for three hypotheses. Section 3 presents the data, empirical designs, and preliminary evidence. Section 4 examines private information. Section 5 studies investor attention and stock-price responses to Polymarket order flow. Section 6 examines earnings incorporation, and Section 7 discusses implications for blockchain market design, information users beyond the platform, and regulation.

# 2 Theoretical Framework

Why would informed traders choose a prediction market with limited trading activity, and under what conditions would stock investors learn from its prices? We develop a framework to formalize informed participation and subsequent learning from prediction-market signals. Appendix 4 provides the derivations.

## 2.1 Stylized Facts of Earnings Prediction Markets

Earnings prediction contracts let traders bet on whether a firm will beat analysts' earnings forecasts. On Polymarket, each contract asks whether reported EPS will exceed a fixed threshold based on the analyst consensus. A "Yes" token pays 1 USDC if it does and zero otherwise; a "No" token pays in the opposite state. The Yes price can be interpreted as a market-implied probability of an earnings beat (Wolfers and Zitzewitz, 2004, 2006). Polymarket and Stocktwits announced a partnership in September 2025 that began with earnings markets.[3] By June 2026, our sample includes 996 closed earnings contracts (Figure 2, Panel A).

Earnings contracts use the same trading infrastructure as other Polymarket markets. Traders submit signed orders to a central limit order book, where orders are matched off-chain. A maker posts a limit order at a specified price, supplying liquidity to the book. A taker executes against an available quote and consumes that liquidity. Taker trades are thus more aggressive because they seek immediate execution rather than waiting for another trader to accept a posted quote. Executed trades settle on Polygon.

Settlement creates public records of trade prices, quantities, timing, and participating wallets (Appendix Table A2 and Figure A3). Wallet addresses generally conceal legal identities while allowing observers to trace each wallet's trading history. Public profiles make this history easier to access, displaying usernames, positions, and past trades; some also report view counts (Appendix Figure A2). Stock investors can therefore follow a wallet's trading and assess its track record without knowing who controls it.

Earnings markets are recurring but relatively short-lived. In the broader platform sample reported in Appendix Table A1, 80.67% of earnings events belong to repeated series; their median duration is 12.41 days and median trading volume is 6,311.59 USDC.[4] Figure 2, Panels B–D, shows the duration and volume distributions in our earnings-contract sample and the concentration of trading near market close.

[3] See the Stocktwits partnership announcement.

[4] Small markets can nevertheless produce useful forecasts. The 1988 Iowa Political Stock Market operated from June 1 to November 9 with 192 traders and a total investment of $4,967. Its prices forecast election results more accurately than contemporaneous polls (Forsythe et al., 1992).

Four features of earnings prediction markets guide the model. (1) Contract prices signal the probability of a positive earnings surprise, which is also relevant to stock-market participants. (2) Trading volume is small relative to the underlying stock market. (3) Standard registration on Polymarket International does not require know-your-customer (KYC) identity verification, and participants trade under wallet pseudonyms without publicly disclosing their legal identities. (4) Public transaction records and user profiles make trading activity visible and accessible to stock investors, facilitating signal acquisition.

## 2.2 Institutional Foundations and Simplifying Assumptions

We extend the one-period model of Kyle (1985) to study informed traders' choice between two venues and stock investors' subsequent learning from prediction-market signals. The sequence has three stages. First, informed traders compare expected trading rents net of identity costs and choose a venue. Second, informed and noise orders generate total order flow, market prices, and a public Polymarket trading record. Third, stock investors decide whether to acquire the public Polymarket signal $\boldsymbol{s}_P$; those who do update their earnings beliefs and can incorporate them into stock prices before the announcement.

Denote the two venues by $P$ (Polymarket) and $S$ (the stock market). The common fundamental is an earnings surprise, $\epsilon \sim N(0, \sigma_\epsilon^2)$, relative to expectations based on public information before trading. We normalize the stock's earnings-related payoff to $v_S = \epsilon$ and use a continuous claim on the same surprise for Polymarket, with price $\boldsymbol{p}_P$.[5]

In each venue $m \in \{P, S\}$, an informed trader observes $\epsilon$ and submits $x_m(\epsilon)$. Noise orders $u_m \sim N(0, \sigma_{\ell,m}^2)$ are independent of $\epsilon$. A competitive, risk-neutral market maker observes total order flow $z_m = x_m(\epsilon) + u_m$ and sets price equal to the conditional expected payoff. Within this common, normalized benchmark, we assume $\sigma_{\ell,P} < \sigma_{\ell,S}$: the stock market's larger pool of noise traders provides more scope for informed trading before prices move. Trader $i$ incurs an expected identity-exposure cost $C_{m,i}$, which can be lower under Polymarket's wallet pseudonymity.

All stock-market participants, including noise traders, can acquire the public signal $\boldsymbol{s}_P = \boldsymbol{p}_P + \eta_P$. Observation noise $\eta_P \sim N(0, \tau_P^2)$ is independent of earnings news and noise orders and captures difficulty interpreting the trading record. Investor $j$ evaluates the signal relative to existing information $\mathcal{I}_{S,j}$ and updates earnings beliefs by Bayes' rule if acquisition is worthwhile.

[5]The actual contract pays $Y = 1\{\epsilon > \kappa\}$ for threshold $\kappa$ and has probability price $\boldsymbol{q}_P = \Pr(\epsilon > \kappa \mid \mathcal{I}_P)$, where $\mathcal{I}_P$ is the information embedded in Polymarket trading. The Kyle benchmark instead prices a continuous claim on $\epsilon$; Appendix 4 relates its price $\boldsymbol{p}_P$ to the binary price $\boldsymbol{q}_P$.

## 2.3 Why Private Information Enters Polymarket

In the Kyle benchmark, noise orders conceal part of the informed order, yielding expected gross profits

$$R_m = \frac{1}{2}\sigma_\epsilon \sigma_{\ell,m}, \qquad R_P < R_S. \tag{1}$$

Lower liquidity therefore makes Polymarket less attractive on gross profits alone (Kyle, 1985). Its potential advantage is a lower cost of exposing the trader's identity, research activity, or professional relationships. Let $C_{m,i}$ denote trader $i$'s expected cost of identity exposure when trading in venue $m$. Comparing net profits gives the participation condition

$$\underbrace{C_{S,i} - C_{P,i}}_{\text{identity-cost savings}} \geq \underbrace{R_S - R_P}_{\text{trading-rent gap}}. \tag{2}$$

The left-hand side of equation (2), $C_{S,i} - C_{P,i}$, is the trader's identity-cost savings from choosing Polymarket. The right-hand side, $R_S - R_P$, is the trading-rent gap: the expected gross profits forgone by trading there instead of in stocks. Limited trading activity therefore need not prevent informed participation. Polymarket can attract private earnings information when the identity-cost savings compensate some informed traders for the lower trading rents. Traders whose savings fall short of this gap prefer the stock market.

**H1 (Existence of Private Information).** *Despite its limited trading activity, Polymarket can attract traders with private earnings information when their identity-cost savings, $C_{S,i} - C_{P,i}$, are at least as large as the trading-rent gap, $R_S - R_P$.*

## 2.4 How Stock Investors Learn from Polymarket

Private earnings information in Polymarket creates an opportunity for stock investors to learn. Learning can transmit noise alongside information (Goldstein et al., 2026), and public records remain costly to acquire and interpret (Hu et al., 2023). Investors therefore weigh the expected benefit of learning against these costs.

Formally, let $\epsilon$ denote the earnings innovation remaining after conditioning on common stock-market information, with prior mean zero. The Kyle price is $\boldsymbol{p}_P = \epsilon/2 + \lambda_P u_P$, where $\lambda_P$ measures price impact. The signal available for acquisition is therefore

$$\boldsymbol{s}_P = \underbrace{\epsilon/2}_{\text{earnings information}} + \underbrace{\lambda_P u_P}_{\text{trading noise}} + \underbrace{\eta_P}_{\text{observation noise}}. \tag{3}$$

Trading noise $\lambda_P u_P$ is already embedded in the price. Observation noise $\eta_P$ captures error

in the investor's observation or interpretation of that price signal, with variance $\tau_P^2$. The investor observes the combined signal rather than its separate components.

Let $B_j(\boldsymbol{s}_P \mid \mathcal{I}_{S,j})$ denote the expected gain from decisions based on updated beliefs rather than $\mathcal{I}_{S,j}$ alone, evaluated before observing the signal and deducting costs. With acquisition and interpretation cost $K_j$, investor $j$ acquires the signal when

$$\underbrace{B_j(\boldsymbol{s}_P \mid \mathcal{I}_{S,j})}_{\text{expected benefit from learning}} > \underbrace{K_j}_{\substack{\text{acquisition and} \\ \text{interpretation cost}}}. \tag{4}$$

The anticipated benefit comes from using the posterior beliefs obtained after acquisition. The expected earnings innovation changes from zero to

$$E[\epsilon \mid \boldsymbol{s}_P] = a_P \boldsymbol{s}_P, \qquad a_P = \frac{\sigma_\epsilon^2/2}{\sigma_\epsilon^2/2 + \tau_P^2} > 0. \tag{5}$$

The Bayesian coefficient $a_P$ governs the response to the signal and falls as observation noise increases. Because investors observe information and noise together, their belief revisions reflect both.

**H2 (Cross-Market Learning).** *When $B_j(\boldsymbol{s}_P \mid \mathcal{I}_{S,j}) > K_j$ and $\boldsymbol{s}_P$ contains incremental earnings information, equity investors use the posterior belief $E[\epsilon \mid \boldsymbol{s}_P] = a_P \boldsymbol{s}_P$ in their trading decisions.*

## 2.5 When Learning Advances Earnings Incorporation

Learning from Polymarket does not necessarily make stock prices more informative about earnings. The information gain requires informed participation under equation (2) and acquisition of an incremental signal under equation (4); lower observation noise makes that gain larger. Conditional on these participation and acquisition decisions, Bayesian updating reduces earnings uncertainty from the prior variance $\sigma_\epsilon^2$, measured after accounting for common stock-market information, to

$$\mathrm{Var}(\epsilon \mid \boldsymbol{s}_P) = \sigma_\epsilon^2 - \underbrace{\frac{\sigma_\epsilon^4}{2\sigma_\epsilon^2 + 4\tau_P^2}}_{\text{reduction from learning}} < \sigma_\epsilon^2. \tag{6}$$

The subtracted term is strictly positive for $\sigma_\epsilon^2 > 0$ and finite $\tau_P^2$, so the posterior variance is lower than the prior variance. With no observation noise, uncertainty falls by half; trading noise accounts for the uncertainty that remains. As observation noise increases, the reduction

becomes smaller and approaches zero in the limit.

When investors acquire and use this signal before the announcement, stock prices can reflect part of the earnings innovation sooner. The information gain is larger when observation noise is low relative to the variance of the earnings innovation.

**H3 (Earnings Information Incorporation).** *If H1 and H2 hold and observation noise $\tau_P^2$ is sufficiently low, stock prices incorporate an economically meaningful amount of earnings information before the announcement.*

These signals may add little for corporate insiders who already know earnings. For ordinary investors, Bayesian learning can narrow information disadvantages even shortly before disclosure and reduce informed traders' rents (Appendix 4), offering potential investor protection. Without investor-level stock-market data to measure this protection directly, we do not propose a separate testable hypothesis.

# 3 Data and Empirical Design

## 3.1 Data and Sample Construction

Our Polymarket data comprise event metadata, trade records, and user profiles. We obtain contract information, settlement outcomes, and account attributes from Polymarket's APIs. We retrieve historical trades from The Graph's V1 and V2 orderbook subgraphs and combine them across versions. We also scrape public profile pages for cumulative profile views, which measure trader visibility.

A participant may trade through multiple wallets, so wallet addresses need not correspond to distinct users. We use token-transfer records from PolygonScan to identify funding links outside Polymarket and group connected wallets into inferred users, excluding transfers involving official platform addresses. This consolidates positions, trading histories, and profits across linked wallets. Appendix 3 describes the clustering procedure and results.

We link these contracts to I/B/E/S earnings forecasts and actual EPS, CRSP stock returns, Compustat firm characteristics, FactSet institutional ownership, and RavenPack news. TAQ quotes and trades provide high-frequency stock prices and trading measures for aligning equity responses with Polymarket order flow.

The sample covers 996 closed earnings contracts from September 2025 to June 2026. We remove terminal trades after prices have effectively converged to the realized outcome, as these trades contribute little to price discovery. The trade-level sample contains 475,855 execution records, and aggregation across linked wallets yields 144,235 inferred user–market

observations. Table 1, Panel A, shows substantial dispersion in trading activity and profits: mean trade value is 34.32 USDC, compared with a median of 4.65 USDC, while trade profits average approximately zero and range from −10,050 to 10,300 USDC. Panel B summarizes 2,522 earnings announcements, of which 27.8% are covered by Polymarket.

Appendix 1 defines all variables. Appendix 3 details data collection, sample construction, and processing.

## 3.2 Empirical Design

**Existence of Private Information.** Public information can account for much of price discovery; our tests ask whether private earnings information also contributes. Wallet pseudonyms do not reveal traders' identities or information sources. We therefore use detailed wallet trading histories to identify patterns that are difficult to explain by observed public information and general trading skill but consistent with private earnings signals. We control for non-earnings and historical trading skill and separately exclude observed public-news days. Market and day fixed effects absorb stable contract differences and common daily shocks where specified. These tests assess private-information explanations without ruling out all forms of public-information processing.

**Stock-Market Learning from Polymarket.** We first examine the price impact of order imbalance (OIB) from skilled traders ($S$), insider candidates ($I$), and noise traders ($N$) within Polymarket; Appendix 3 defines the groups and OIB. Price dynamics help assess the information content of their trades. Informative flow should generate persistent price changes, with subsequent continuation if information is initially underincorporated. Noise flow may also move prices contemporaneously, but its impact should subsequently reverse. The information-diffusion and overreaction mechanisms in Hong and Stein (1999) motivate our interpretation of continuation and reversal. We estimate:

$$\Delta p^h_{m,t} = \alpha_h + \sum_{g\in\{S,I,N\}} \beta_{g,h} OIB^g_{m,t} + \gamma_h' X_{m,t} + \varepsilon^h_{m,t}, \tag{7}$$

where $m$ indexes earnings contracts, $t$ is the signal interval's start, $g$ denotes trader group, and $h$ indexes the outcome interval. The dependent variable is the change in the Polymarket YES probability in percentage points. OIB is fixed over the six hours beginning at $t$; outcomes cover that interval and the next two separate six-hour intervals, rather than cumulative windows. We also measure price discovery as the reduction in absolute settlement error, $PD^h = 100(|p_a - Y| - |p_b - Y|)$, where $Y$ is the final YES payoff and $(a, b)$ are the outcome

endpoints. Its regression replaces $\Delta p^h$ with $PD^h$ and current signed OIB with $|OIB^g|$, retaining signed lagged OIB controls. Together, price dynamics and settlement accuracy provide evidence on the groups' information content, rather than directly identifying traders' information sources.

We next examine whether Polymarket order flow reaches stock prices, focusing on noise-group flow. Nonfundamental price pressure can help distinguish cross-market learning from responses to common fundamental news (Edmans et al., 2012; Goldstein et al., 2026). Noise flow must move Polymarket prices; if stock investors cannot fully distinguish information from noise, these price movements can also affect stocks. This provides the relevance mechanism, while a causal interpretation additionally requires that the underlying pressure be unrelated to earnings fundamentals and other stock-return shocks. We estimate:

$$r^h_{i,t} = a_h + \sum_{g\in\{S,I,N\}} \theta_{g,h} OIB^g_{i,t} + \delta_h' Z_{i,t} + u^h_{i,t}, \tag{8}$$

where $i$ indexes the corresponding stock and $r^h_{i,t}$ is its log return in percent. We aggregate Polymarket liquidity-taking tokens across contracts for the same stock before constructing group OIB. To match stock-market trading hours, signals span either a regular session, from open to close, or an overnight session, from close to the next trading day's open. Outcomes cover the signal session and the next two separate sessions, holding the original OIB fixed. The first coefficient measures a contemporaneous response; subsequent coefficients assess whether order flow predicts returns after the signal interval ends.

In equation (7), $X$ includes opening probability, lagged probability change, lagged log volume, signed lagged group OIBs, and log hours to cutoff. In equation (8), $Z$ includes return, stock OIB, log volume, and opening spread from the latest completed regular session before signal start. Both regressions include an intercept and no fixed effects. Standard errors are two-way clustered by event and signal-start date in equation (7), and by ticker and signal-start date in equation (8), using New York dates. Samples vary by horizon; Appendix 3 provides construction details.

**Earnings Information Incorporation into Stock Prices.** Following Hirshleifer and Sheng (2022), we test whether stock prices reflect more earnings information before the announcement when Polymarket covers the event:

$$\begin{aligned} BHAR^w_{i,q} = \alpha + \beta_1 SUE_{i,q} + \beta_2 Covered_{i,q} + \beta_3 \left(SUE_{i,q} \times Covered_{i,q}\right) \\ + \Gamma' X_{i,q} + f_i + f_q + \varepsilon^w_{i,q}, \end{aligned} \tag{9}$$

For firm $i$ and year-quarter $q$, $BHAR^w$ is the CAPM-adjusted buy-and-hold return in the pre-announcement $[-5,-1]$ or announcement $[0,+2]$ window. $SUE$ is actual EPS minus the analyst consensus, scaled by the pre-announcement price and converted to a normalized decile rank using prior-quarter breakpoints. *Covered* identifies a matching Polymarket contract. The earnings response coefficient (ERC) is $\beta_1$ for uncovered events and $\beta_1 + \beta_3$ for covered events; $\beta_3$ measures the incremental ERC associated with coverage. H3 predicts $\beta_3 > 0$ before the announcement, when prediction-market signals can help investors anticipate earnings. The announcement window provides a comparison, with no imposed sign on its interaction coefficient.

We use propensity-score matching (PSM) to improve comparability on observed characteristics; Appendix Table A4 reports the matching procedure and balance. Controls $X_{i,q}$ include analyst coverage, firm size, lagged ROA and leverage, prior institutional ownership, news coverage, and news sentiment. Firm and year-quarter fixed effects, $f_i$ and $f_q$, absorb stable firm differences and common time shocks. Standard errors are clustered by firm.

# 4 Private Information in Polymarket Earnings Markets

We first examine information aggregation and informed trading at the market and order levels. We then test whether measured trading skill and observed public news explain these patterns. Finally, we examine profit concentration, trading concealment, and potential information channels for further evidence of private earnings information.

## 4.1 Price Informativeness and Informed Trading

We first test whether Polymarket prices aggregate earnings information. Figure 3 plots the distribution of transaction-level price error, measured as the absolute difference between the Yes-token price and its realized payoff. Errors are concentrated near zero, with a median of 0.18 and a mean of 0.282. Accuracy also improves with transaction volume and as trading progresses. Appendix Figure A7 illustrates these patterns, and Appendix Table A7 confirms them in regressions with market and calendar-day fixed effects.

We next examine whether traders scale their orders with their information advantage. In the one-period Kyle framework, a larger gap between the informed trader's valuation and the market's prior valuation implies both a larger order and a higher expected profit per unit traded (Kyle, 1985). We assess the informativeness of these trading decisions using realized profit per token, calculated from the difference between the settlement value and

the execution price, adjusted for trade direction. Table 2 relates this measure to log trade value or log token quantity, separately for liquidity-taking and liquidity-providing orders, with market and calendar-day fixed effects. Larger liquidity-taking orders earn higher unit profits, while larger liquidity-providing orders earn lower unit profits. These patterns are consistent with informed traders taking liquidity and exposing liquidity providers to adverse selection (Glosten and Milgrom, 1985). These results do not imply that taking liquidity is generally more profitable. Appendix Table A8 shows lower average unit profits for liquidity-taking orders, while Appendix Figure A8 shows no monotonic relation between traders' liquidity-taking shares and profitability. The evidence of informed trading instead concerns how profitability varies with order size. The following tests examine the sources of these information advantages.

## 4.2 Private Information versus General Skill

General skill, such as effective trading strategies or superior processing of public information, could explain the observed information advantages without requiring private information about particular earnings events. We adapt the skill-versus-luck approach of Gomez Cram et al. (2026) to identify wallets whose profits are unlikely to arise from chance. We use non-earnings markets and trading histories through September 15, 2025 as separate reference samples. Profitability in these samples may reflect general skill that carries over to earnings trading. Appendix 3 describes the classification procedure.

