High Volume Traders and Price Disruptions in Modern Prediction Markets

Ignacio Mijares Santibañez

Undergraduate Thesis

Bachelor of International Economics (BIE)

University of British Columbia

Advisor: Prof. Scot Orr

Abstract

This paper studies how high-volume traders, Whales, affect prediction market prices across varying time, sizes, and categories. Prediction markets are a novel form of market, allowing users to buy and sell event contracts with "Yes" or "No" outcomes. With a wide variety of possible topics, trade volumes, and participants, they combine market dynamics with probabilities and beliefs in ways that other trading spaces do not. Understanding a whale's relationship with price in this new environment is key to understanding these novel markets. To do so, this paper uses 'Polymarket', the leading event betting platform, and data on 40 resolved markets from 2020 onwards. To estimate the effect of whales on prices, changes in price after a whale purchase, relative to the market, time, and position, were calculated. Results indicate that whales have no statistically significant influence on price, regardless of the time frame observed or the size threshold for whales, suggesting that trade size is not a meaningful proxy for informational advantage and that prices efficiently incorporate available information.

Introduction

Prediction markets are open markets that enable users to predict specific outcomes on real-world events via financial incentives. They follow an exchange-traded market format, with binary outcomes on events often being tied to a price between \$0 and \$1, expiring at one of the extremes once a market reaches resolution. Prices thus serve as an aggregation of beliefs on the likelihood of an event occurring. Prediction markets then serve as a form of crowdsourcing opinions on unknown future events on any kind of topic. Although the concept of prediction markets is reported to have been around since the 16th century, it is only since 2020 that modern platforms such as ‘Kalshi’ and ‘Polymarket’ have popularized the format of exchange.

Financial markets play a key role in modern economics by facilitating the allocation of capital and information. Their importance stems from the efficient market hypothesis, the belief that asset prices reflect all available information, making market mechanisms for data processing and aggregation, taking all the available, yet dispersed, knowledge from all market participants to center on a singular price. This functionality extends beyond just traditional financial markets; crypto, real estate, and prediction markets all encode aggregate expectations. Thus, the resilience of a financial market to disruption, whether from geopolitical shocks, policy changes, or adverse practices, is critical, as it determines its ability to continue performing its core functionality. Robust markets are not just a luxury tool for efficient exchanges; they are essential to information efficiency and price stability.

Prediction markets offer a novel extension of financial market dynamics. Participants effectively bet on their confidence in specific outcomes, allocating capital where they believe they have an informational advantage. This creates a distinct environment where prices are even more explicitly tied to information and probabilities than in traditional markets.

In modern markets, such as crypto, “Whales”, or participants who control disproportionately large amounts of capital and thus may possess the ability to influence outcomes via market power, have become a source of disruption. These large trades can shift prices by altering perceived probabilities and information signals. Given high fragmentation and relatively lower liquidity in many crypto-based platforms, a whale's impact is found to be more profound than in traditional markets. Given the direct parallels between prediction and crypto markets as decentralized digital currency platforms protected via blockchain technology, and explicit ties between probabilities, prices, and information in prediction markets, it is natural to question how they may affect prediction markets instead. In particular, do prediction markets follow crypto-like dynamics, where whales can exert outsized influence, or does the inherently stronger informational structure of these markets act to dissipate such market power?

This paper, therefore, asks: How do high-volume traders 'Whales' affect prices in prediction markets, or is their impact limited by efficient information aggregation? Given the lack of a strict definition of a Whale in regards to size in modern literature, the paper defines them as trades larger than 5% of total market volume, with robustness checks testing results for values between 3% & 20%. Additionally, the markets explored are between $5k and $500k in size due to API limitations.

Literature Review

Literature on modern information markets has focused on analysing their accuracy, efficiency, bias, and arbitrage opportunities. Reichenbach & Walther (2025) explore the aggregation of dispersed information into probabilistic forecasts that guide markets, particularly in decentralized blockchain platforms like Polymarket, which provide users with real-time trading data. As far as accuracy is concerned, they found that market prices closely track realized probabilities and outperformed bookmaker odds. The paper measures accuracy by comparing implied probabilities to realized event frequencies. They also found evidence of skilled traders by testing whether individual gains per bet persistently exceed random expectations via Welch t-tests. They explored differences in behaviors and characteristics between these skilled traders and less skilled participants. Saguillo et al. (2025) explore arbitrage opportunities on Polymarket via a large-scale empirical analysis of intra-market rebalancing arbitrage and inter-market combinatorial arbitrage. The prevalence of arbitrage opportunities indicates limits to informational efficiency even in modern blockchain-based platforms, creating systematic opportunities for informed participants to profit despite accurate information aggregation. Reichenbach & Walthers (2025) methods for identifying skilled traders & Saguillo et al.'s (2025) framework for identifying mispricing could also be of interest when creating an instrument for exogeneity in my own exploration.