Table 3 reports the size–profitability regressions for liquidity-taking orders. Panel A excludes wallets classified as skilled under each measure separately. The trade-value coefficients are 0.0062 and 0.0063 under the non-earnings and historical classifications, respectively; the token-quantity coefficients are 0.0087 and 0.0088. All four are significant at the 1% level. Panel B reports positive but less precisely estimated coefficients among skilled traders; only the token-quantity coefficient under the non-earnings classification, 0.0061, is significant at the 10% level. The positive size–profitability relation persists after excluding wallets classified as skilled under either measure. Measured general skill therefore does not fully explain the order-level evidence of informed trading.

## 4.3 Private Information beyond Public Information

Superior processing of public information could also explain the positive relation between liquidity-taking order size and profitability. To address this concern, we examine whether this relation persists outside important public-information releases. Specifically, we separately exclude macroeconomic-news days and days with firm-specific I/B/E/S EPS forecasts, then

repeat the order-level tests in Table 2.

Table 4 supports a role for public information without attributing the entire size–profitability relation to it. The relation is strongest on analyst EPS forecast days, suggesting that analyst forecasts provide signals for profitable trading. Nevertheless, the trade-value and token-quantity coefficients remain positive outside these releases—0.0059 and 0.0070 after excluding analyst-forecast days, and 0.0069 and 0.0092 after separately excluding macroeconomic-news days—all significant at the 1% level. The persistence of these patterns suggests that trading around observed public news is only part of the explanation.

## 4.4 Profit Concentration and Information Advantages

We next examine whether gains among top winners are more concentrated than losses among top losers. Publicly available information allows more traders to compete for the same opportunities, whereas private information may concentrate gains among a smaller group. Within each market, we separate winners from losers and calculate the shares of total gains and total losses attributable to the top 5% and 10% of each group:

$$\text{TopShare}^{X}_{m,g}(q) = \frac{\sum_{u \in T_q(m,g)} X_{u,m}}{\sum_{u \in U_{m,g}} X_{u,m}}. \tag{10}$$

Here, $m$ denotes a market, $g$ denotes winners or losers, and $U_{m,g}$ contains the users in that group. $T_q(m,g)$ selects the highest-profit winners or largest-loss losers, with $q \in \{5, 10\}$. $X_{u,m}$ denotes absolute profit or trading volume. Using the same selected users for both measures allows us to compare their shares of gains or losses with their shares of trading activity.

Figure 4 shows that top winners generally capture a larger share of total gains than top losers account for of total losses. Profits are therefore more concentrated, while losses are more dispersed, providing preliminary support for the conjecture that information advantages accrue to a relatively small group. Top winners also account for a substantial share of trading volume, although their volume shares tend to be somewhat lower than their profit shares. Their disproportionate gains thus reflect both substantial trading activity and higher profits per unit of trading value, consistent with larger-scale trading accompanied by more favorable trading outcomes.

## 4.5 Strategic Concealment and Trading Behavior

In Kyle (1985), informed traders have an incentive to keep their information from being revealed too quickly through trading. We examine two trading patterns consistent with

this incentive. Spreading trades over time may limit the information revealed by any single transaction, while using multiple wallets may make a trader's overall activity harder to track. We capture these patterns with *Multi-trade* and *Multi-wallet*, indicators for repeated transactions and multiple wallets within a user–market pair, respectively. To test whether these patterns are associated with stronger information advantages, we interact each indicator with order size in the liquidity-taking order sample:

$$\begin{aligned} \textit{PnL per Token}_{i,m,d} = \alpha + \beta_1 \ln(1 + \textit{val traded}_{i,m,d}) + \beta_2 \textit{Hide}_{i,m} \\ + \beta_3 \big[ \ln(1 + \textit{val traded}_{i,m,d}) \times \textit{Hide}_{i,m} \big] + f_m + f_d + \varepsilon_{i,m,d}, \end{aligned} \tag{11}$$

where $\textit{Hide}_{i,m}$ denotes either *Multi-wallet* or *Multi-trade*, and $f_m$ and $f_d$ are market and calendar-day fixed effects. If these patterns reflect informed traders concealing their information advantages, we expect $\beta_3 > 0$: the positive relation between order size and unit profitability should be stronger for trades associated with multiple wallets or repeated trading.

Table 5 reports positive and statistically significant interaction coefficients for both measures. Larger liquidity-taking orders therefore exhibit a stronger size–profitability relation when users spread their trading across wallets or transactions. These results are consistent with informed traders using concealment strategies to preserve information rents and manage price impact.

## 4.6 Potential Channels of Private Information

We next examine whether unusually accurate predictions cluster around potential channels of private information: common auditors, headquarters cities, and director relationships. Professional roles and local contacts may provide access to nonpublic earnings information. We first calculate each auditor's, city's, and director's observed accuracy by averaging the accuracy of its associated earnings markets. To construct a random benchmark, we shuffle the accuracy scores across all 996 markets while keeping each node's set of associated markets unchanged. We then recalculate each node's average accuracy using the reassigned scores and repeat this procedure 10,000 times.

Figure 5 highlights unusually high accuracy in parts of the auditor and director networks. The highest-accuracy auditor lies in the upper tail of the random-assignment benchmark. The director network contains multiple high-accuracy nodes, with observed counts in several upper-tail bins well above the average simulated counts. Some cities also exhibit high accuracy relative to the benchmark. These patterns are consistent with information advantages concentrated around particular professional and geographic connections.

We cannot directly establish who traded on private information or how they obtained it. Appendix Table A9 provides complementary accounts from platform participants. One comment alleges that traders work for an auditing firm and already know the earnings outcome. The auditor mentioned in that anonymized comment corresponds to the highest-accuracy auditor in Figure 5. This correspondence supports the possibility that professional relationships provide a channel for private earnings information, without verifying the allegation.

# 5 Stock-Market Learning from Polymarket

The preceding evidence suggests that Polymarket contains private earnings information. H2 predicts learning when these signals contain incremental information and the expected benefit exceeds acquisition and interpretation costs. To test this hypothesis, we examine whether investors extract information from Polymarket signals and, particularly, whether stock prices respond to them.

## 5.1 Investor Attention to Polymarket Signals

Investors' acquisition and use of available information underpin the efficient-market benchmark in which prices reflect that information (Fama, 1970). Acquisition and integration costs, however, can impede this process even in transparent markets (Hu et al., 2023). Polymarket's public interface makes traders' positions and trading histories accessible and traceable over time, potentially reducing the costs of acquiring and interpreting trading signals.[6] If investors learn from these signals, we expect them to direct attention toward potentially informative traders. We measure ex post attention using wallet profile views (Appendix Figure A2) and assess the potential informativeness of trading histories using realized profit per token. We therefore expect a positive association between attention and trading profitability.

Table 6 shows a positive association between profile views and realized profit per token across all orders and both liquidity-taking and liquidity-providing subsamples, significant at the 1% level in each case. The coefficient is larger for liquidity-taking orders than for liquidity-providing orders, at 0.006 versus 0.002. Figure 6 further shows a pronounced upward pattern across attention bins, with higher profitability among more visible traders. These patterns are consistent with investors rationally directing attention toward potentially informative traders, although profile views do not identify whether those investors

[6]Appendix Figure A10 illustrates a public tutorial explaining how to use Polymarket earnings markets for stock trading by tracking transaction sizes, timing, and trader histories. Released in October 2025, it had received more than 22,900 views by April 2026, providing an anecdotal example of how methods for reading these signals circulate.

participate in the stock market.

## 5.2 Stock-Price Responses to Polymarket Order Flow

The attention results do not establish whether stock investors learn from Polymarket. We therefore use equations (7) and (8) to examine what Polymarket order flow reveals and whether it is reflected in stock prices.

On the one hand, we assess the information content of trades by the three groups identified in Appendix 3: skilled traders, insider candidates, and noise traders. Table 7, Panel A, shows that skilled traders trade more than insider candidates but earn less. They account for 17.7% versus 1.3% of participant token volume, yet earn 73,125 versus 103,272 USDC in aggregate profits. Their profits per token are also lower, at 1.043 versus 19.388 cents. This pattern is consistent with skilled traders earning smaller unit advantages through more trading. Noise traders account for 80.9% of volume and lose 176,397 USDC. Panels B and C assess the groups' information content through price dynamics and settlement accuracy. Skilled flow moves prices and improves accuracy during the trading interval, with no significant later continuation or reversal. Insider-candidate flow has numerically larger coefficients for both price impact and accuracy improvement. Prices also continue moving in the same direction afterward, consistent with initial underreaction. These findings reinforce the earlier evidence that general skill alone does not explain informed trading. Noise flow, by contrast, moves prices without significantly improving accuracy during the trading interval, and its price impact partly reverses in both subsequent intervals. Its main contribution appears to be temporary price pressure rather than price discovery.

On the other hand, Table 8 shows that the positive stock-return associations are concentrated outside normal stock-market hours. Skilled and noise flow overnight—between the stock market's close and its next opening—are positively associated with stock returns over that interval. Neither group has a significant contemporaneous association during the trading day. This timing is consistent with the framework's venue-choice mechanism. The more liquid stock market attracts informed trading when open, while Polymarket may become a more important signal source outside regular hours. Daytime noise flow is also negatively associated with the following overnight stock return. The other groups require different interpretations. Skilled-flow associations may reflect common public news or learning from stock prices, leaving the direction of transmission unresolved. Insider-candidate flow is associated with a positive stock response one overnight interval later, consistent with delayed incorporation. However, classification based on ex post settlement profits introduces look-ahead bias.

The noise-flow results provide the clearest evidence of stock-market learning in this comparison. Our interpretation draws on the use of nonfundamental price pressure in Edmans et al. (2012) and the information-and-noise transmission mechanism in Goldstein et al. (2026). Table 7 links this flow to temporary price pressure without significant contemporaneous accuracy gains. Table 8 shows that it is nevertheless reflected in overnight stock returns. Together, these results provide evidence that stock investors learn from Polymarket earnings prediction markets, absorbing noise alongside information. Whether this learning helps stock prices incorporate meaningful earnings information requires further examination.

# 6 Earnings Information Incorporation into Stock Prices

Having found evidence of private earnings information and stock-market learning, we next examine whether investors learn economically meaningful information. This outcome is not automatic: Polymarket signals contain noise, and investors may have difficulty interpreting them. H3 predicts that, when H1 and H2 hold and observation noise is sufficiently low, stock prices incorporate earnings information before its public release. We test this prediction in equation (9) by comparing pre-announcement earnings response coefficients across announcements with and without Polymarket coverage.

## 6.1 Baseline Evidence on Earnings Information Incorporation

Polymarket coverage may favor firms that attract investor attention and trading demand. The platform asks market proposals to provide evidence of demand for trading, suggesting that coverage is unlikely to be random.[7] This motivates matching on analyst coverage, firm size, and news coverage as proxies for attention and potential trading demand, together with lagged ROA and leverage to account for differences in financial characteristics. Appendix Table A4 shows that matching reduces mean standardized bias from 75.5% to 1.3%, with no statistically significant matched mean differences. Firm and year-quarter fixed effects and additional controls further account for differences across firms and over time.

Appendix Table A6 shows that the full-sample results are consistent with the longstanding evidence that earnings announcements convey information to stock prices (Ball and Brown, 1968). Table 9 shows a stronger pre-announcement earnings response for events covered by Polymarket. The interaction between SUE and coverage is positive and significant at the 5% level in both the full and matched samples over $[-5, -1]$. In the matched sample, coverage is associated with a 2.85-percentage-point larger return difference between the

[7] See Polymarket's market-creation guidance.

highest and lowest earnings-surprise ranks for covered events. The total pre-announcement ERC for covered events is approximately 0.0276, about 28% of its announcement-window counterpart. This magnitude indicates economically meaningful earnings incorporation before disclosure, while leaving a substantial response to the announcement itself. By contrast, the coverage interaction is insignificant over $[0, +2]$. This pattern is consistent with the framework: prediction-market signals facilitate learning before disclosure, while the earnings announcement makes their earnings information redundant.

The pre-announcement result is robust to several changes in measurement and sample construction (Appendix Table A5). Shortening the return window to $[-4, -1]$ or extending it to $[-6, -1]$ shows that the finding does not depend on the exact five-day window. Separately removing analyst EPS forecast days and macroeconomic-news days from the return calculation leaves the interaction positive and significant, suggesting that returns on these public-information days do not fully explain the result. Using Fama–French three-factor-adjusted BHAR or CAPM-adjusted cumulative abnormal returns addresses sensitivity to the expected-return model and the method of aggregating abnormal returns. Changing the matching ratio from 1:5 to 1:4 or 1:6 checks whether the finding depends on the number of comparison events selected for each covered announcement. Across these specifications, the estimated incremental ERC ranges from approximately 0.019 to 0.037 and remains significant at least at the 10% level. Permuting coverage within firms or year-quarters rarely produces coefficients as large as the observed estimate (Appendix Figure A6). This suggests that the stronger pre-announcement earnings response is tied to which firms Polymarket covers and when, rather than being typical of random coverage assignments with the same firm-level or quarterly coverage frequencies.

## 6.2 Heterogeneity in Earnings Information Incorporation

The framework predicts that stock investors learn more when prediction-market signals are easier to observe and more informative. Signal visibility facilitates acquisition, while information content determines how much investors can learn about earnings. We therefore examine whether these characteristics are associated with stronger pre-announcement earnings incorporation, as implied by H2 and H3.

We use market trading volume and total wallet profile views to proxy for signal visibility. More active markets and more visible traders may attract greater attention from stock investors. To capture information content, we use the trading-volume shares associated with liquidity-taking, repeated-trading, and multi-wallet activity. These measures draw on our earlier evidence that larger liquidity-taking trades are more profitable, particularly

when participants spread their activity across transactions or wallets. For each measure, we split covered announcements into high and low groups at the matched-sample median and estimate the following specification. The coefficients $\beta_H$ and $\beta_L$ are the groups' pre-announcement ERCs. Controls, fixed effects, and standard errors follow equation (9). The learning prediction is $\beta_H > \beta_L$, which we assess using tests of coefficient equality.

$$BHAR_{i,q}^{[-5,-1]} = \alpha + \beta_H \left(Rank_SUE_{i,q} \times HighGroup_{i,q}\right) + \beta_L \left(Rank_SUE_{i,q} \times LowGroup_{i,q}\right) + \delta\, HighGroup_{i,q} + \Gamma' X_{i,q} + f_i + f_q + \varepsilon_{i,q}. \tag{12}$$

Table 10 shows stronger pre-announcement earnings responses in the high group across all five characteristics. The high-group ERCs range from 0.035 to 0.047 and are significant at the 1% level; the low-group estimates are small and insignificant. All five high–low differences are significant at the 5% level. Stock prices therefore reflect more earnings information before announcements when Polymarket markets attract greater trading activity and profile attention, and when their trading composition suggests greater information content. These patterns are consistent with the framework's Bayesian-learning prediction: investors incorporate more earnings information when prediction-market signals are easier to observe and more informative.

# 7 Conclusion

Can small prediction markets inform much larger stock markets? Our framework and evidence suggest that they can. Polymarket's earnings markets contain information advantages beyond measured general skill and observed public news, and stock investors appear to learn from their signals. This learning has economically meaningful consequences: stock prices incorporate more earnings information before announcements, particularly when prediction-market signals are more visible and informative. The framework explains how these outcomes depend on platform features. Wallet pseudonymity can encourage informed participation by lowering identity costs, while public trading records make the resulting signals accessible to outside investors. Learning can also transmit noise, so a stock-price response alone does not establish improved informativeness. Taken together, the findings suggest that prediction markets can create public value by supplying useful information beyond their own participants.

The findings inform the CFTC's question about the public value of informed trading in prediction markets. Debates over insider trading weigh informational benefits against

concerns about fairness and market integrity (Bhattacharya, 2014). In our sample, the largest trade-level profit is 10,300 USDC, while insider candidates earn aggregate profits of 103,272 USDC. These limited gains are consistent with the framework's prediction that thin liquidity constrains information rents. Yet the information revealed through their trading can reach investors in much larger stock markets. Regulatory assessments should therefore consider both the distribution of gains within prediction markets and the information benefits to outside investors.

Our setting also offers an information-design perspective on real-world asset tokenization (Cong et al., 2026). Earnings contracts tokenize state-contingent claims tied to real-world corporate outcomes, linking blockchain trading to firms' information environments. They combine pseudonymous participation with publicly observable transactions, a feature described by Nakamoto (2008). Our framework shows why this combination matters: identity protection can encourage informed participation, while trading transparency facilitates outside learning. The value of tokenized markets may therefore include the information they generate for related markets, alongside improvements in trading and settlement.

# References


Abdi, F. and B. Wu (2018). Pre-FOMC information asymmetry. NYU Stern School of Business. Available at SSRN.

Acharya, V. V. and T. C. Johnson (2007). Insider trading in credit derivatives. *Journal of Financial Economics 84*(1), 110–141.

Akey, P., V. Grégoire, and C. Martineau (2022). Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics 143*(3), 1162–1184.

Aktuğ, E. and O. Torul (2026). Political capital: Real-time electoral probabilities and the value of corporate political alignment. *Working Paper*. March 2026.

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

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research 6*(2), 159–178.

Bartlett, R. and M. O'Hara (2026). Adverse selection in prediction markets: Evidence from kalshi. Cornell SC Johnson College of Business Research Paper, forthcoming; Stanford Public Law Working Paper, forthcoming; Available at SSRN.

Bernard, V. L. and J. K. Thomas (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research 27*, 1–36.

Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni (2025). The impact of generative AI on information processing: Evidence from the ban of ChatGPT in italy. *Journal of Accounting and Economics 80*(1), 101782.

Bhattacharya, U. (2014). Insider trading controversies: A literature review. *Annu. Rev. Financ. Econ. 6*(1), 385–403.

Blankespoor, E., E. deHaan, and I. Marinovic (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics 70*(2–3), 101344.

Cao, C., Z. Chen, and J. M. Griffin (2005). Informational content of option volume prior to takeovers. *The Journal of Business 78*(3), 1073–1109.

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

Chang, A. Y., X. Dong, X. Martin, and C. Zhou (2026). Ai democratization and trading inequality. *Journal of Accounting Research 64*(3), 1287–1331.

Chau, J. (2025). Accounting information usage and trading by retail investors: Evidence from integrated trading platform. *Journal of Accounting Research 63*(4), 1377–1452.

Cong, L. W., S. Mayer, and D. Rabetti (2026, May). Tokenizing real-world assets. White paper, Wharton Initiative on Financial Policy and Regulation.

Cowgill, B. and E. Zitzewitz (2015). Corporate prediction markets: Evidence from google, ford, and firm x. *The Review of Economic Studies 82*(4), 1309–1341.

Diercks, A. M., J. D. Katz, and J. H. Wright (2026). Kalshi and the rise of macro markets. NBER Working Paper 34702, National Bureau of Economic Research.

Easley, D., M. O'Hara, and P. S. Srinivas (1998). Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance 53*(2), 431–465.

Edmans, A., I. Goldstein, and W. Jiang (2012). The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance 67*(3), 933–971.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance 25*(2), 383–417.

Feigenbaum, J. J. and C. A. Shelton (2013). The vicious cycle: Fundraising and perceived viability in us presidential primaries. *Quarterly Journal of Political Science 8*(1), 1–40.

Feng, C. (2026). Minority report: Contrarian traders, prediction markets, and the return of post-earnings drift. Available at SSRN.

Forsythe, R., F. Nelson, G. R. Neumann, and J. Wright (1992). Anatomy of an experimental political stock market. *American Economic Review 82*(5), 1142–1161.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Goldstein, I., Y. Li, and C. Wang (2026). Learning from prediction markets: The transmission of information and noise to traditional assets. *Available at SSRN 7198379*.

Gomez Cram, R., Y. Guo, T. I. Jensen, and H. Kung (2025). Financial prediction markets: A new measure of earnings expectations. Available at SSRN.

Gomez Cram, R., Y. Guo, T. I. Jensen, and H. Kung (2026). Prediction market accuracy: Crowd wisdom or informed minority? Available at SSRN.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hirshleifer, D. and J. Sheng (2022). Macro news and micro news: Complements or substitutes? *Journal of Financial Economics 145*(3), 1006–1024.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance 54*(6), 2143–2184.

Hu, D., S. Markovich, and V. Zhang (2023). Information processing in a smart-contract-

based market: Evidence from a defi protocol. *Available at SSRN 3962517*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Pan, J. and A. M. Poteshman (2006). The information in option volume for future stock prices. *Review of Financial Studies 19*(3), 871–908.

Rabetti, D., J. Shao, and C. Zhang (2026). Beating the earnings game: Why do prediction markets outperform professional analysts? Available at SSRN.

Schafhäutle, S. G. and D. Veenman (2024). Crowdsourced forecasts and the market reaction to earnings announcement news. *The Accounting Review 99*(2), 421–456.

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

Wolfers, J. and E. Zitzewitz (2006). Interpreting prediction market prices as probabilities.

Ying, C. (2020). The pre-fomc announcement drift and private information: Kyle meets macro-finance. *Available at SSRN 3644386*.

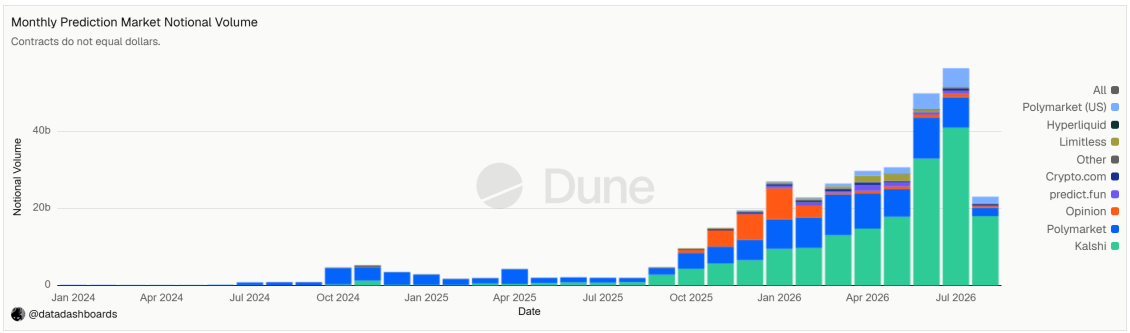
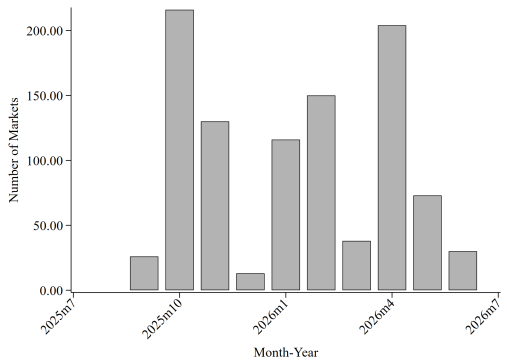
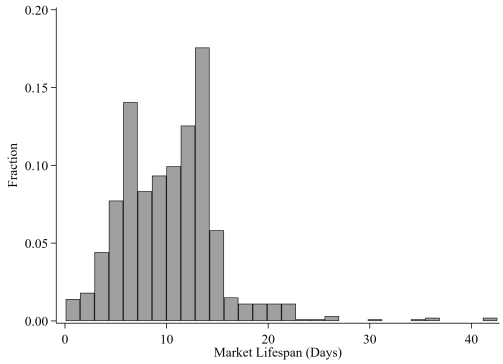



**Figure 1.** Monthly Prediction Market Notional Volume

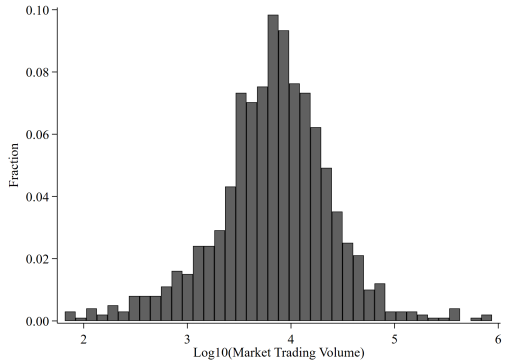
This figure plots monthly notional trading volume across major prediction-market platforms from January 2024 to June 2026. The underlying data are downloaded from Dune Analytics and then aggregated and visualized by the authors at the monthly frequency. The data cover Polymarket, Polymarket US, Kalshi, Limitless, Opinion, Predict, Crypto.com, Hyperliquid, and other prediction-market venues. For CTF-style order-book markets, notional volume is computed from executed liquidity-providing or liquidity-taking amounts, normalized by token decimals and converted into U.S. dollars when applicable; for platforms with reporting tables, the data use their reported daily or monthly notional volume. Because the figure combines on-chain transaction data with platform-level reporting datasets, measurement conventions may differ across venues. The figure should therefore be interpreted as a descriptive comparison of aggregate market activity rather than a fully harmonized measure of trading volume. Source: dune.com/queries/6689536/10534249.



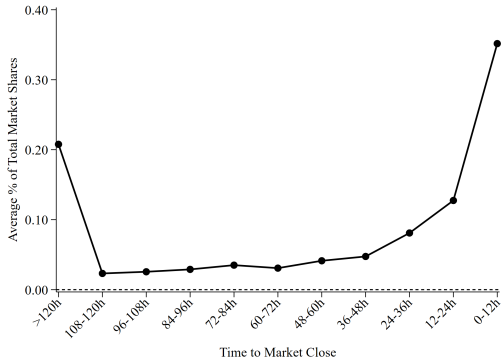

Panel A. Distribution of market start dates by month




Panel B. Contract duration distribution




Panel C. Distribution of trading volume




Panel D. Trading share by time-to-expiry bins

**Figure 2.** Trading Patterns of Polymarket Earnings Markets

This figure describes the coverage and trading activity of Polymarket earnings contracts. Panel A plots the distribution of market start dates by month. Panel B reports contract duration in days. Panel C shows the distribution of trading volume. Panel D reports the share of trading across time-to-expiry bins.

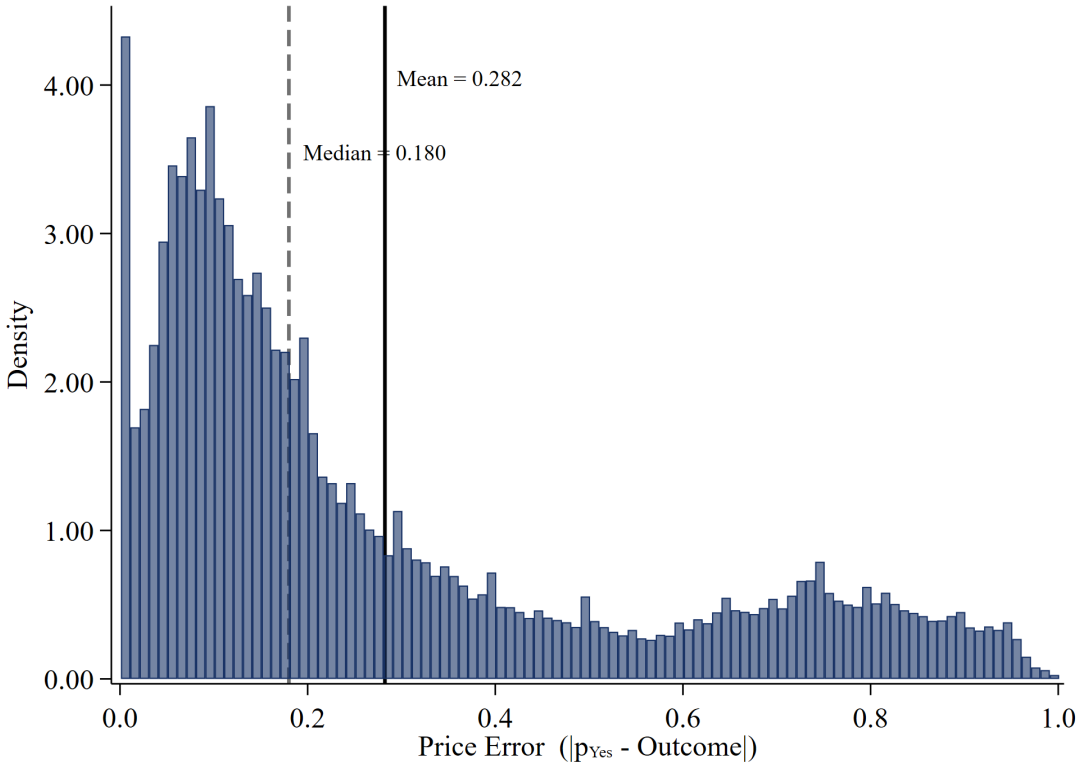



**Figure 3.** Distribution of Price Error

This figure plots the distribution of price error in Polymarket earnings markets. Price error is measured as the absolute deviation between the transaction price of the "Yes" token and the ex-post settlement value. Specifically, $Price\ Error_{i,m} = |p_{i,m}^{Yes} - Pay_m^{Yes}|$, where $Pay_m^{Yes} \in \{0, 1\}$. Lower error indicates greater accuracy. The sample excludes terminal sequences of near-certain trades, as described in Appendix 3.

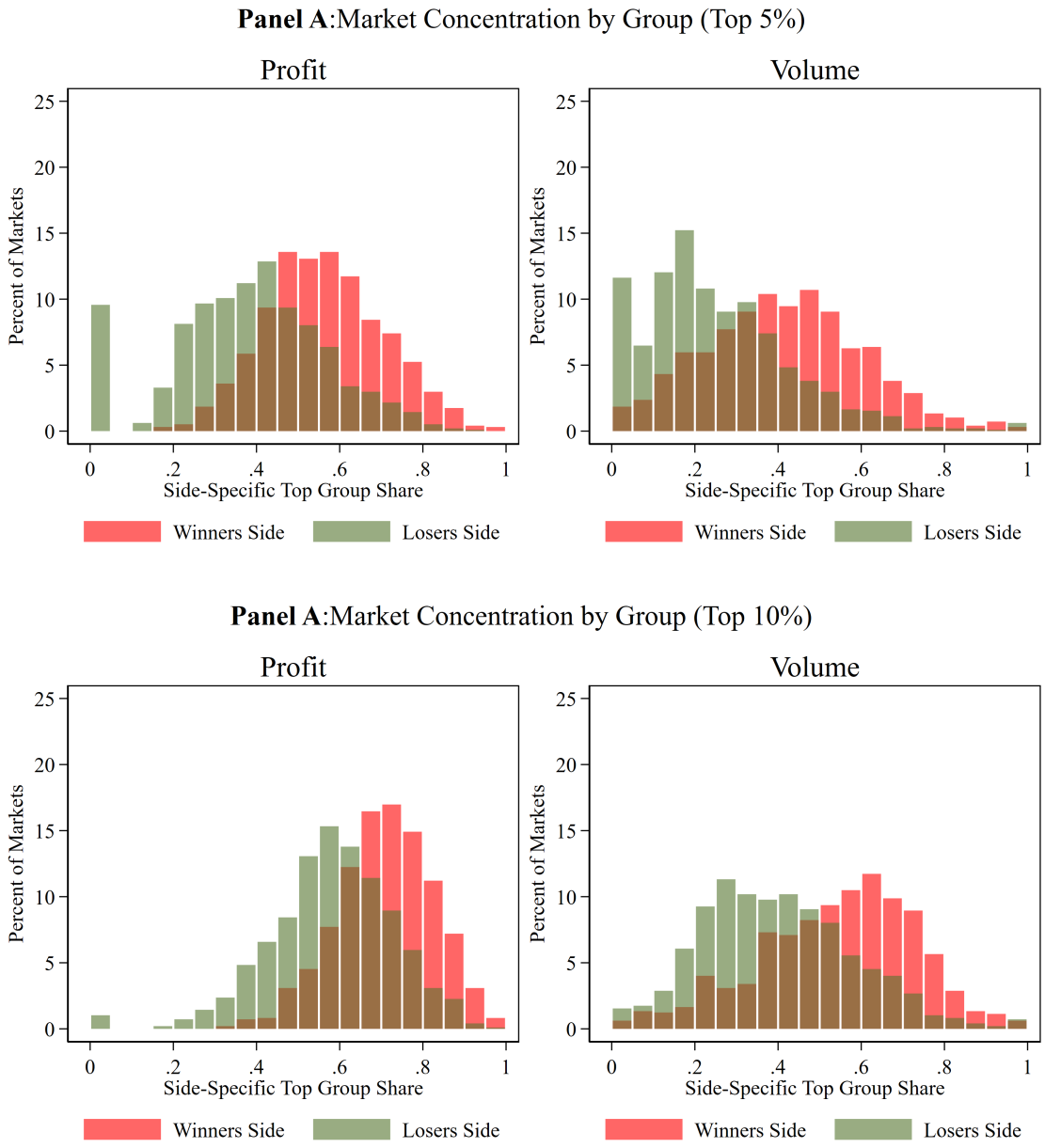



**Figure 4.** User Concentration on Winning and Losing Sides

This figure reports within-side user concentration across earnings prediction markets. Panel A uses the top 5% of users, and Panel B uses the top 10%. In each panel, the left plot reports profit concentration and the right plot reports volume concentration. User-level profit is the sum of trade-level PnL, and user-level volume is the sum of trade value. Winners are users with positive user-market profits, and Losers are users with negative user-market profits. Within each market and side, top users are selected by profit: the highest-profit Winners and the largest-loss Losers. Concentration is the top users' absolute PnL or trading volume divided by the corresponding side total. The sample is restricted to markets with at least 10 users on both sides.

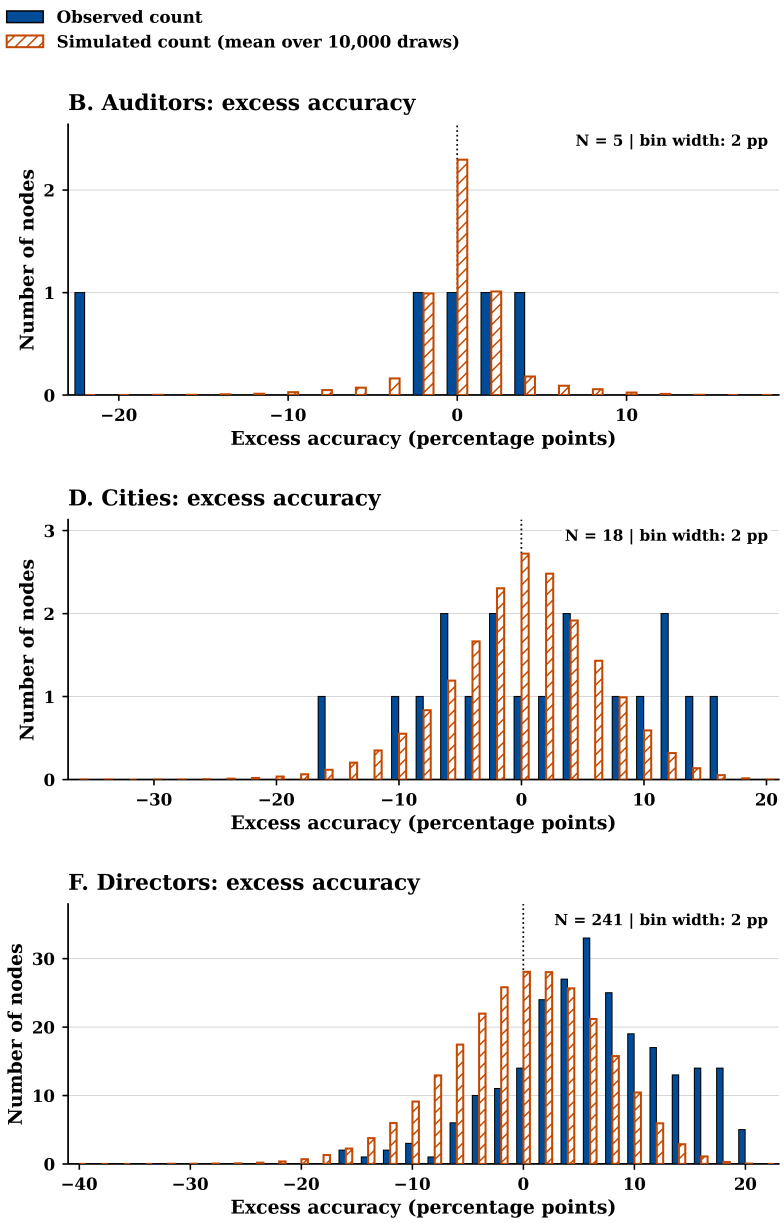



**Figure 5.** Market Accuracy across Auditor, City, and Director Networks

Panels B, D, and F compare observed and mean simulated distributions of excess market accuracy across 5 auditor, 18 city, and 241 director nodes, respectively; original panel identifiers are retained. Each node links at least 10 distinct markets, and node accuracy is the equal-weighted mean of their market accuracies. Market accuracy is $1-|\bar{p}_m - Pay_m^{Yes}|$, where $Pay_m^{Yes} \in \{0,1\}$ is the realized payoff and $\bar{p}_m$ equally weights available hourly closing YES prices over up to 72 hours ending at the last retained trade. Terminal sequences of near-payoff trades are removed except for their first trade, as described in Appendix 3. Empty hours carry the preceding close forward; markets with shorter histories are retained. Excess accuracy subtracts the full-market mean of 70.45%. Blue bars give observed node counts in two-percentage-point bins; hatched bars give mean counts across 10,000 random permutations without replacement of accuracy scores over all 996 eligible markets, separately for each network. Node–market links and the number of markets per node remain fixed; nodes sharing a market receive the same reassigned accuracy for that market in each draw. The highest-accuracy auditor, in the rightmost occupied observed bin in Panel B, is the auditor referenced in the anonymized discussion in Appendix Table A9, based on the author-provided correspondence.

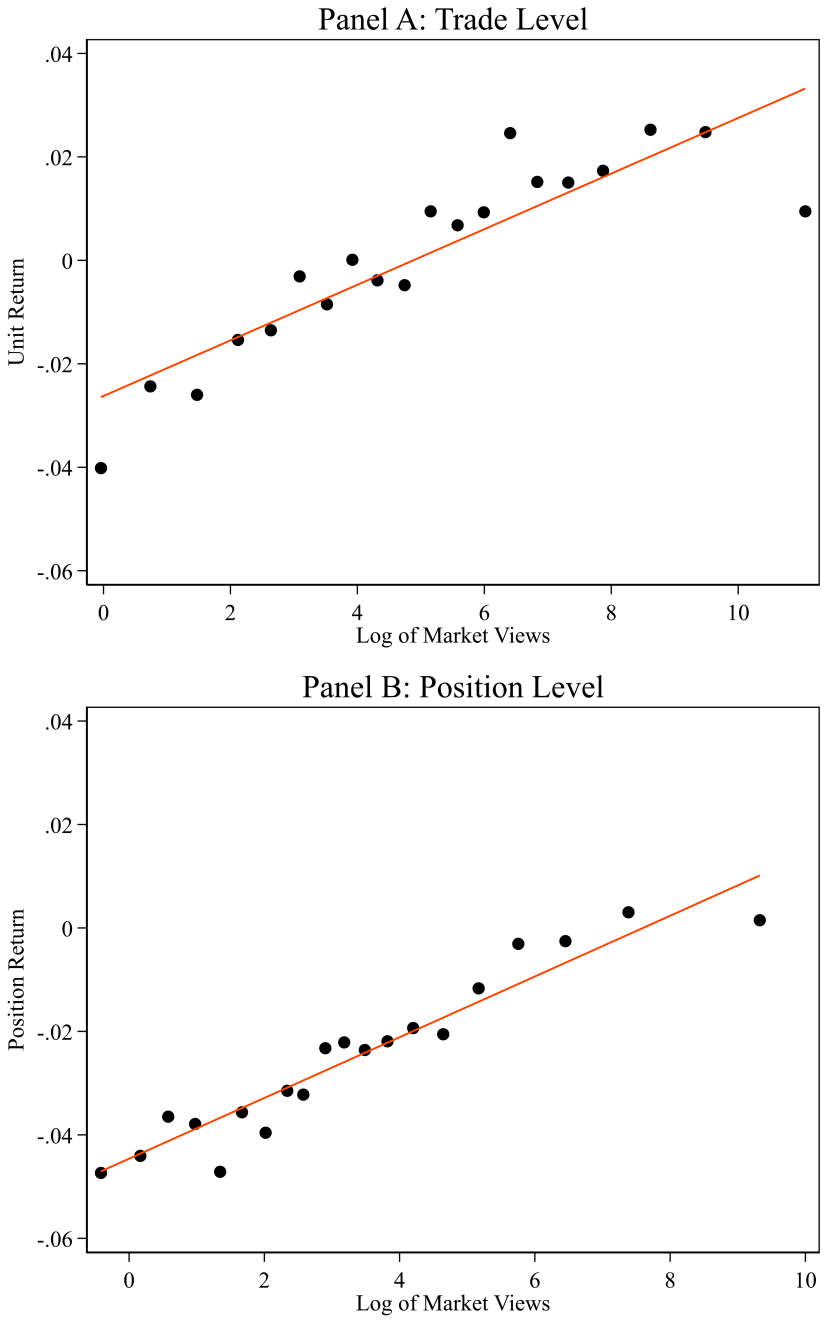



**Figure 6.** Market Attention and Trading Profitability

This figure presents binned scatterplots of trading profitability against ln(1 + Views), adjusting both variables for market fixed effects. Views are cumulative wallet-profile visits measured at data collection. Panel A uses order-level realized profit per token, in USDC per token. Panel B uses inferred user–market observations and measures position return as total realized profit divided by total trade value, with values below −1 set to −1; attention uses the maximum profile-view count across the user's linked wallets. Points show bin averages of the adjusted variables, and lines show linear fits.