Much of the existing literature on prediction markets has centered on political predictions. Clinton and Huang (2024) analyzed 2,500 contracts traded in the final weeks of the 2024 U.S presidential election across several platforms. Their paper measured accuracy and informational efficiency via cross-market price comparisons and arbitrage analysis. They found prices diverged across platforms, often without correlation, allowing for arbitrage opportunities and inefficiencies in information aggregation. In contrast, Cutting et al. (2025) found Polymarket data specifically to outperform traditional polling in predicting the outcome of the 2024 election, particularly in swing states, citing aggregating beliefs as a key reason. Goodell, McGee, & McGroarty (2020) evaluate prediction market probabilities during the 2020 elections, finding that odds reflected shifts in policy and financial market expectations. Further back, Graefe (2017) evaluates of market performance in the 2016 elections and concludes that markets underperformed relative to their theoretical promise, attesting to structural and behavioral constraints to the underwhelming accuracy results. When evaluating this literature, it is important to consider both time and platform effects on their findings. Clinton & Huang (2024) used data from multiple platforms, while Cutting et al. (2025) focused their research on Polymarket, and discrepancies in accuracy may therefore reflect platform-specific design features rather than fundamental differences in the predictive power of markets themselves. Additionally, structural and behavioral limits in prediction markets mentioned by both Clinton and Graefe, including trader sophistication and the prevalence of arbitrage opportunities, likely evolved as prediction markets matured.

Chernoff & Jagtiani (2024) explore how high-net-worth traders -whales- influence prices and outcomes in cryptocurrency markets. They identify differences in trading behavior between whales and small volume traders and assess how they affect price formation, volatility, and market efficiency. They observed a consistent outsized impact from whales, highlighting limits to efficiency in crypto markets due to whale manipulation distorting prices and liquidity in comparison to traditional financial markets. Shen & Shi (2025) explore these behaviours further, constructing an Artificial Bitcoin Market (ABM) to reproduce observed behaviour from the real Bitcoin market and simulate the entry of Whales. They found a sharp increase in daily price volatility, upwards of 100%, when simulating a 5% increase in the proportion of whale investors. Although not identical, crypto and event markets share many structural and behavioral characteristics, and current literature on whale behaviour and effects creates a base expectation for effects on prices in prediction markets. Many of the empirical methods used to identify whales and calculate their effects may be of use in this thesis.

Although novel, there is plenty of literature regarding prediction market prices, accuracy, and efficiency; however, many of this precedes modern platforms that have revolutionized the prediction market space. Literature on whale behaviour in the crypto market also provides a close representation of their behaviour on speculative, information-driven markets on easy-to-enter digital platforms. However, despite many instances of research on Whales in different financial markets, there is no existing literature exploring whale behaviour in prediction markets themselves. The existing literature provides a solid baseline for this analysis to take from and build upon to fill this knowledge gap.

Data Summary

Sources and Scope

The data being used for this analysis is provided by Polymarket, a decentralized event betting platform. Data was downloaded from their servers using their Gamma and Data API's. A data set represents a market throughout its entire life span, with each observation in the set representing an individual trade. An observation includes the associated wallet (proxy identifier), the position taken, the outcome chosen, the size of the position, and the price per contract. It additionally includes a timestamp of the transaction, market/asset identifiers, resolution details, and user identifiers. A representative observation is shown in Figure 1.

Figure 1. Example Observation

proxyWallet	**side**	**asset**	**conditionId**	**size**	**price**	**timestamp**	**title**	**slug**	**icon**	**eventSlug**	**outcome**	**outcomeIndex**	**...**
0xea1cf96f16d	BUY	115423132994	0xd5893d2696	2487.89	1.00	1772348349.0	Fact Check: is	fact-check-is-la	https://polymar	fact-check-is-la	No	1.00	...

Prediction markets on Polymarket are accessible to users globally; however, due to the anonymous nature of blockchain transactions, geographic information on participants is not observable. Each market's dataset spans its entire trading period, with observations reflecting real-time trading activity.

For the purposes of this analysis, the key variables include price, volume, and position indicators, which together capture the direction and magnitude of each transaction. Timestamps are particularly important, as they allow us to place each of a whales transactions on a moment in time, enabling the calculation of subsequent price dynamics relative to these events across different time horizons.