## **Table 1.** Summary Statistics

Panel A reports individual trade records, the smallest observation unit in our trade data, with monetary values in USDC. An order denotes one execution record, rather than a unique order hash. Panel B reports unweighted statistics for the 2,522 earnings announcements from 2025 onward in the matched pre-announcement regression sample, after complete-case and fixed-effect singleton exclusions. BHAR is in decimal units; Covered is an indicator; SUE ranges from zero to one. Analysts, Firm Size, and News Coverage use $\ln(1+x)$; ROA, leverage, and institutional ownership are fractions; News Sentiment is in levels. Appendix 1 defines all variables.

**Panel A. Trade-level observations**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | SD | Min | P25 | P50 | P75 | Max |
| Is correct | 475,855 | 0.490 | 0.500 | 0 | 0 | 0 | 1 | 1 |
| PnL | 475,855 | 0.001 | 161.235 | -10,050 | -2.500 | -0.034 | 2.358 | 10,300 |
| PnL per Token | 475,855 | -0.000 | 0.386 | -0.999 | -0.177 | -0.004 | 0.180 | 0.999 |
| Return on capital | 475,855 | 0.125 | 10.654 | -999 | -1.000 | -0.010 | 0.408 | 999.000 |
| Value traded | 475,855 | 34.315 | 285.187 | 0.000 | 1.012 | 4.650 | 16.000 | 42,999.840 |
| ln(1 + value traded) | 475,855 | 1.956 | 1.429 | 0.000 | 0.699 | 1.732 | 2.833 | 10.669 |
| Tokens traded | 475,855 | 70.444 | 496.825 | 0.004 | 5.000 | 13.280 | 44.444 | 50,000.000 |
| ln(1 + tokens traded) | 475,855 | 2.828 | 1.449 | 0.004 | 1.792 | 2.659 | 3.816 | 10.820 |
| Views | 459,186 | 16,936.880 | 1,143,865 | 0 | 16 | 138 | 1,100 | 125,000,000 |
| ln(1 + Views) | 459,186 | 5.042 | 2.937 | 0 | 2.833 | 4.934 | 7.004 | 18.644 |
| Days to Market Close | 475,855 | 3.368 | 4.414 | 0 | 0.481 | 1.641 | 4.750 | 42.548 |
| Wallet Age | 475,855 | 196.556 | 227.437 | 0.005 | 25.976 | 103.798 | 318.634 | 2,008.180 |

**Panel B. Earnings announcements: PSM sample**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | SD | Min | P25 | P50 | P75 | Max |
| BHAR $[-5,-1]$ | 2,522 | -0.003 | 0.043 | -0.141 | -0.027 | -0.003 | 0.022 | 0.124 |
| BHAR $[0,+2]$ | 2,522 | -0.002 | 0.097 | -0.476 | -0.051 | -0.005 | 0.044 | 0.516 |
| SUE | 2,522 | 0.505 | 0.237 | 0.000 | 0.333 | 0.444 | 0.667 | 1.000 |
| Covered | 2,522 | 0.278 | 0.448 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| Analysts | 2,522 | 2.279 | 0.624 | 0.693 | 1.946 | 2.398 | 2.708 | 3.401 |
| Firm Size | 2,522 | 10.218 | 1.622 | 5.185 | 9.133 | 10.305 | 11.341 | 14.210 |
| $\text{ROA}_{t-1}$ | 2,522 | 0.012 | 0.034 | -0.195 | 0.003 | 0.012 | 0.024 | 0.116 |
| $\text{Leverage}_{t-1}$ | 2,522 | 0.337 | 0.248 | 0.004 | 0.159 | 0.307 | 0.452 | 1.347 |
| News Coverage | 2,522 | 6.901 | 1.139 | 4.394 | 6.118 | 6.665 | 7.514 | 10.337 |
| $\text{Institutional Ownership}_{t-1}$ | 2,522 | 0.796 | 0.216 | 0.045 | 0.726 | 0.840 | 0.928 | 1.172 |
| News Sentiment | 2,522 | 0.110 | 0.149 | -0.329 | 0.022 | 0.129 | 0.216 | 0.393 |

**Table 2.** Trading Aggressiveness, Size, and Profitability

This table reports order-level regressions of PnL per Token on trade size, separately for liquidity-taking and liquidity-providing orders. The specification is $PnL\ per\ Token_{i,m,d} = \alpha + \beta Size_{i,m,d} + f_m + f_d + \varepsilon_{i,m,d}$, where $i$, $m$, and $d$ denote order, market, and calendar day. PnL per Token is settlement profit in USDC divided by the absolute number of tokens traded. Size is measured by the natural logarithm of one plus trade value (USDC) or tokens traded. All regressions include market and calendar-day fixed effects. Market-clustered standard errors are in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%.

| *Dependent variable: PnL per Token* | Liquidity-Taking Orders | | Liquidity-Providing Orders | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| ln(1 + value traded) | 0.007*** | | -0.010** | |
| | (0.001) | | (0.004) | |
| ln(1 + tokens traded) | | 0.009*** | | -0.010*** |
| | | (0.003) | | (0.003) |
| Market FE | Yes | Yes | Yes | Yes |
| Calendar FE | Yes | Yes | Yes | Yes |
| Observations | 186,358 | 186,358 | 289,496 | 289,496 |
| Adjusted $R^2$ | 0.065 | 0.065 | 0.035 | 0.035 |

## **Table 3.** Informed Trading Beyond Skill

This table estimates order-level size–profitability regressions for liquidity-taking orders. Panel A retains traders whose skill indicator equals zero; Panel B retains those whose indicator equals one. Non-Earnings uses skill measured in non-earnings markets; Pre-September 15 uses the historical skill flag through September 15, 2025. Skill follows the event-level sign-randomization procedure in Appendix 3. PnL per Token is in USDC per token. Size is the natural logarithm of one plus trade value or tokens traded. All regressions include market and calendar-day fixed effects. Parentheses report market-clustered standard errors. ***, **, and * denote significance at 1%, 5%, and 10%.

**Panel A. Non-skilled sample**

| *Dependent variable: PnL per Token* | Non-Earnings | | Pre-September 15 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| ln(1 + value traded) | 0.0062*** | | 0.0063*** | |
| | (0.0012) | | (0.0011) | |
| ln(1 + tokens traded) | | 0.0087*** | | 0.0088*** |
| | | (0.0032) | | (0.0031) |
| Market FE | Yes | Yes | Yes | Yes |
| Calendar-Day FE | Yes | Yes | Yes | Yes |
| Observations | 176,646 | 176,646 | 181,394 | 181,394 |
| Adjusted $R^2$ | 0.070 | 0.071 | 0.067 | 0.068 |

**Panel B. Skilled sample**

| *Dependent variable: PnL per Token* | Non-Earnings | | Pre-September 15 | |
|---|---|---|---|---|
| | (5) | (6) | (7) | (8) |
| ln(1 + value traded) | 0.0037 | | 0.0048 | |
| | (0.0030) | | (0.0051) | |
| ln(1 + tokens traded) | | 0.0061* | | 0.0082 |
| | | (0.0034) | | (0.0060) |
| Market FE | Yes | Yes | Yes | Yes |
| Calendar-Day FE | Yes | Yes | Yes | Yes |
| Observations | 9,535 | 9,535 | 4,699 | 4,699 |
| Adjusted $R^2$ | 0.080 | 0.081 | 0.105 | 0.106 |

**Table 4.** Informed Trading on Days With and Without Identified Public-Information Releases

This table estimates size–profitability regressions for liquidity-taking orders on days without (Panel A) and with (Panel B) the specified release. Macroeconomic News Releases follow the macroeconomic news calendar; Analyst EPS Forecasts identify an I/B/E/S EPS forecast for the firm. Each split uses its own indicator, not the joint absence or presence of both categories. Orders use New York dates with daylight-saving adjustment; unmatched dates or firm-dates are coded zero, which does not establish the absence of other public information. PnL per Token is in USDC per token. Size is the natural logarithm of one plus trade value or tokens traded. All regressions include market and calendar-day fixed effects. Parentheses report market-clustered standard errors. ***, **, and * denote significance at 1%, 5%, and 10%.

**Panel A. Days without the Specified Release**

| *Dependent variable: PnL per Token* | Macroeconomic News Releases | | Analyst EPS Forecasts | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| ln(1 + value traded) | 0.0069*** | | 0.0059*** | |
| | (0.0011) | | (0.0014) | |
| ln(1 + tokens traded) | | 0.0092*** | | 0.0070*** |
| | | (0.0034) | | (0.0024) |
| Market FE | Yes | Yes | Yes | Yes |
| Calendar-Day FE | Yes | Yes | Yes | Yes |
| Observations | 161,964 | 161,964 | 118,895 | 118,895 |
| Adjusted $R^2$ | 0.068 | 0.068 | 0.081 | 0.081 |

**Panel B. Days with the Specified Release**

| *Dependent variable: PnL per Token* | Macroeconomic News Releases | | Analyst EPS Forecasts | |
|---|---|---|---|---|
| | (5) | (6) | (7) | (8) |
| ln(1 + value traded) | 0.0055** | | 0.0071*** | |
| | (0.0028) | | (0.0014) | |
| ln(1 + tokens traded) | | 0.0081** | | 0.0113** |
| | | (0.0041) | | (0.0047) |
| Market FE | Yes | Yes | Yes | Yes |
| Calendar-Day FE | Yes | Yes | Yes | Yes |
| Observations | 24,344 | 24,344 | 67,428 | 67,428 |
| Adjusted $R^2$ | 0.076 | 0.076 | 0.051 | 0.052 |

## **Table 5.** Hiding Strategies and Profitability

This table examines whether strategic concealment is associated with higher profitability among traders. Columns (1)–(2) report order-level regressions for liquidity-taking orders. Multi-wallet equals one if the clustered user employs more than one wallet address in a market. Multi-trade equals one if the user executes multiple transactions in a market. PnL per Token is in USDC per token. Both columns include market and calendar-day fixed effects. Standard errors, clustered at the market level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

| *Dependent variable: PnL per Token* | Order-Level Liquidity-Taking Orders | |
|---|---|---|
| | (1) | (2) |
| $\ln(1 + \text{value traded})$ | 0.005*** | 0.002 |
| | (0.001) | (0.002) |
| Multi-wallet | -0.003 | |
| | (0.009) | |
| $\ln(1 + \text{value traded}) \times$ Multi-wallet | 0.010*** | |
| | (0.003) | |
| Multi-trade | | 0.019** |
| | | (0.009) |
| $\ln(1 + \text{value traded}) \times$ Multi-trade | | 0.006*** |
| | | (0.002) |
| Market FE | Yes | Yes |
| Day FE | Yes | Yes |
| Observations | 186,358 | 186,358 |
| Adjusted $R^2$ | 0.065 | 0.066 |

**Table 6.** Market Attention and Trading Profitability

This table examines the relation between public attention and trading profitability. The dependent variable is PnL per Token, in USDC per token. Columns (1)–(3) use order-level observations for all executed orders, liquidity-taking orders, and liquidity-providing orders, respectively. ln(1 + Views) is the natural logarithm of one plus the cumulative number of views displayed on the trader's wallet profile at data collection on the Polymarket website, as illustrated in Appendix Figure A2. All specifications include market fixed effects. Standard errors, clustered at the market level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

| *Dependent variable: PnL per Token* | All Executed Orders | Liquidity-Taking Orders | Liquidity-Providing Orders |
|---|---|---|---|
| | (1) | (2) | (3) |
| ln(1 + Views) | 0.005*** | 0.006*** | 0.002*** |
| | (0.001) | (0.001) | (0.001) |
| Market FE | Yes | Yes | Yes |
| Observations | 459,186 | 173,212 | 285,973 |
| Adjusted $R^2$ | 0.001 | 0.058 | 0.028 |

## Table 7. Trader Characteristics and Polymarket Price Discovery

Panel A uses the full classification sample; token volume counts both trade sides, and users may span classes. Gross PnL is settlement profit before fees (USDC); PnL per Token is in cents, not a capital return. Panels B–C use liquidity-taking OIB over $t \to t+6h$. Panel B reports $100(p_b - p_a)$ in percentage points (pp), avoiding the small-denominator sensitivity of relative changes near zero. Panel C reports $100(|p_a - Y| - |p_b - Y|)$, where $Y$ is the YES settlement value; positive values indicate improved accuracy, and current OIB enters in absolute value. Controls are opening probability, lagged probability change, lagged log volume, log hours to cutoff, and signed lagged OIBs. Models include an intercept, no fixed effects, and column-specific samples. Parentheses contain standard errors clustered by event and New York signal-start date. Classification and Price Discovery are ex post. ***, **, and * denote 1%, 5%, and 10% significance.

**Panel A. Trading Activity and Profitability**

| | Skilled | Insider candidates | Noise traders |
|---|---|---|---|
| Users | 1,890 | 983 | 30,822 |
| Trades | 72,862 | 6,438 | 416,767 |
| Participant token volume | 7,011,778 | 532,651 | 31,992,336 |
| Token share (%) | 17.735 | 1.347 | 80.918 |
| Gross PnL (USDC) | 73,125 | 103,272 | -176,397 |
| PnL per Token (cents) | 1.043 | 19.388 | -0.551 |

**Panel B. Polymarket Price Responses**

| *Dependent variable:* *Price Change (pp)* | (1) $t \to t+6h$ | (2) $t+6h \to t+12h$ | (3) $t+12h \to t+18h$ |
|---|---|---|---|
| $OIB^S$: Skilled | 4.5535*** | 0.1070 | 0.1149 |
| | (0.4654) | (0.2462) | (0.2413) |
| $OIB^I$: Insider candidates | 5.2982*** | 0.5027 | 1.5375*** |
| | (0.5704) | (0.5906) | (0.4750) |
| $OIB^N$: Noise traders | 3.4485*** | -0.8037*** | -0.1857*** |
| | (0.1321) | (0.0792) | (0.0677) |
| Observations | 26,281 | 25,331 | 24,392 |
| $R^2$ | 0.1417 | 0.0070 | 0.0015 |
| Controls | Yes | Yes | Yes |

**Panel C. Price Discovery**

| *Dependent variable:* *Price Discovery (pp)* | (1) $t \to t+6h$ | (2) $t+6h \to t+12h$ | (3) $t+12h \to t+18h$ |
|---|---|---|---|
| $\lvert OIB^S \rvert$: Skilled | 1.0096** | -0.1870 | -0.4690 |
| | (0.4732) | (0.3110) | (0.2951) |
| $\lvert OIB^I \rvert$: Insider candidates | 3.0796*** | -0.7855 | -0.1519 |
| | (0.5740) | (0.5692) | (0.4092) |
| $\lvert OIB^N \rvert$: Noise traders | 0.0852 | -0.2151 | -0.4872** |
| | (0.2083) | (0.2228) | (0.1944) |
| Observations | 26,281 | 25,331 | 24,392 |
| $R^2$ | 0.0245 | 0.0154 | 0.0173 |
| Controls | Yes | Yes | Yes |

## Table 8. Polymarket Order Flow and Stock Returns

OIB aggregates liquidity-taking tokens across same-ticker contracts using a common denominator across groups. Signals are $O_d \to C_d$ in Panel A and $C_d \to O_{d+1}$ in Panel B, fixed across columns; $O_d$ and $C_d$ denote the trading-day open and close on the exchange calendar. Returns are $100 \log(P_b/P_a)$ (%), using TAQ NBBO midpoints strictly before endpoints and no older than 300 seconds. Column (1) is contemporaneous with the signal. Controls are stock return, stock OIB, log share volume, and opening spread from the latest completed regular session before signal start. All three OIBs enter jointly; models include an intercept, no fixed effects, and column-specific ticker–session samples. Parentheses contain standard errors clustered by ticker and New York signal-start date. Insider classification uses ex post settlement profits and entails look-ahead bias; skilled-flow associations may reflect common fundamental news. The cutoff proxies for announcement time; corporate-action adjustment remains unverified (Appendix 3). ***, **, and * denote 1%, 5%, and 10% significance.

**Panel A. Regular-Session Signals**

| *Dependent variable:* *Stock Log Return (%)* | (1) $O_d \to C_d$ | (2) $C_d \to O_{d+1}$ | (3) $O_{d+1} \to C_{d+1}$ |
|---|---|---|---|
| $OIB^S$: Skilled | 0.2463 | -0.0983 | -0.2914 |
| | (0.2969) | (0.3254) | (0.2967) |
| $OIB^I$: Insider candidates | -0.0158 | 0.6441 | -0.0551 |
| | (0.6103) | (0.4678) | (0.7966) |
| $OIB^N$: Noise traders | -0.0214 | -0.1316** | -0.1010 |
| | (0.0698) | (0.0661) | (0.0667) |
| Observations | 4,328 | 4,262 | 4,305 |
| $R^2$ | 0.0039 | 0.0020 | 0.0018 |
| Controls | Yes | Yes | Yes |

**Panel B. Overnight-Session Signals**

| *Dependent variable:* *Stock Log Return (%)* | (1) $C_d \to O_{d+1}$ | (2) $O_{d+1} \to C_{d+1}$ | (3) $C_{d+1} \to O_{d+2}$ |
|---|---|---|---|
| $OIB^S$: Skilled | 1.1633*** | 0.2802 | -0.0388 |
| | (0.3634) | (0.3260) | (0.1277) |
| $OIB^I$: Insider candidates | 0.2025 | -0.3813 | 2.3995*** |
| | (0.5796) | (0.6788) | (0.6021) |
| $OIB^N$: Noise traders | 0.1765** | 0.0534 | 0.0182 |
| | (0.0796) | (0.0735) | (0.0657) |
| Observations | 5,714 | 5,709 | 5,660 |
| $R^2$ | 0.0072 | 0.0023 | 0.0029 |
| Controls | Yes | Yes | Yes |

**Table 9.** Polymarket Coverage and Earnings Response Coefficients

Columns (1)–(2) report pre-announcement returns and columns (3)–(4) report announcement-window returns. Within each pair, the first column uses the Full Sample of eligible complete-case observations and the second uses the PSM Sample from 1:5 nearest-neighbor matching with a logit model, a 0.05 caliper, and common support. Matching uses analyst coverage, firm size, lagged ROA, lagged leverage, and news coverage. Outcomes are CAPM-adjusted buy-and-hold abnormal returns. SUE is the prior-quarter-based decile rank normalized to $[0, 1]$; Covered identifies an eligible earnings contract settling within four calendar days of the announcement. Variable definitions appear in Appendix 1. All specifications include analyst coverage, firm size, lagged ROA, lagged leverage, previous-quarter institutional ownership, news coverage, news sentiment, and firm and year-quarter fixed effects. Regressions are unweighted. Parentheses contain firm-clustered $t$-statistics. ***, **, and * denote significance at the 1%, 5%, and 10% levels.

| *Dependent variable: BHAR* | Pre-announcement $[-5, -1]$ | | Announcement $[0, +2]$ | |
|---|---|---|---|---|
| | (1) Full Sample | (2) PSM Sample | (3) Full Sample | (4) PSM Sample |
| SUE | 0.006*** | -0.001 | 0.091*** | 0.085*** |
| | (3.529) | (-0.136) | (21.482) | (5.919) |
| Covered | -0.013** | -0.013** | -0.002 | -0.008 |
| | (-2.331) | (-2.070) | (-0.144) | (-0.632) |
| SUE × Covered | 0.020** | 0.028** | 0.025 | 0.013 |
| | (2.040) | (2.477) | (1.078) | (0.526) |
| Analysts | 0.002 | 0.003 | -0.006** | 0.005 |
| | (1.397) | (0.846) | (-2.060) | (0.769) |
| Firm Size | -0.018*** | -0.026*** | -0.074*** | -0.065*** |
| | (-5.574) | (-4.473) | (-13.868) | (-4.493) |
| $\text{ROA}_{t-1}$ | 0.032* | 0.013 | -0.034 | -0.049 |
| | (1.721) | (0.177) | (-1.134) | (-0.356) |
| $\text{Leverage}_{t-1}$ | 0.008 | -0.018 | 0.035 | 0.014 |
| | (0.613) | (-0.490) | (1.046) | (0.140) |
| $\text{Institutional Ownership}_{t-1}$ | 0.007 | -0.052* | -0.002 | 0.074 |
| | (0.535) | (-1.942) | (-0.111) | (1.153) |
| News Coverage | -0.000 | 0.004 | -0.010*** | -0.008 |
| | (-0.293) | (1.097) | (-3.504) | (-1.000) |
| News Sentiment | -0.006 | 0.007 | -0.033*** | -0.039* |
| | (-1.205) | (0.782) | (-3.535) | (-1.800) |
| Firm FE | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Observations | 16,453 | 2,522 | 16,448 | 2,522 |
| Adjusted $R^2$ | 0.063 | 0.067 | 0.088 | 0.074 |

**Table 10.** Heterogeneity in Pre-announcement Price Discovery

The dependent variable is pre-announcement BHAR over $[-5, -1]$, in decimal units. Volume is total market trading volume in USDC, and Total Views sums public profile views across wallet–market observations. Liquidity-Taking is the market trading-volume share of wallets with more than 90% of token volume from liquidity-taking orders. Multi-Trade restricts this share to such wallets with at least two distinct transaction hashes in the market; Multi-Wallet restricts it to such wallets belonging to an inferred user cluster with at least two wallets in the market. These characteristics are measured for the treated group, namely firm-quarter observations with Polymarket coverage. Within the treated group, High and Low split nonmissing values at the matched-working-sample median before regression-specific exclusions (ties enter High). Both equal zero for missing values; these observations remain eligible for estimation but enter neither SUE interaction. SUE: High is the interaction between SUE and the High indicator, and SUE: Low is the interaction between SUE and the Low indicator. The reported $p$-value tests equality of the High and Low coefficients. Regressions include the seven controls in Table 9, firm and year-quarter fixed effects, and firm-clustered standard errors. Control coefficients are omitted for brevity. ***, **, and * denote significance at 1%, 5%, and 10%.

| *Dependent variable: BHAR* | (1) Volume | (2) Total Views | (3) Liquidity-Taking | (4) Multi-Trade | (5) Multi-Wallet |
|---|---|---|---|---|---|
| SUE: High | 0.039*** | 0.047*** | 0.035*** | 0.035*** | 0.037*** |
| | (0.013) | (0.012) | (0.012) | (0.012) | (0.013) |
| SUE: Low | 0.005 | 0.003 | 0.005 | 0.005 | 0.004 |
| | (0.007) | (0.007) | (0.006) | (0.007) | (0.007) |
| $p$: High = Low | 0.016 | 0.001 | 0.027 | 0.028 | 0.024 |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Firm FE | Yes | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,522 | 2,522 | 2,522 | 2,522 | 2,522 |
| Adjusted $R^2$ | 0.067 | 0.070 | 0.066 | 0.066 | 0.067 |

# Appendix 1: Variable Definitions

Definitions distinguish execution records, wallets, inferred users, markets, and earnings announcements. Code fields appear in monospaced type. Appendix 3 provides construction procedures.

## Stock-Market and Corporate Control Variables

$\texttt{BHAR}^{w}_{i,q}$**:** Buy-and-hold abnormal return for firm $i$ in fiscal quarter $q$, in decimal units, over window $w$ relative to the announcement (day 0). The windows are $[-5,-1]$ and $[0,+2]$. The baseline benchmark is CAPM; robustness tests use Fama–French three-factor returns. Factor loadings are estimated over $[-250,-21]$.

`CAR`**:** Cumulative abnormal return, calculated by summing daily actual returns minus CAPM expected returns over the specified event window. Used over $[-5,-1]$ as an alternative to compounded BHAR.

$\texttt{SUE}_{i,q}$**:** Standardized unexpected earnings for firm $i$ in quarter $q$. Raw SUE is actual reported GAAP/non-GAAP EPS minus the latest I/B/E/S consensus available before disclosure, scaled by the pre-announcement stock price. In the descriptive statistics and regressions, SUE denotes the normalized decile rank of this raw measure: we use preceding-calendar-quarter breakpoints and transform decile $d$ to $(d-1)/9$, for $d = 1,\ldots,10$. The rank therefore does not rely on earnings realizations announced later in the current quarter.

$\texttt{Covered}_{i,q}$**:** A binary indicator equal to one if the settlement date of an eligible Polymarket earnings contract is within four calendar days of firm $i$'s earnings announcement date in quarter $q$, that is, $|date_gap| \leq 4$, and zero otherwise.

`Analysts`**:** The natural logarithm of one plus the number of analysts contributing to the I/B/E/S earnings consensus, $\ln(1 + numest)$ (`num_analyst` in the code).

**Firm Size:** The natural logarithm of one plus market equity capitalization, $\ln(1 + mcap)$ (`size` in the code).

$\texttt{ROA}_{t-1}$**:** Net income divided by total assets in the prior fiscal quarter.

$\texttt{Leverage}_{t-1}$**:** Lagged financial leverage, defined as the sum of short-term and long-term debt divided by total assets from the prior fiscal quarter.

$\texttt{Institutional Ownership}_{t-1}$**:** The previous quarter's fraction of outstanding shares held by institutional investors, obtained from FactSet (`inst_ownership`).

`News Sentiment`**:** Story-count-weighted mean of available RavenPack event sentiment scores over calendar days $[-90,-7]$ relative to the Compustat earnings announcement date. The numerator sums daily mean scores multiplied by scored-story counts; the denominator is the total scored-story count. Entered in levels and missing when no scored

stories are available.

`News Coverage`: The natural logarithm of one plus the number of firm-specific news reports in RavenPack over calendar days $[-90, -7]$ relative to the earnings announcement, $\ln(1 + \textit{news_count_pre90_7})$ (`news_coverage`). The count includes reports on both endpoint days.

## Order-Level Variables (Polymarket)

**Correct:** A binary indicator equal to one if the traded token direction aligns with the ex post settlement outcome, and zero otherwise (`Is_correct`); for example, a long position in a winning YES contract is coded one.

`PnL`: Order-level profit or loss denominated in USDC, calculated as the signed quantity multiplied by the difference between the terminal contract settlement value ($1 or $0) and the transaction price.

**PnL per Token:** Order-level unit profitability, defined as PnL divided by absolute tokens executed in that order, in USDC per token (`PnL_per_Token`). This is not a return on invested capital.

`val_traded`: The executed trade value in USDC.

$\ln(1 + \textbf{\textit{val_traded}})$: The natural logarithm of one plus trade value in USDC.

`tokens_traded`: The absolute number of tokens in the execution record.

$\ln(1 + \textbf{\textit{tokens_traded}})$: The natural logarithm of one plus the absolute number of tokens traded.

`Views`: Cumulative visits to the trader's public profile at data retrieval.

$\ln(1 + \textbf{\textit{Views}})$: The natural logarithm of one plus the profile view count. This measures public attention at retrieval, not attention known before each order.

**Days to Market Close:** Days from order execution to scheduled market closure or settlement (`tenure_mkt`). This is a remaining duration, not elapsed market age.

**Wallet Age:** Days between the executing wallet's account creation date and the transaction timestamp (`user_exp`). This order-time measure differs from Account-Date Gap in the user–market classification.

## Trading Behavior and Skill

`PnL`: The clustered user's cumulative realized profit or loss within a specific contract market, aggregated across all constituent wallet addresses and transactions, denominated in USDC.

**PnL per Token:** User–market profit divided by the user's cumulative absolute tokens in

that market. In the high-frequency classification ledger, profit is gross trade-attributed settlement PnL before fees (`pnl_per_token`). Appendix Table A3 reports the ratio in cents per token by multiplying it by 100.

`n_trades`**:** The total number of individual transactions executed by the user cluster within a specific market.

`n_wallets`**:** The number of active wallets assigned to an inferred user within a market using ex post direct-transfer links on Polygon.

**Multi-wallet:** An indicator for more than one active wallet belonging to the same inferred user within a contract market (`multi_wallet`; `n_wallets` $> 1$).

**Multi-trade:** An indicator for multiple transactions by the same inferred user within a contract market (`multi_trade`; `n_trades` $> 1$).

**Liquidity-Taking Share:** The share of a user's absolute token volume executed through liquidity-taking orders within a contract (`Taker_Ratio`). The denominator includes both liquidity-taking and liquidity-providing orders.

`vol_sum`**:** The user cluster's total absolute token volume (sum of buy and sell token quantities) within the contract market.

`Skill Before Earnings`**:** An indicator for trading skill measured using activity through September 15, 2025. Eligibility requirements and the event-level sign-randomization procedure appear in Appendix 3.

`Skill Non-Earnings`**:** An indicator for trading skill measured exclusively in non-earnings markets and assigned to linked wallets and their orders. Excluding earnings contracts does not impose a historical cutoff. The classification procedure and its high-frequency application appear in Appendix 3.

**Directionality:** Absolute net YES-equivalent tokens divided by cumulative absolute tokens for an inferred user within a market (`directionality`). Buying YES or selling NO is positive; selling YES or buying NO is negative.

**Account-Date Gap:** The market's first observed trade date minus the earliest available profile-creation or first-outward-transfer date across the user's linked wallets, in New York calendar days (`age_user_days`). This is not Wallet Age at order execution.

## Public-Information Day Indicators

The indicators below use the order's New York calendar date, converted from UTC with daylight-saving adjustments. Unmatched dates or firm-dates are coded zero. Each indicator defines a separate split between days with and without the identified public-information release in Section 4.3.

**Macroeconomic News Releases:** One if the date appears in the macroeconomic news

calendar used in the analysis, and zero otherwise.

**Analyst EPS Forecasts:** One if the I/B/E/S unadjusted EPS detail file records at least one forecast announcement for the firm's eight-character CUSIP on that date, and zero otherwise.

## High-Frequency Signals and Returns

**Trader classes:** Skilled traders, insider candidates, and noise traders are mutually exclusive user–market groups used to compare trading characteristics and price responses. Complete rules appear in the classification table in Appendix 3. The classification uses ex post histories and settlement profits and is not a signal known at trade time.

**Group order imbalance (OIB):** Net YES-equivalent liquidity-taking tokens for a group divided by total liquidity-taking tokens across all groups in a six-hour Polymarket bin or a regular/overnight stock session. Stock-level signals aggregate tokens across contracts for the same ticker before division.

**Polymarket price change:** $100(p_b - p_a)$, the change in YES probability in percentage points. Unlike a relative price change, it does not divide by the initial probability.

**Price Discovery:** $100(|p_a - Y| - |p_b - Y|)$, the reduction in absolute error relative to the final YES settlement value $Y$. Positive values indicate improved accuracy. The measure is ex post; its regression uses absolute current group OIB and signed lagged OIB controls.

**High-frequency stock return:** $100 \log(P_{\text{end}}/P_{\text{start}})$, where $P$ is the TAQ NBBO midpoint. Returns are in percent. Appendix 3 defines the signal and return windows.

## Market and Network Accuracy

**Hourly Accuracy:** One minus the absolute difference between the equally weighted mean hourly closing YES price and the realized binary payoff. The window contains up to 72 elapsed hours ending at the last retained trade after terminal-trade cleaning; empty hours carry forward the preceding close, and shorter histories are retained.

**Node Accuracy:** The unweighted mean of hourly market accuracy across distinct markets linked to an auditor, headquarters city, or director identifier. Nodes must have at least ten markets. Director links use the supplied BoardEx company-person mapping.

**Excess Accuracy:** Node accuracy minus the unweighted mean accuracy across all 996 eligible markets. Observed and simulated excess accuracy use the same reference mean.

# Appendix 2: Platform Background and Data Illustrations

## Earnings Markets within Polymarket

**Table A1.** Polymarket Event-Level Summary by Main Topic

Topics are assigned from event tags and titles using text embeddings, clustering, and LLM labels. N Events counts unique events; Repeat is the share in repeated series; Duration is median days; Volume is median USDC volume. Earnings events are identified from titles and reported separately. The descriptive universe is measured as of August 18, 2026.

| Category | (1) N Events | (2) Repeat (%) | (3) Duration | (4) Volume |
|---|---|---|---|---|
| Crypto | 644,830 | 4.10 | 1.00 | 4,071.23 |
| Sports | 145,070 | 21.79 | 7.66 | 8,369.42 |
| Esports | 23,475 | 8.72 | 1.09 | 16,940.86 |
| Finance | 7,412 | 70.75 | 6.98 | 10,434.73 |
| Weather | 10,831 | 95.07 | 2.30 | 61,000.22 |
| Politics | 9,795 | 14.41 | 34.29 | 42,762.49 |
| Geopolitics | 2,809 | 34.14 | 36.23 | 163,675.95 |
| Culture | 2,787 | 23.14 | 10.10 | 30,709.75 |
| Technology | 5,524 | 75.09 | 1.33 | 6,823.95 |
| Earnings | 1,380 | 80.67 | 12.41 | 6,311.59 |

# Contract Interface and Public Profiles

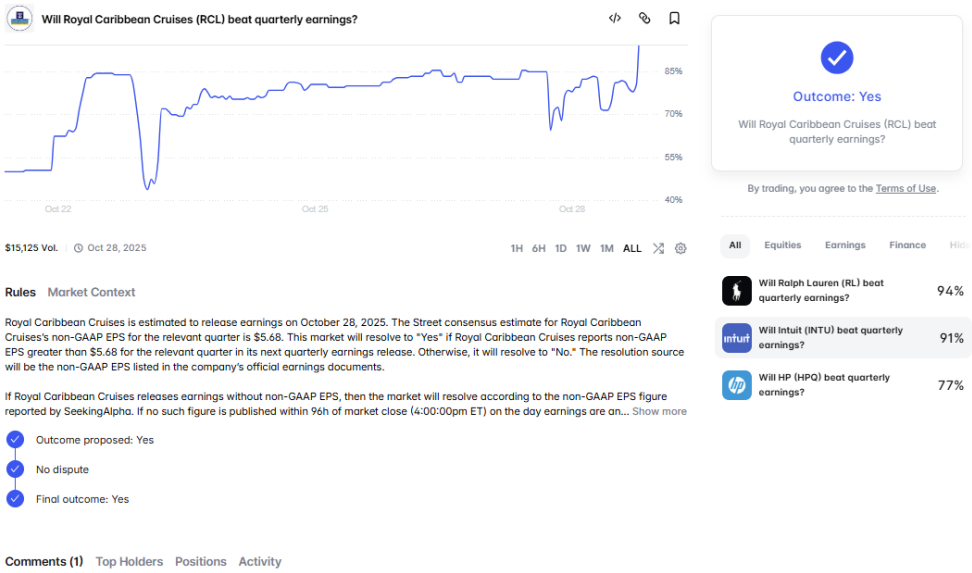



**Figure A1.** Interface Snapshot of the Earnings Prediction Market
The Royal Caribbean contract displays its YES probability, price history, EPS benchmark and resolution source, oracle proposal and dispute status, discussions, holders, trades, and positions. Source: Polymarket contract page.

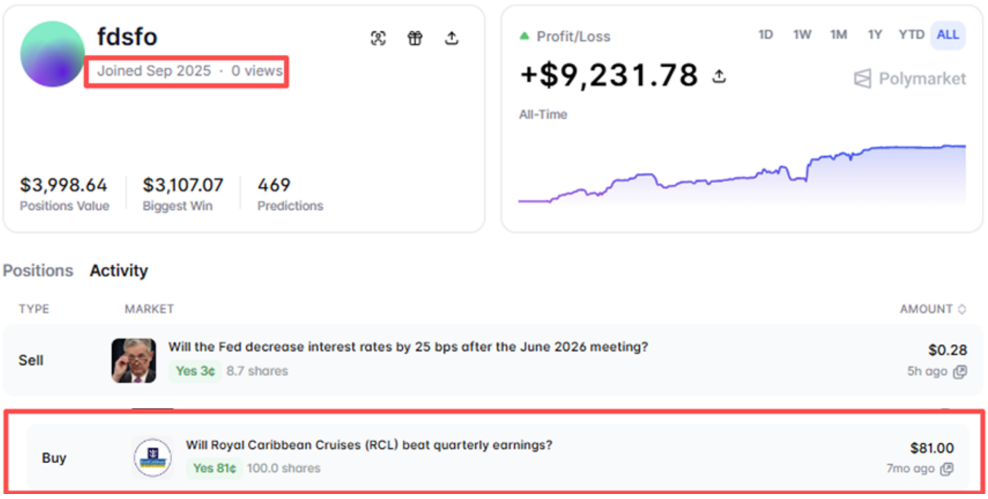



**Figure A2.** User Profiles on the Polymarket Platform
Retrieved May 7, 2026. Profiles display usernames, views, positions, and histories. Wallet 0x23 bought 100 YES tokens for 81 USDC in the validation example. Its March 2026 profile showed "Yktvrev" and 187 views; the May snapshot shows "fdsfo" and zero views. Wallet addresses remain fixed, but usernames and profile paths can change and may affect view counts. Source: wallet profile.

## Transaction Records and Validation

The table and screenshot document the same transaction. The final figure illustrates off-platform funding transfers used in clustering.

**Table A2.** A Case of Filled Order Records from The Graph

*Notes:* Identifiers use four-character prefixes: condition 0x0c, transaction 0x5b, and wallets 0x07, 0x3d, 0xa3, and 0x23. Address 0x4b is the Polymarket Operator. Under Conditional Token Framework clearing, raw `maker` is the EIP-712 limit-order signer; raw `taker` can be the operator initiating settlement, so these fields require interpretation before assigning economic liquidity roles. Here `makerAssetId`=0 denotes USDC supplied to buy tokens. See the subgraph specification and the linked transaction below for complete identifiers.

| (1) Condition ID | (2) Transaction hash | (3) Timestamp | (4) Token | (5) Price | (6) Stake (USDC) | (7) Shares (Token) | (8) Side | (9) `maker`/ `taker` |
|---|---|---|---|---|---|---|---|---|
| 0x0c | 0x5b | 2025-10-21 23:47:26 | NO | 0.19 | 5.7 | 30 | Buy | 0x07 / 0x23 |
| 0x0c | 0x5b | 2025-10-21 23:47:26 | NO | 0.19 | 3.8 | 20 | Buy | 0x3d / 0x23 |
| 0x0c | 0x5b | 2025-10-21 23:47:26 | NO | 0.19 | 9.5 | 50 | Buy | 0xa3 / 0x23 |
| 0x0c | 0x5b | 2025-10-21 23:47:26 | YES | 0.81 | 81.0 | 100 | Buy | 0x23 / 0x4b |

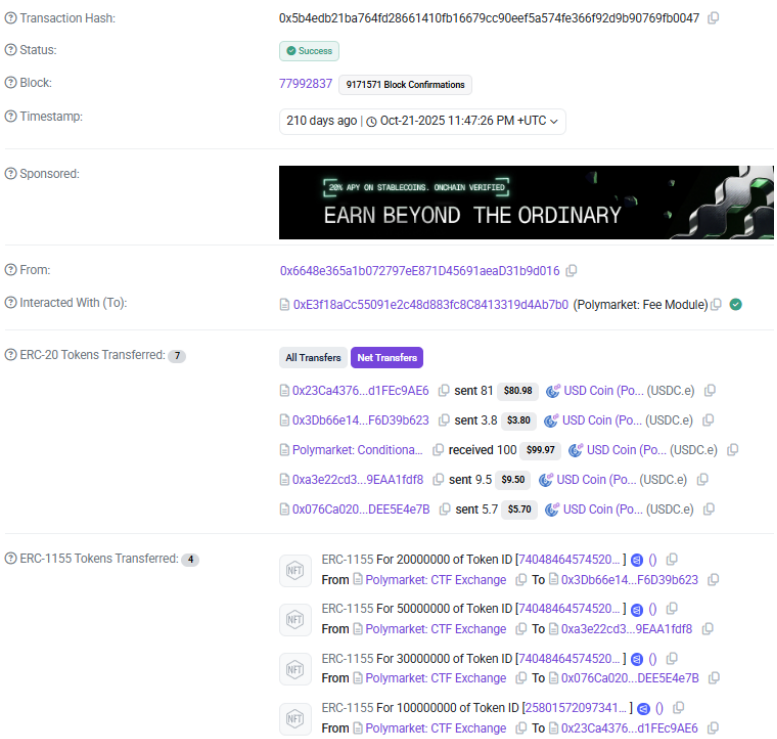



**Figure A3.** Example of a Transaction Record on Polygonscan

Block 77,992,837, October 21, 2025, 23:47:26 UTC. Wallets 0x07, 0x3d, and 0xa3 supply NO-side liquidity; 0x23 supplies YES-side collateral. Their 100 USDC creates 100 complete sets: 100 YES tokens go to 0x23, and 100 NO tokens are split among the other three wallets. Operator 0x4b settles the transaction. Source: PolygonScan transaction 0x5b.