Data Selection

Markets chosen for this study vary in subject, category, and size. They consist of 40 randomly selected markets evenly distributed across each of the major subject categories on Polymarkets. Totalling roughly 30,000 individual observations. Market size varies between $5,000 and $500,000 USD. By varying market parameters and types, I aim to mitigate distortions and biases that may originate from certain categories. For instance, certain markets may attract emotionally driven traders (e.g., sporting markets), information asymmetries (e.g., financial & political markets), or specialized traders (e.g., niche topic markets). By incorporating a diverse set of market types, the analysis aims to reduce the influence of said distortions and better isolate the effects of Whales across varying market types.

Limitations

Although Polymarkets API's provide highly informative data points, there are two main constraints with the data. One, it can not be guaranteed that whales are trading from one proxy wallet; it would not be hard for coordinated trading efforts to move large amounts of money across markets to avoid leaving a traceable trade. Thus, many of the more elaborate whales may go undetected or have their true trading volume underestimated. Two, the Polymarkets API only allows a user to download up to 3,500 transactions per market, meaning any market with more transactions than that would be incomplete. In order to observe resolved markets in their entirety, the analysis focuses on the previously specified size range ($5k-$500k USD).

Data Transformations

To prepare the data for analysis, unused user and identifier variables were dropped. A wallet statistics data frame, including total volume, count, etc was created by grouping market data by *proxyWallet*. Total market volume was then calculated to identify whales, defined as traders with greater than a 5% share of market volume. This provides a list of *proxyWallet* IDs for whales in each market. By mapping this list to the original market data, each transaction made by a whale is flagged. Additionally, *timestamp* values are converted to dates and used to sort data. Price is then calculated at time horizons: 1 hour, 6 hours, 1 day, and 3 days after a whale purchase, as well as before the next whale purchase, to create differences in price values for each of those time intervals. Data for each market is then categorized and aggregated into a single data frame to perform analysis on.

To prepare the data for analysis, I begin by importing the raw trade-level datasets and removing non-essential metadata (e.g., userID, profileImage, etc). Prices are then standardized into a common outcome, adjusting observations such that all price movements reflect the probability of the reference outcome. This ensures consistency across different positions. Timestamp variables are parsed into datatime format and sorted chronologically to preserve the market's structure. A Whale threshold is defined as a fixed share of the total market volume by resolution; trades exceeding this threshold are marked accordingly. For each whale trade, prices are calculated after multiple time horizons (1 hour, 6 hours, 12 hours, 1 day, 3 days), using nearest-neighbor (chronologically) to calculate post-trade price change. Raw price changes are then adjusted for the direction, creating “Signed” changes. Calculations for the price between consecutive whale trades were also performed. The dataset is then structured into a cleaned and standardized table for comparative analysis.

Data Preview

To best understand the structure and characteristics of the final dataset, the analysis begins with an exploratory data analysis. The dataset consists of transaction-level data for all trades. Below are summary statistics for all relevant variables for both all observations and identified whale trades at the 5% threshold.

Table 1. Complete data summary

	Size	**Price**	**Time to Resolution**	**Outcome Index**	**Price of Outcome**	**Whale Dummy**	**Signed Change 12h**	**Signed Change 1d**	**Signed Change 2d**	**Signed Change 3d**
count	29167	29167	29167	29167	29167	29167	29167	29167	29167	29167
mean	139.426	0.597	983.228	0.470	0.443	0.003	0.003	0.008	0.011	0.011
min	0.010	0.001	0.000	0.000	0.001	0.000	-0.956	-0.956	-0.956	-0.956
25%	3.940	0.300	104.993	0.000	0.041	0.000	-0.010	-0.012	-0.020	-0.020
50%	16.930	0.640	491.133	0.000	0.390	0.000	0.000	0.000	0.000	0.000
75%	81.979	0.970	1398.662	1.000	0.830	0.000	0.013	0.020	0.029	0.029
max	55871	0.999	7664.198	1.000	0.999	1.000	0.969	0.950	0.960	0.960

Table 2. Whales (5% Threshold) data summary

	Size	**Price**	**Time to Resolution**	**Outcome Index**	**Price of Outcome**	**Whale Dummy**	**Signed Change 12h**	**Signed Change 1d**	**Signed Change 2d**	**Signed Change 3d**
count	77	77	77	77	77	77	77	77	77	77
mean	11493.202	0.677	665.032	0.519	0.596	1.000	-0.001	0.008	0.008	0.008
min	498.166	0.001	0.000	0.000	0.001	1.000	-0.201	-0.201	-0.489	-0.489
25%	1256.540	0.327	2.101	0.000	0.020	1.000	-0.002	-0.005	-0.008	-0.008
50%	3644.250	0.949	79.361	1.000	0.874	1.000	0.000	0.000	0.000	0.000
75%	10186.390	0.994	400.095	1.000	0.991	1.000	0.007	0.007	0.009	0.009
max	55871.000	0.999	6025.655	1.000	0.999	1.000	0.125	0.719	0.815	0.815
std	16481.234	0.406	1542.937	0.503	0.433	0.000	0.044	0.094	0.121	0.121