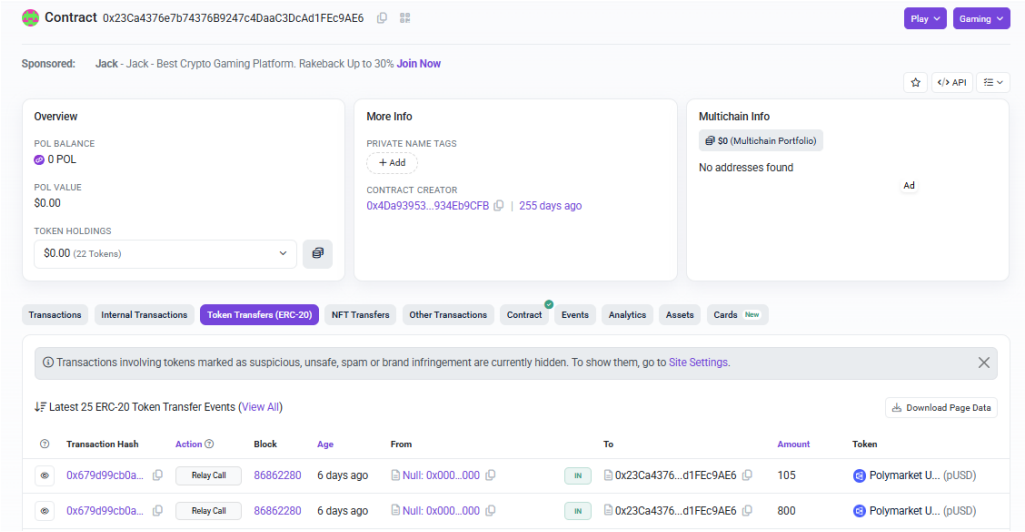



**Figure A4.** Sample of Transaction Records Used for Wallet Clustering
PolygonScan token transfers show wallet inflows and outflows. Clustering retains direct funding links outside Polymarket after excluding official platform addresses. Source: wallet 0x23 transfer history.

# Appendix 3: Data Construction and Empirical Methods

## Polymarket Data

We combine event metadata, execution records, and wallet profiles. Appendix Table A2 illustrates the raw execution data.

The Gamma API supplies `event_id`, `market_id`, `condition_id`, `clobTokenIds`, descriptions, start and end dates, outcome labels and prices, and market status.[8] We merge executions on `condition_id` and use settlement fields to calculate payoffs, accuracy, and profits. We select questions matching "Will [Company Name] ([Ticker]) beat quarterly earnings? and parse names and tickers for links to Compustat, CRSP, and I/B/E/S.

An order denotes one execution-data row, not a unique order hash. We retrieve historical filled orders separately from The Graph's V1 and V2 Polymarket orderbook subgraphs, normalize their fields to a common format, and combine the records by contract.[9] For each filled order, we observe the `condition_id`, wallet addresses for liquidity-providing and liquidity-taking orders, transaction price, traded quantity, timestamp, token side, and transaction hash.

For terminal-trade cleaning, we work backward from each market's final transaction and identify consecutive trades with absolute YES-price error below one percentage point. We retain only the first trade in this sequence. The remaining trades contribute little to price discovery and can mechanically inflate measured accuracy during the delay before oracle settlement.

Wallet addresses identify participants before clustering. The Gamma API supplies account creation timestamps and usernames when available; public profile pages supply cumulative views at retrieval.[10] Appendix Figure A2 illustrates these fields and the sensitivity of view counts to profile changes.

The final sample contains 996 closed earnings markets (September 2025–June 2026), 475,855 execution records, and 144,235 inferred user–market observations. Figure 2 reports coverage, duration, volume, and trading timing. Appendix Table A1 covers a distinct platform-wide descriptive universe.

## Corporate Data

From I/B/E/S Detail History, we retain the latest quarterly EPS forecast issued by each analyst before the announcement (FPI 6 or 7). For each firm-quarter, we select the primary or diluted reporting basis with the highest analyst coverage.

[8] `https://gamma-api.polymarket.com/events`.

[9] The queried orderbook subgraphs are V1 and V2.

[10] Profiles use URLs of the form `https://polymarket.com/profile/\{wallet_address\}`.

SUE is actual EPS minus consensus, scaled by the pre-announcement stock price. We convert it to normalized decile ranks using preceding-calendar-quarter breakpoints, as defined in Appendix 1.

CRSP daily prices and returns incorporate adjustment factors for splits and distributions. We calculate BHAR over $[-5, -1]$ and $[0, +2]$, using CAPM expected returns and Fama–French three-factor returns as a robustness benchmark. Factor loadings use $[-250, -21]$ relative to the announcement.

The stock regressions include previous-quarter institutional ownership from FactSet and prior news coverage and sentiment from RavenPack, alongside analyst coverage, firm size, lagged ROA, and lagged leverage. Appendix 1 defines the measures, transformations, and news windows.

For the order-level news-day tests, we match trades to macroeconomic news dates and firm-specific I/B/E/S EPS forecast dates. Trade dates use New York time with daylight-saving adjustments; macro news is merged by date and analyst news by eight-character CUSIP and date. Unmatched observations receive a zero indicator. Each category defines a separate sample split, estimated with market and calendar-day fixed effects and market-clustered standard errors.

For the network analysis in Section 4.6, we link firms to auditors using the WRDS audit-fee records, headquarters cities from Compustat, and company-person relationships from BoardEx. Auditor and BoardEx links are matched through CIK; the director network retains the supplied company-person relationships without restricting the sample to the information-access flag. These are firm-link groupings, not links between corporate insiders and trading wallets. We average the hourly market-accuracy measure across each node's distinct markets and require at least ten markets per node.

## Earnings-Announcement Matching and PSM

We link eligible Polymarket earnings contracts to firm earnings announcements by GVKEY and require settlement within four calendar days of the announcement. We exclude markets with less than 100 USDC in total volume and those at or below the 5th percentile of contract duration (3.83 days). When multiple contracts match an announcement, we select the nearest settlement and break ties by volume. This procedure yields 812 covered announcements.

We use 1:5 nearest-neighbor propensity-score matching with a 0.05 caliper and common support, based on analyst coverage, firm size, news coverage, lagged ROA, and leverage. The matched outcome regressions contain 2,522 observations in each window. Panel B of Table 1 uses the pre-announcement estimation sample. Following the saved specification, pre-announcement BHAR and the seven controls are winsorized at the 1st and 99th percentiles within the matched working sample; descriptive statistics reflect these transformed values.

Equation (9) uses firm and year-quarter fixed effects and the seven controls defined in

Appendix 1. Appendix 5 reports matching balance and additional earnings-response tests.

## Wallet Clustering and User Identity

We use PolygonScan normal-transaction and ERC20 token-transfer histories to infer users from direct funding links outside Polymarket. Official Polymarket addresses are excluded, and links must connect two wallets in the earnings-market sample. Appendix Figure A4 illustrates the source records.

Each sample wallet is a node. Counterparties in the `from` and `to` fields form undirected edges when both wallets belong to the sample. We exclude wallets with more than 500 connections to limit public hubs, then assign connected components to flow clusters. Components with at least two wallets are multi-wallet clusters. Links require direct transfers; shared third-party counterparties do not establish a link.

The resulting network contains 27,915 Polymarket wallets and 1,782 direct wallet-pair links. We identify 280 multi-wallet clusters containing 978 wallets in total, with the largest cluster containing 232 wallets. Figure A5 visualizes the direct-transfer network for these identified clusters.

Each connected component receives a user identifier, with unmatched wallets retained as separate users. Within a market, we count active wallets and flag users with more than one wallet. Transfer links approximate common control or coordination; missed or spurious links can affect the grouping.

The high-frequency ledger contains 515,405 owner-order records before excluding 37 records in 29 exceptional groups, leaving 515,368. These counts differ from the original filled-order sample, which applies its own terminal-trade screen. High-frequency regressions further use horizon-specific samples.

## Trader Classification and Order Imbalance

The table below defines three mutually exclusive user–market groups using the full retained trading history. Flow-linked wallets form inferred users; unlinked wallets remain separate users. Skilled status takes precedence, followed by insider candidates and the residual noise group.

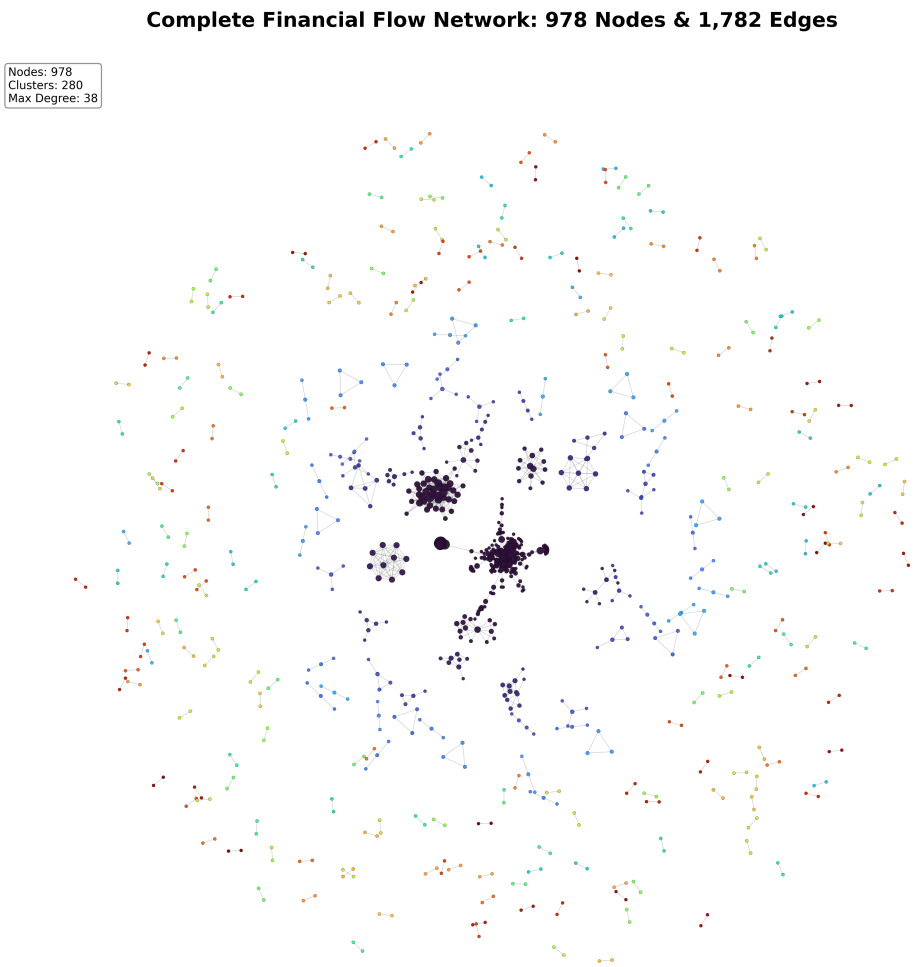



**Figure A5.** Direct On-Chain Fund Transfers within Identified Clusters
Nodes are the 978 wallets in 280 multi-wallet components; edges are the 1,782 direct transfer links after excluding official Polymarket addresses. Colors distinguish components and node size denotes degree.

Trader Classification: Definitions

| Trader Class | Classification Rule | Interpretation |
|---|---|---|
| **Skilled traders** | • Based on non-earnings markets.<br>• At least 10 distinct events.<br>• At least 10 trades.<br>• Cumulative volume of at least 100 tokens.<br>• Outside the market-maker exclusion group.<br>• Positive total PnL.<br>• One-sided $p < 0.01$ from 10,000 event-level sign randomizations. | Trading or information-processing advantages beyond a particular contract; rejects the random-luck benchmark at the 1% level. |
| **Insider candidates** | • Not classified as skilled.<br>• Gross settlement PnL $> 0$.<br>• Directionality $> 0.90$.<br>• Account-Date Gap $< 7$ days and nonmissing.<br>• At least two distinct trades **or** two active wallets within the same market. | Trading behavior consistent with event-specific private information. |
| **Noise traders** | Users not classified as skilled traders or insider candidates. | Remaining group. |

All listed conditions must hold within each class; the final insider-candidate condition permits either multiple trades or multiple wallets. Definitions and implementation details follow below.

**Skill.** We adapt the skill-versus-luck approach of Gomez Cram et al. (2026) to identify wallets whose profits are unlikely to arise from chance. In our implementation, we aggregate non-earnings profits across linked wallets to user–event observations. The table lists eligibility thresholds. The market-maker exclusion requires all three conditions: liquidity-providing volume share $> 75\%$, two-sided trading in at least 75% of markets, and market participation at or above the 95th percentile for the relevant clustering method and skill sample. Ineligible or excluded users are not tested.

For eligible user $u$, let $\pi_{ue}$ be event profit and $\Pi_u = \sum_e \pi_{ue}$. Independently flipping event-profit signs with equal probability gives

$$\Pi_u^{(b)} = \sum_e \xi_{ue}^{(b)} \pi_{ue}, \quad \xi_{ue}^{(b)} \in \{-1, +1\}, \quad p_u = \frac{1 + \sum_{b=1}^{10,000} \mathbf{1}\{\Pi_u^{(b)} \geq \Pi_u\}}{10,001}. \tag{A1}$$

Positive total profit and $p_u < 0.01$ identify skill relative to this random-luck benchmark. We copy user-level flags to linked wallets; a flow-based user is skilled if any linked wallet carries the flag. Non-earnings skill need not be known at the earnings trade date; the separate historical measure uses activity through September 15, 2025.

**Insider candidates.** The insider screen requires gross settlement PnL $> 0$ and Directionality $> 0.90$, in addition to the age and activity conditions in the table. Gross settlement PnL sums signed YES-equivalent tokens times the difference between the realized YES payoff and execution price, before fees; PnL per Token divides this by gross tokens traded. Directional concentration is

$$Directionality_{u,m} = \frac{|\text{Net YES-equivalent tokens}_{u,m}|}{\text{Gross tokens traded}_{u,m}}. \tag{A2}$$

Account-Date Gap is the market's first observed trade date minus the earliest available profile-creation or first-outward-transfer date across linked wallets, in New York calendar days. Nonmissing negative gaps qualify; this is not wallet age at execution. Distinct trades are market–transaction-hash pairs within a user; active wallets must actually trade. The activity screen allows multiple trades *or* multiple wallets within the same market, building on profitability-based classifications in Goldstein et al. (2026).

Insider candidates use ex post histories and settlement profits. Their activity and price responses are descriptive: treating the labels as available at trade time introduces look-ahead bias, and positive-profit selection prevents profitability from independently validating private information. The label does not establish a trader's identity or illegal insider trading.

**Order imbalance.** For group $g$, contract $m$, and signal interval $t$, liquidity-taking orders define

$$OIB^{g}_{m,t} = \frac{\text{Group } g \text{ net YES-equivalent tokens in } (m,t)}{\text{All groups' total liquidity-taking tokens in } (m,t)}. \tag{A3}$$

Buying YES or selling NO contributes positive flow; selling YES or buying NO contributes negative flow. The common denominator captures both direction and group activity share. Stock-level OIB sums group net tokens and total tokens across contracts for the same ticker before dividing.

**Group characteristics.** Table 7, Panel A, describes the full classification sample. Appendix Table A3 instead covers liquidity-taking orders in complete six-hour bins before each contract's cutoff, prior to regression-specific exclusions. Its token shares and unit profits should not be mixed with full-sample statistics.

## Cross-Market Timing and Regressions

**Polymarket.** Panels B–C of Table 7 fix OIB over the six-hour interval $t \to t + 6h$ and measure outcomes over $t \to t + 6h$, $t + 6h \to t + 12h$, and $t + 12h \to t + 18h$. Subsequent outcomes are separate segments, not cumulative changes from $t$. Bins are aligned in UTC; clustering dates use New York time. Boundary prices are the latest valid YES-equivalent

**Table A3.** Trading Activity and Profitability in the OIB Construction Sample

Liquidity-taking orders in complete six-hour bins before contract cutoffs and regression-specific exclusions; this sample differs from Table 7, Panel A. Shares use all groups' tokens. Gross PnL is trade-attributed settlement profit before fees; unit profit is 100 × PnL/tokens in cents, not a capital return. Excluding liquidity providers means aggregate PnL need not sum to zero. The ex post classification selects profitable insider candidates, so profits do not independently validate private information.

| | Skilled | Insider candidates | Noise traders |
|---|---|---|---|
| Token volume | 1,522,746 | 287,365 | 13,513,680 |
| Token share (%) | 9.937 | 1.875 | 88.188 |
| Gross PnL (USDC) | 28,326 | 66,595 | -486,112 |
| PnL per Token (cents) | 1.860 | 23.174 | -3.597 |

liquidity-providing fill prices strictly before each boundary, token-weighted at tied timestamps and carried through empty bins without a staleness limit or backfilling before the first trade.

For endpoints $(a, b)$, probability change is $100(p_b - p_a)$ and Price Discovery is $100(|p_a - Y| - |p_b - Y|)$, where $Y \in \{0, 1\}$ is the realized YES payoff. Both are in percentage points; positive Price Discovery denotes reduced absolute settlement error. It is an ex post accuracy measure. Probability changes use signed current OIB; Price Discovery uses absolute current OIB. Both retain opening probability, preceding-bin probability change, preceding-bin log liquidity-taking volume, log hours to cutoff, and signed preceding-bin group OIBs. All three classes enter jointly. Models include an intercept and no fixed effects, with standard errors clustered by event and New York signal-start date.

**Stock sessions.** Table 8 aggregates eligible half-hour liquidity-taking tokens by ticker and session, summing across contracts before dividing by all-group gross tokens. Bins must precede each contract's cutoff, so session coverage may be partial. Regular and overnight signals each explain returns over the original session and the next two separate sessions. Only the first column is contemporaneous. The calendar accounts for weekends, holidays, daylight-saving changes, and early closes.

Stock returns are $100 \log(P_b/P_a)$. TAQ NBBO midpoints must strictly precede each endpoint by at most 300 seconds; missing endpoints are excluded. Controls—return, stock OIB, log(1 + share volume), and opening spread—come from the latest completed regular session before signal start and remain fixed across outcomes. Missing TAQ inputs are not replaced with older sessions. Each regression includes all three group OIBs and an intercept, no fixed effects, and its own complete sample. Standard errors are clustered by ticker and New York signal-start date.

**Measurement limits.** Both analyses use market-close metadata when verified earnings timestamps are unavailable, so pre-cutoff signals and responses may include public disclosure. Without a corporate-action factor file, TAQ cross-day price adjustments remain unverified.

Settlement-based insider classification and Price Discovery do not validate real-time signals.

# Appendix 4: Derivations for the Theoretical Framework

This appendix provides the setup and derivations underlying Section 2. A one-period Kyle comparison describes how an identity-cost advantage can compensate informed traders for lower liquidity. A conditional signal-extraction calculation then shows how public trading records can inform equity investors who acquire them. The venue comparison takes gross trading rents before cross-market learning as given; the learning calculation conditions on signal acquisition. These complementary benchmarks support the participation, learning, and earnings-incorporation mechanisms in H1–H3 without jointly solving for venue entry and subsequent learning.

## Setup

There is one payoff-relevant latent earnings shock,

$$\epsilon \sim N(0, \sigma_\epsilon^2), \tag{A4}$$

where $\epsilon$ represents unexpected earnings news. The stock-market payoff is normalized as $v_S = \epsilon$.

The actual Polymarket earnings contract is binary. Let

$$Y = 1\{\epsilon > \kappa\}, \tag{A5}$$

where $\kappa$ is the relevant earnings threshold. The Polymarket price is a probability,

$$\boldsymbol{q}_P = \Pr(Y = 1 \mid \mathcal{I}_P) = \Pr(\epsilon > \kappa \mid \mathcal{I}_P), \tag{A6}$$

where $\mathcal{I}_P$ denotes the information embedded in Polymarket trading. We use a continuous latent-shock approximation and interpret Polymarket prices as signals about $\epsilon$. Under normal beliefs,

$$\boldsymbol{q}_P = \Phi\left(\frac{E[\epsilon \mid \mathcal{I}_P] - \kappa}{\sqrt{\mathrm{Var}(\epsilon \mid \mathcal{I}_P)}}\right), \tag{A7}$$

so the binary price summarizes a threshold probability that depends on both the posterior mean and variance. The linear benchmark below uses a continuous-payoff approximation to isolate the participation and learning mechanisms.

There are two trading venues:

$$P = \text{Polymarket}, \qquad S = \text{stock market}. \tag{A8}$$

In venue $m \in \{P, S\}$, an informed trader observes $\epsilon$ and submits a market order $x_m(\epsilon)$. Noise trading is

$$u_m \sim N(0, \sigma^2_{\ell,m}), \tag{A9}$$

where $\sigma^2_{\ell,m}$ is liquidity noise, and $u_m$ is independent of $\epsilon$. Total order flow is

$$z_m = x_m(\epsilon) + u_m. \tag{A10}$$

A competitive risk-neutral market maker observes $z_m$ and sets price equal to the conditional expectation of the payoff. Privately informed traders have access to the payoff-relevant signal.

Within the common, normalized continuous-claim benchmark, we assume that aggregate noise order flow has a lower standard deviation in Polymarket:

$$\sigma_{\ell,P} < \sigma_{\ell,S}. \tag{A11}$$

The intuition is that the stock market has a larger pool of noise traders whose liquidity demands provide more scope for informed trading before prices move. The model captures this difference through aggregate noise order flow. Polymarket's public record is accessible to all equity-market participants, including noise traders, who may choose to acquire a signal from it. Here $u_S$ represents exogenous liquidity demand; access to public information is not restricted by that trading motive. Conditional on acquisition, investors observe Polymarket activity with public-observation noise:

$$\boldsymbol{s}_P = \boldsymbol{p}_P + \eta_P, \qquad \eta_P \sim N(0, \tau^2_P), \tag{A12}$$

where $\boldsymbol{p}_P$ is the latent Polymarket price signal and $\boldsymbol{s}_P$ is the signal available to equity-market investors. The parameter $\tau^2_P$ measures how difficult it is for outside investors to decode Polymarket's public record. Lower $\tau^2_P$ corresponds to greater post-trade transparency.

The timing is as follows. First, privately informed traders choose whether to trade in Polymarket or in the stock market. Second, informed trading in Polymarket generates a price and public trading record. Third, equity-market investors decide whether to acquire $\boldsymbol{s}_P$; those who acquire an incrementally informative signal can incorporate it into stock prices. This sequence separates the participation and learning decisions analytically.

## Trading Rents, Sorting, and Concealment

Consider a standard one-period Kyle equilibrium in venue $m$. The informed trader uses a linear strategy $x_m = \beta_m \epsilon$, and the market maker sets a linear price $\boldsymbol{p}_m = \lambda_m z_m$. Given a positive $\lambda_m$, the trader solves

$$\max_{x_m} E[(\epsilon - \boldsymbol{p}_m)x_m \mid \epsilon] = \max_{x_m}(\epsilon - \lambda_m x_m)x_m. \tag{A13}$$

The first-order condition gives $x_m = \epsilon/(2\lambda_m)$, so $\beta_m = 1/(2\lambda_m)$. Competitive pricing and joint normality imply

$$\lambda_m = \frac{\mathrm{Cov}(\epsilon, z_m)}{\mathrm{Var}(z_m)} = \frac{\beta_m \sigma_\epsilon^2}{\beta_m^2 \sigma_\epsilon^2 + \sigma_{\ell,m}^2}. \tag{A14}$$

Combining these two conditions yields $\beta_m^2 \sigma_\epsilon^2 = \sigma_{\ell,m}^2$ and $\lambda_m \beta_m = 1/2$. Expected gross rent is therefore $R_m = E[(\epsilon - \boldsymbol{p}_m)x_m] = \beta_m \sigma_\epsilon^2/2$. Taking the positive solution gives

$$\beta_m = \frac{\sigma_{\ell,m}}{\sigma_\epsilon}, \qquad \lambda_m = \frac{1}{2}\frac{\sigma_\epsilon}{\sigma_{\ell,m}}, \qquad R_m = \frac{1}{2}\sigma_\epsilon \sigma_{\ell,m}, \tag{A15}$$

where $R_m$ is the informed trader's expected gross rent. Since $\sigma_{\ell,P} < \sigma_{\ell,S}$,

$$R_P < R_S, \qquad \beta_P < \beta_S, \qquad \lambda_P > \lambda_S. \tag{A16}$$

The stock market offers higher gross rents because noise trading provides more camouflage. Polymarket, by contrast, has higher price impact.

We allow identity exposure to impose costs on informed traders and assume that these costs can be lower under Polymarket's wallet-level pseudonymity. Such costs may reflect the disclosure of a trader's research activity, professional relationships, or information advantage. The sorting mechanism depends on a difference in these identity-based costs across venues.

Let $C_{S,i}$ and $C_{P,i}$ denote informed trader $i$'s expected identity-based cost of trading in the stock market and Polymarket, respectively. Define

$$\Delta C_i = C_{S,i} - C_{P,i}. \tag{A17}$$

Trader $i$ chooses Polymarket if

$$R_P - C_{P,i} \geq R_S - C_{S,i}, \tag{A18}$$

or equivalently,

$$\Delta C_i \geq R_S - R_P = \frac{1}{2}\sigma_\epsilon \left(\sigma_{\ell,S} - \sigma_{\ell,P}\right). \tag{A19}$$

If $\Delta C_i$ has continuous distribution function $F$, the mass of privately informed traders who sort into Polymarket is

$$\pi_P = 1 - F\left[\frac{1}{2}\sigma_\epsilon\left(\sigma_{\ell,S} - \sigma_{\ell,P}\right)\right]. \tag{A20}$$

This expression delivers partial sorting: informed traders with sufficiently high stock-market identity costs choose Polymarket, while others remain in the stock market.

**Informed participation.** The sorting condition illustrates H1: an identity-cost advantage must compensate for lower gross rents. The distribution $F$ determines the share of traders for whom this condition holds. The resulting price impact also motivates the empirical analysis of how traders manage information exposure through their trading patterns. For the cross-market learning result, note that the Polymarket price can be written as

$$\boldsymbol{p}_P = \lambda_P z_P = \lambda_P(\beta_P \epsilon + u_P) = \frac{1}{2}\epsilon + \lambda_P u_P. \tag{A21}$$

Therefore,

$$\mathrm{Cov}(\epsilon, \boldsymbol{p}_P) = \frac{1}{2}\sigma_\epsilon^2, \qquad \mathrm{Var}(\boldsymbol{p}_P) = \frac{1}{2}\sigma_\epsilon^2. \tag{A22}$$

## Transparency and Stock-Market Learning

Let $\mathcal{I}_{S,j}$ denote equity investor $j$'s existing information, $K_j$ the cost of acquiring and interpreting a Polymarket signal, and $B_j(\boldsymbol{s}_P \mid \mathcal{I}_{S,j})$ the ex ante expected gain from using the resulting posterior beliefs to make better investment decisions, relative to decisions based on $\mathcal{I}_{S,j}$ alone. This benefit is evaluated before observing the signal and before deducting the acquisition and interpretation cost. The investor acquires the signal only if

$$B_j(\boldsymbol{s}_P \mid \mathcal{I}_{S,j}) > K_j. \tag{A23}$$

This participation condition states the trade-off in H2, taking investors' signal values and acquisition costs as given. The posterior calculations below describe the information improvement that investors can use in their decisions. The model does not specify an investor objective that converts the reduction in earnings uncertainty into a numerical decision benefit $B_j$. The observation-noise variance $\tau_P^2$ describes signal quality conditional on acquisition, while $K_j$ is the fixed cost of obtaining and interpreting it.

For the conditional calculation, interpret $\epsilon$ as the earnings innovation remaining after conditioning on a common stock-market information set. Assume the acquired Polymarket signal contains information about that innovation. The calculation therefore isolates learning

from a signal that adds to the stock market's existing information.

Conditional on acquisition, equity-market investors observe a signal about Polymarket activity,

$$\boldsymbol{s}_P = \boldsymbol{p}_P + \eta_P, \qquad \eta_P \sim N(0, \tau_P^2), \tag{A24}$$

where $\eta_P$ is independent of $\epsilon$ and $u_P$. Observation noise $\eta_P$ captures imprecision in the signal extracted from the public record, and $\tau_P^2$ is its variance. When Polymarket records make prices, trades, timestamps, order direction, and wallet histories easier to decode, $\tau_P^2$ is lower. Trading noise $\lambda_P u_P$ remains embedded in $\boldsymbol{p}_P$ even when observation noise is zero.

Using equation (A22),

$$\mathrm{Cov}(\epsilon, \boldsymbol{s}_P) = \frac{1}{2}\sigma_\epsilon^2, \qquad \mathrm{Var}(\boldsymbol{s}_P) = \frac{1}{2}\sigma_\epsilon^2 + \tau_P^2. \tag{A25}$$

Because $(\epsilon, \boldsymbol{s}_P)$ are jointly normal with zero means, the conditional mean is

$$E[\epsilon \mid \boldsymbol{s}_P] = \frac{\mathrm{Cov}(\epsilon, \boldsymbol{s}_P)}{\mathrm{Var}(\boldsymbol{s}_P)}\boldsymbol{s}_P = \frac{\sigma_\epsilon^2/2}{\sigma_\epsilon^2/2 + \tau_P^2}\boldsymbol{s}_P. \tag{A26}$$

This is the updating relation in equation (5). The corresponding conditional variance is

$$\begin{aligned}
\mathrm{Var}(\epsilon \mid \boldsymbol{s}_P) &= \sigma_\epsilon^2 - \frac{\mathrm{Cov}(\epsilon, \boldsymbol{s}_P)^2}{\mathrm{Var}(\boldsymbol{s}_P)} \\
&= \sigma_\epsilon^2 - \frac{\left(\frac{1}{2}\sigma_\epsilon^2\right)^2}{\frac{1}{2}\sigma_\epsilon^2 + \tau_P^2} \\
&= \sigma_\epsilon^2 \frac{\sigma_\epsilon^2 + 4\tau_P^2}{2\sigma_\epsilon^2 + 4\tau_P^2}.
\end{aligned} \tag{A27}$$

The reduction from the prior variance in equation (A27) is $\sigma_\epsilon^4/(2\sigma_\epsilon^2 + 4\tau_P^2)$, which is strictly positive for $\sigma_\epsilon^2 > 0$ and finite $\tau_P^2$. This is the reduction displayed in equation (6). If $\tau_P^2 = 0$, then

$$\mathrm{Var}(\epsilon \mid \boldsymbol{s}_P) = \frac{1}{2}\sigma_\epsilon^2. \tag{A28}$$

If $\tau_P^2 \to \infty$, then

$$\mathrm{Var}(\epsilon \mid \boldsymbol{s}_P) \to \sigma_\epsilon^2. \tag{A29}$$

Thus, in this benchmark an acquired signal reduces residual uncertainty when it contains incremental information and has finite observation noise. If investors do not acquire the signal, their information set and conditional variance remain unchanged. Public availability alone does not imply learning.

Define the residual private information after equity-market investors observe Polymarket

as

$$\widetilde{\epsilon} = \epsilon - E[\epsilon \mid \boldsymbol{s}_P], \tag{A30}$$

with variance

$$\sigma^2_{\widetilde{\epsilon}} = \mathrm{Var}(\epsilon \mid \boldsymbol{s}_P) = \sigma^2_\epsilon \frac{\sigma^2_\epsilon + 4\tau^2_P}{2\sigma^2_\epsilon + 4\tau^2_P}. \tag{A31}$$

Privately informed traders who remain in the stock market trade on this residual information advantage. Conditional on $\boldsymbol{s}_P$, the stock-market trading problem is again a Kyle problem, now with payoff innovation $\widetilde{\epsilon}$ and liquidity noise $\sigma^2_{\ell,S}$. Therefore,

$$R^P_S = \frac{1}{2}\sigma_{\widetilde{\epsilon}}\sigma_{\ell,S}. \tag{A32}$$

Without Polymarket, the benchmark rent of an informed stock-market trader is

$$R^0_S = \frac{1}{2}\sigma_\epsilon\sigma_{\ell,S}. \tag{A33}$$

Combining equations (A31)–(A33),

$$\frac{R^P_S}{R^0_S} = \frac{\sigma_{\widetilde{\epsilon}}}{\sigma_\epsilon} = \sqrt{\frac{\sigma^2_\epsilon + 4\tau^2_P}{2\sigma^2_\epsilon + 4\tau^2_P}}. \tag{A34}$$

When $\tau^2_P = 0$, $R^P_S/R^0_S = 1/\sqrt{2}$. When $\tau^2_P \to \infty$, $R^P_S/R^0_S \to 1$. Greater post-trade transparency therefore strengthens the cross-market information channel.

**Conditional learning and empirical interpretation.** The variance and rent calculations show how acquisition of an incrementally informative signal reduces residual uncertainty and informed trading rents in the benchmark. Our empirical tests examine the associated information-diffusion channel through stock-price responses, pre-announcement earnings sensitivity, and variation with signal visibility and trading composition.

# Appendix 5: Additional Evidence on Earnings Incorporation

## Covariate Balance

**Table A4.** Propensity Score Matching: Covariate Balance

U and M denote samples before and after 1:5 nearest-neighbor matching using a logit model, a 0.05 caliper, and common support. Matching covariates are analysts, firm size, lagged ROA, lagged leverage, and news coverage. Institutional ownership and news sentiment enter outcome regressions but not matching or the reported balance tests. Variables follow Table 9. Balance statistics use matching weights; outcome regressions are unweighted. Bias is the standardized mean difference in percent. Reported $p$-values are rounded to three decimals. ***, **, and * denote significance at the 1%, 5%, and 10% levels.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Variable | Sample | Treated | Control | % Bias | $t$ | $p$ |
| Analysts | U | 2.2480 | 1.5982 | 97.1 | 26.70 | 0.000*** |
| | M | 2.2432 | 2.2536 | -1.6 | -0.30 | 0.762 |
| Firm Size | U | 10.182 | 7.7347 | 123.8 | 33.19 | 0.000*** |
| | M | 10.162 | 10.195 | -1.7 | -0.34 | 0.736 |
| $\text{ROA}_{t-1}$ | U | 0.01074 | -0.01639 | 18.5 | 3.74 | 0.000*** |
| | M | 0.01033 | 0.01027 | 0.0 | 0.02 | 0.983 |
| $\text{Leverage}_{t-1}$ | U | 0.34990 | 0.30162 | 16.2 | 4.54 | 0.000*** |
| | M | 0.35063 | 0.34511 | 1.9 | 0.33 | 0.745 |
| News Coverage | U | 7.0080 | 5.5840 | 121.9 | 39.87 | 0.000*** |
| | M | 6.9931 | 6.9746 | 1.6 | 0.27 | 0.789 |

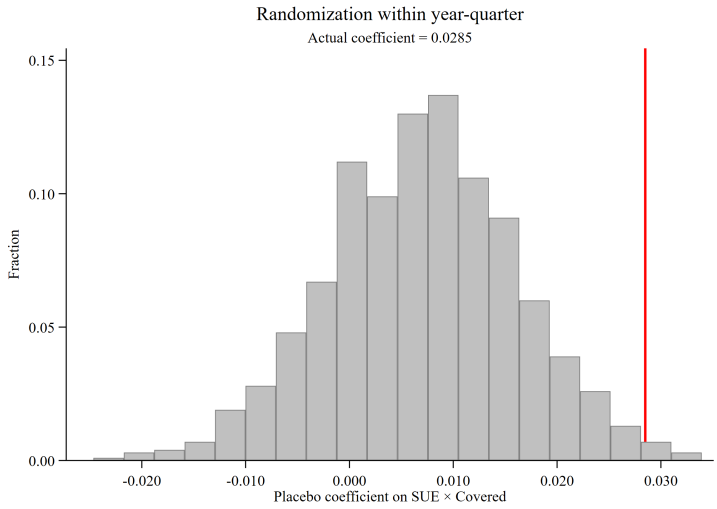
| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Sample | Pseudo-$R^2$ | Mean Bias | Median Bias | Rubin's $B$ | Rubin's $R$ | LR $p$ |
| Unmatched | 0.208 | 75.5 | 97.1 | 137.8 | 1.41 | 0.000 |
| Matched | 0.000 | 1.3 | 1.6 | 4.5 | 1.05 | 0.978 |

**Table A5.** Robustness of Pre-announcement Earnings Responses

| | Panel A. Event windows and public-information days | | | | Panel B. Return measures and matching ratios | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Dependent variable | | | CAPM BHAR | | FF3 BHAR | CAPM CAR | CAPM BHAR | CAPM BHAR |
| | $[-4,-1]$ | $[-6,-1]$ | No Analyst EPS Forecast Days | No Macro News Days | PSM 1:5 | PSM 1:5 | PSM 1:4 | PSM 1:6 |
| SUE | 0.000 | 0.002 | -0.001 | -0.007 | -0.005 | -0.002 | -0.012 | 0.004 |
| | (0.045) | (0.220) | (-0.182) | (-0.996) | (-0.643) | (-0.260) | (-1.599) | (0.652) |
| Covered | -0.010* | -0.008 | -0.011 | -0.017*** | -0.015** | -0.013* | -0.018** | -0.012* |
| | (-1.655) | (-1.099) | (-1.646) | (-3.034) | (-2.310) | (-1.937) | (-2.575) | (-1.911) |
| SUE × Covered | 0.019* | 0.024* | 0.024** | 0.033*** | 0.028** | 0.028** | 0.037*** | 0.022* |
| | (1.825) | (1.756) | (1.996) | (3.370) | (2.401) | (2.314) | (3.065) | (1.946) |
| Analysts | 0.002 | 0.004 | 0.003 | 0.002 | 0.000 | 0.002 | 0.004 | 0.000 |
| | (0.838) | (1.025) | (0.877) | (0.608) | (0.059) | (0.719) | (1.199) | (0.107) |
| Firm Size | -0.021*** | -0.034*** | -0.027*** | -0.019*** | -0.023*** | -0.026*** | -0.028*** | -0.023*** |
| | (-3.726) | (-4.523) | (-4.111) | (-3.246) | (-3.181) | (-3.931) | (-3.998) | (-4.344) |
| $\text{ROA}_{t-1}$ | 0.043 | 0.023 | -0.003 | -0.036 | 0.018 | 0.017 | -0.042 | -0.031 |
| | (0.580) | (0.278) | (-0.046) | (-0.502) | (0.243) | (0.219) | (-0.512) | (-0.434) |
| $\text{Leverage}_{t-1}$ | -0.031 | -0.047 | -0.012 | -0.005 | -0.035 | -0.035 | -0.030 | -0.026 |
| | (-0.840) | (-1.091) | (-0.305) | (-0.120) | (-0.828) | (-0.805) | (-0.852) | (-0.773) |
| $\text{Institutional Ownership}_{t-1}$ | -0.015 | -0.069** | -0.061** | -0.036 | -0.064** | -0.056* | -0.050* | -0.048* |
| | (-0.609) | (-2.032) | (-2.084) | (-1.339) | (-2.121) | (-1.955) | (-1.721) | (-1.802) |
| News Coverage | 0.004 | 0.006 | 0.001 | 0.001 | 0.003 | 0.005 | 0.001 | 0.006* |
| | (1.234) | (1.534) | (0.385) | (0.410) | (0.770) | (1.262) | (0.194) | (1.772) |
| News Sentiment | -0.002 | 0.015 | 0.008 | -0.006 | -0.002 | 0.004 | 0.009 | 0.002 |
| | (-0.179) | (1.408) | (0.866) | (-0.703) | (-0.217) | (0.432) | (0.940) | (0.290) |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,522 | 2,522 | 2,513 | 2,522 | 2,522 | 2,522 | 2,137 | 2,863 |
| Adjusted $R^2$ | 0.039 | 0.069 | 0.067 | 0.022 | 0.051 | 0.060 | 0.063 | 0.052 |

This table reports four types of robustness tests. Panel A varies the window in columns (1)–(2) and excludes analyst-forecast or macroeconomic-news days from $[-5,-1]$ in columns (3)–(4), without replacing excluded days. Panel B uses FF3 BHAR and CAPM CAR in columns (5)–(6), and CAPM BHAR with 1:4 and 1:6 matching in columns (7)–(8), all over $[-5,-1]$. Other columns use baseline 1:5 matching. Matching uses a logit model, a 0.05 caliper, and common support. All specifications are unweighted and include the seven controls in Table 9, firm and year-quarter fixed effects. Parentheses contain firm-clustered $t$-statistics. ***, **, and * denote significance at 1%, 5%, and 10%.

## Panel A. Randomization within Year-Quarter




## Panel B. Randomization within Firm

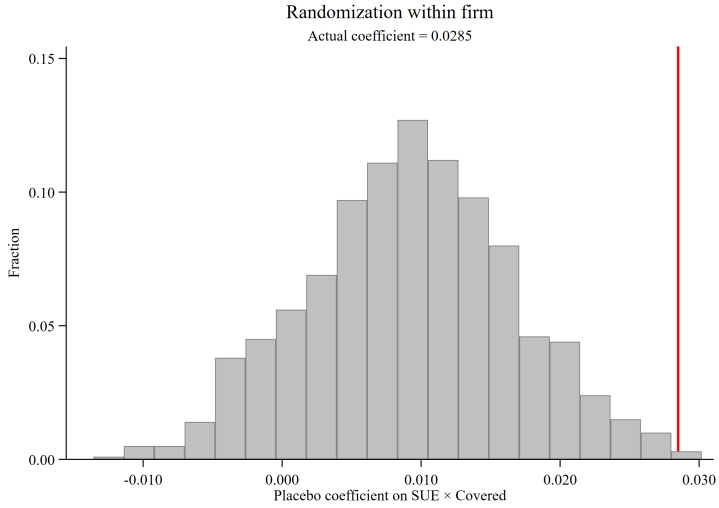



**Figure A6.** Placebo Tests for Polymarket Coverage

The histograms show 1,000 permutations of Polymarket coverage within year-quarters (Panel A) and within firms (Panel B), preserving each stratum's coverage count. Each permutation re-estimates the baseline CAPM BHAR regression over $[-5, -1]$ on its 2,522-observation sample, with the seven controls in Table 9, firm and year-quarter fixed effects, and firm-clustered standard errors. The red line marks the observed SUE × Covered coefficient, 0.0284942 (displayed as 0.0285). The proportions of permutation estimates at least this large are 8/1,000 (0.8%) in Panel A and 2/1,000 (0.2%) in Panel B.

# Earnings Response Coefficients Before and After Matching

**Table A6.** Earnings Response Coefficients Before and After Matching

This table reports SUE slopes without the coverage indicator or its interaction. Columns (1)–(2) use the Full Sample of eligible complete-case observations; columns (3)–(4) use the 1:5 PSM Sample described in Table 9. Outcomes are CAPM BHAR. All specifications include analyst coverage, firm size, lagged ROA, lagged leverage, previous-quarter institutional ownership, news coverage, news sentiment, and firm and year-quarter fixed effects. Regressions are unweighted. Parentheses contain firm-clustered $t$-statistics. ***, **, and * denote significance at the 1%, 5%, and 10% levels.

| *Dependent variable: BHAR* | Full Sample | | PSM Sample | |
|---|---|---|---|---|
| | (1) $[-5,-1]$ | (2) $[0,+2]$ | (3) $[-5,-1]$ | (4) $[0,+2]$ |
| SUE | 0.007*** | 0.093*** | 0.007 | 0.089*** |
| | (3.547) | (22.065) | (1.224) | (6.815) |
| Analysts | 0.002* | -0.001 | 0.003 | 0.005 |
| | (1.664) | (-0.357) | (0.931) | (0.798) |
| Firm Size | -0.017*** | -0.076*** | -0.026*** | -0.064*** |
| | (-7.296) | (-14.152) | (-4.361) | (-4.476) |
| $\text{ROA}_{t-1}$ | 0.032* | -0.032 | 0.011 | -0.050 |
| | (1.694) | (-1.065) | (0.157) | (-0.365) |
| $\text{Leverage}_{t-1}$ | 0.008 | 0.032 | -0.019 | 0.014 |
| | (0.568) | (0.952) | (-0.519) | (0.137) |
| $\text{Institutional Ownership}_{t-1}$ | 0.007 | -0.007 | -0.051* | 0.075 |
| | (0.597) | (-0.299) | (-1.893) | (1.171) |
| News Coverage | -0.001 | -0.003 | 0.004 | -0.008 |
| | (-0.370) | (-1.148) | (1.151) | (-0.984) |
| News Sentiment | -0.006 | -0.033*** | 0.009 | -0.038* |
| | (-1.338) | (-3.586) | (0.965) | (-1.779) |
| Firm FE | Yes | Yes | Yes | Yes |
| Year-Quarter FE | Yes | Yes | Yes | Yes |
| Observations | 16,453 | 16,448 | 2,522 | 2,522 |
| Adjusted $R^2$ | 0.063 | 0.091 | 0.063 | 0.075 |

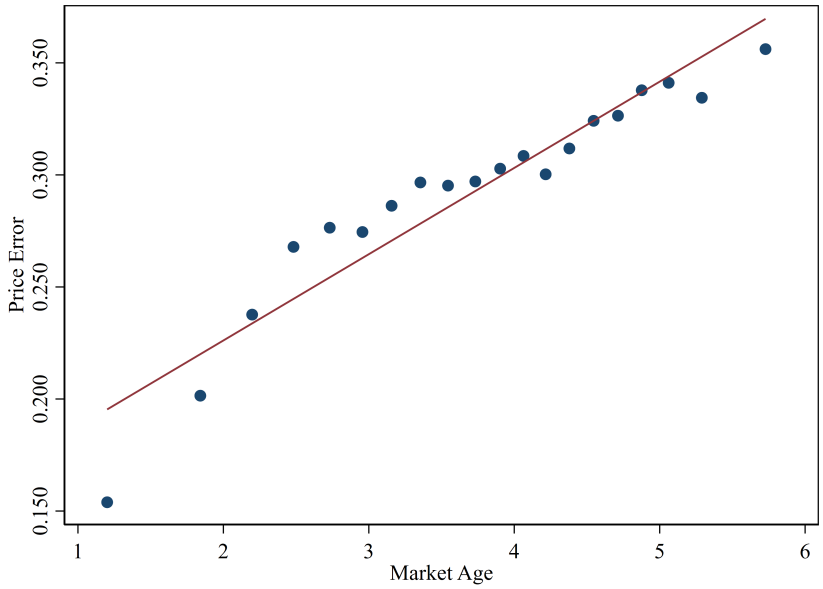
# Appendix 6: Additional Evidence on Information and Learning

## Price Error and Information Aggregation

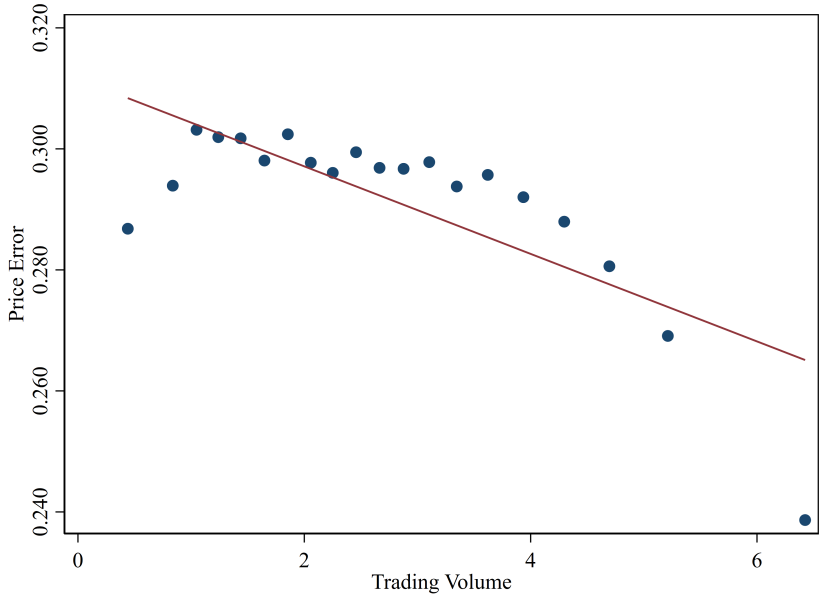
**Table A7.** Trading Volume, Time to Settlement, and Price Error

This table reports OLS regressions of price error on trading volume and time remaining. The specification is $\textit{Price Error}_{i,m,d} = \alpha + \beta_1 \textit{Volume}_{i,m,d} + \beta_2 \textit{Time Remaining}_{i,m,d} + f_m + f_d + \varepsilon_{i,m,d}$, with $f_d$ included only in column (2). Price error is measured as the absolute deviation between the transaction price and the ex-post settlement value. Volume is measured as the log of one plus the number of tokens traded in a transaction. Time Remaining measures the interval until market resolution. Column (1) includes condition fixed effects, and Column (2) additionally includes calendar fixed effects. Standard errors, clustered at the market level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

| *Dependent variable:* *Price Error* | (1) | (2) |
|---|---|---|
| Log Volume | -0.004*** | -0.004*** |
| | (0.001) | (0.001) |
| Time Remaining | 0.038*** | 0.063*** |
| | (0.003) | (0.005) |
| Constant | 0.163*** | 0.069*** |
| | (0.011) | (0.018) |
| Observations | 172,866 | 172,866 |
| Adjusted $R^2$ | 0.771 | 0.784 |
| Condition FE | Yes | Yes |
| Calendar FE | No | Yes |




Panel A. Price Error and market age




Panel B. Price Error and trading volume

**Figure A7.** Price Error, Market Age, and Trading Volume

This figure reports binned scatter plots relating Price Error to market age and trading volume. Price Error is defined as the absolute deviation between the transaction price of the "Yes" token and the ex-post settlement value. Market age is measured as $\ln(1 + \text{Hours to Last Trade})$, where Hours to Last Trade is the number of hours between the current transaction and the final transaction observed in the same market. Panel A plots residualized Price Error against this measure after absorbing market fixed effects. Panel B plots residualized Price Error against log-transformed trading volume per transaction after absorbing market fixed effects.

## Trading Aggressiveness and Profitability

These exhibits compare average profitability across order types and profitability across traders' liquidity-taking shares.

**Table A8.** Performance of Liquidity-Taking and Liquidity-Providing Orders

Liquidity-taking orders execute against existing quotes, while liquidity-providing orders supply those quotes. PnL per Token is settlement profit per token in USDC, rather than a return on invested capital.

| Group | Variable | (1) N | (2) Mean | (3) SD | (4) Min | (5) P25 | (6) P50 | (7) P75 | (8) P90 | (9) Max |
|---|---|---|---|---|---|---|---|---|---|---|
| Liquidity-providing | PnL per Token | 289,497 | 0.013 | 0.383 | -0.999 | -0.170 | -0.035 | 0.180 | 0.610 | 0.999 |
| Liquidity-taking | PnL per Token | 186,358 | -0.021 | 0.389 | -0.999 | -0.190 | 0.050 | 0.170 | 0.400 | 0.999 |

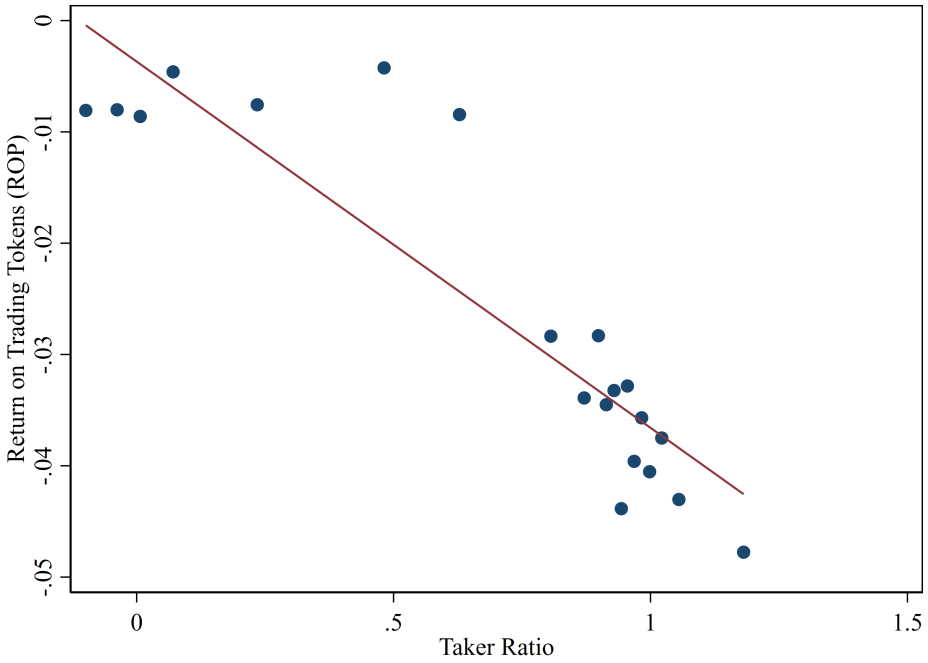



**Figure A8.** Trader Aggressiveness and Performance

This figure plots user-level PnL per Token against the liquidity-taking share, the share of total trading activity executed through liquidity-taking orders within a market. Liquidity-taking intensity is not monotonically associated with profitability, so aggressive trading alone does not identify informed traders.

# Platform Discussions and Potential Information Channels

**Table A9.** Platform Discussions of Potential Information Leakage

This table presents 13 selected comments about potential information leakage from 1,006 unique Polymarket discussion-board comments across 165 markets, identified using keywords such as "insider," "leak," "private info," and "internal source." The auditor allegation refers to the same anonymized auditor as the highest-accuracy auditor in the rightmost occupied observed bin of Panel B of Figure 5. This correspondence complements the price evidence; it does not verify the allegation or identify the traders. The auditor name, timestamp, and company name in that row remain anonymized.

| (1)<br>Comment | (2)<br>Company Name |
|---|---|
| Already leaked – yes. | Citigroup |
| nvm they sold. Insider smh | Tesla |
| Literal real time chart of insider trading | Tesla |
| lol insiders buying up no | Tesla |
| We can see that insider trading is alive and well on Tesla; they will def beat it | Tesla |
| Seems like some insiders are scooping up this yes | Evercore |
| Don't bet against insiders. | SoFi Technologies |
| Polymarket isn't regulated, or even legal, in the U.S. That's why this INSIDER is betting here instead of on the actual stock market. | SoFi Technologies |
| bro why did you buy so much at such a high price, I hope you're an insider; that was a bad decision | JPMorgan Chase |
| For a moment I thought maybe it's an insider with knowledge. | Wells Fargo & Co |
| I think they are insiders who work for XXXX auditors and already know the answer | XXXX |
| Don't forget insiders selling. | Netflix |
| So unfair earnings got leaked | T-Mobile |

## Access to Polymarket Signals

Stocktwits embeds earnings probabilities in a stock-investor interface, while public tutorials explain how to interpret Polymarket trading signals.

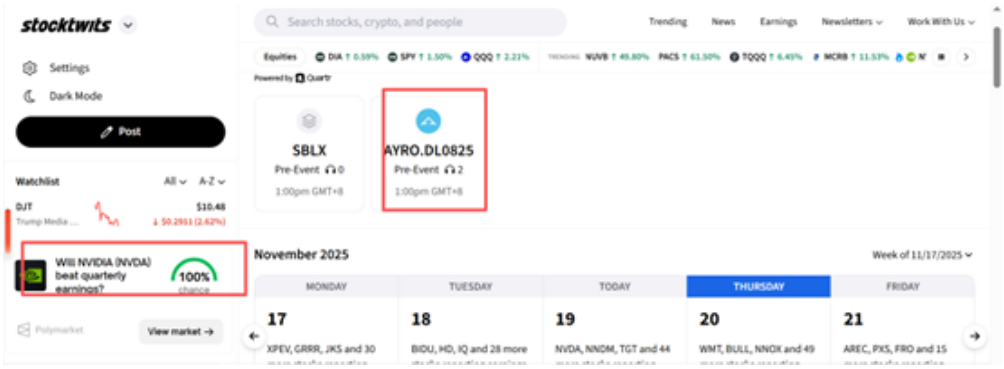



**Figure A9.** Polymarket Earnings Predictions in Stocktwits
The module displays Polymarket earnings probabilities alongside firm-specific investor discussions and links to the contract page. Source: https://stocktwits.com/.

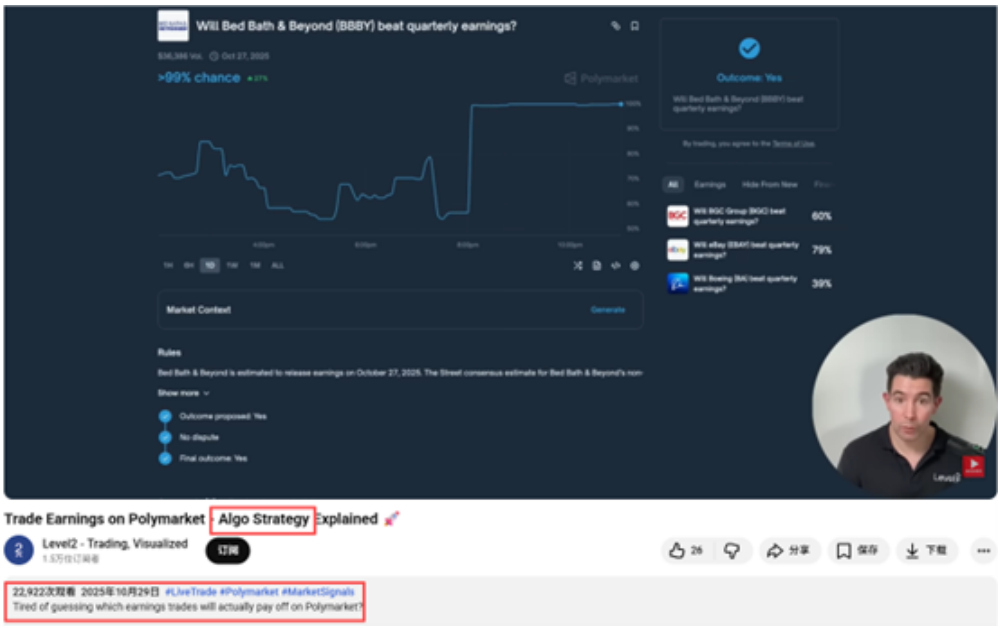



**Figure A10.** Algorithmic Trading Strategies on YouTube
The tutorial combines social sentiment, analyst expectations, market probabilities, and order flow. It tracks transaction size, timing, and trader histories to identify "smart money." Source: https://www.youtube.com/watch?v=r0msI2wJ1BQ.