To further understand the data, we examine pairwise correlations between the main variables, including size, price, and subsequent changes. These correlations provide initial indications of relationships in our data and flag potential multicollinearity amongst explanatory variables. They revealed no clear patterns in our data. Correlation matrices were made for both all trades and whale trades and can be found in the appendix.

Finally, plotting the market price paths and flagging the observed whale trades allows us to visually identify potential outliers or irregular observations in the dataset. By overlaying whale transactions on the evolution of market prices, we can observe whether certain trades coincide with unusually large price movements or broader market events. While such extreme observations may reflect genuine market dynamics, they may also exert disproportionate influence on regression estimates. Thus, these observations are carefully considered before accounting for them in empirical analysis. Below is a visualization of one of the markets I observed; the rest of the markets with identified whales at 5% can be found in the appendix.

Figure 2. Market Visualizations

Estimation Strategy

Estimation Equation & Method

To evaluate the impact of large trades on prediction market prices, we estimate regressions where the dependent variable is the signed change in contract price following a whale trade. Specifically, we examine price changes at multiple time horizons (1 hour, 6 hours, and 1 day) following the transaction. We do so via the following regression:

$$SignedChange_{i,t+h} = \beta\, Whale_{i,t} + \gamma\, TimeToResolution_{i,t} + \delta P_{i,t} + \alpha_m + \varepsilon_{i,t}$$

Where $SignedChange_{i,t+h}$ represents the change in price from time t to horizon $t + h$, adjusted for the direction of trade i (i.e., positive for buys and negative for sells), such that it captures whether prices move in line with the trader's position. The variable $Whale_{i,t}$ is a binary indicator equal to one if the trade is classified as a whale transaction. $TimeToResolution_{i,t}$ controls for the mechanical tendency of prices to converge toward their realized outcomes as the resolution date approaches. The variable $P_{i,t}$ denotes the price at the time of the trade, capturing baseline probability levels, while α_m represents market fixed effects that account for unobserved, time-invariant differences across markets.

The estimation equation is estimated using ordinary least squares (OLS). Because the dependent variable, signed price change, is continuous, OLS provides a straightforward method for estimating the relationship between whale trades and subsequent price movements. To account for differences across markets, fixed effects are included in the specification. Standard errors are clustered at the market level to account for potential correlation of errors within markets over time.

Identification Strategy

The empirical strategy follows an event-study-style regression in which whale trades are treated as information events. For each identified whale transaction, we measure subsequent price movements over several time horizons. If whale traders possess superior information, prices should move in the direction of their trades after the transaction occurs. Conversely, if whale trades merely reflect temporary liquidity shocks, price changes should revert over time.

Importantly, price movements following whale trades should also be influenced by how other market participants perceive these trades. If other traders interpret large transactions as signals of informed trading, they may update their beliefs and trade in the same direction, amplifying the price movement. By examining systematic price responses following whale trades across multiple markets, we can assess whether these traders contain additional informational content to the average trader and how others perceive them via how price movements reflect broader market reactions to perceived informed trading. The coefficient of interest, β, captures the average difference in price movements between whale and non-whale trades. If statistically significant, β it would indicate whale trades contain informational content or are large enough to substantially affect liquidity.

Results

Table 3. Effect of Whale Trades on Signed Price Changes

	2.5%			5%			10%			15%			20%		
Horizon	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**
12 hours	0.0028	0.0066	0.4176	0.0012	(0.0081)	0.1500	(0.0040)	0.0129	(0.3116)	(0.0027)	0.0048	(0.5542)	0.0148	0.0243	0.6090
1 day	0.0041	0.0088	0.4677	0.0060	(0.0127)	0.4700	(0.0124)	0.0129	(0.9613)	(0.0034)	0.0036	(0.9552)	0.0144	0.0213	0.6794
2 days	0.0026	0.0097	0.2661	0.0012	(0.0133)	0.0900	(0.0100)	0.0170	(0.5851)	(0.0047)	0.0051	(0.9235)	0.0187	0.0295	0.6302
3 days	0.0026	0.0097	0.2661	0.0012	(0.0133)	0.0900	(0.0100)	0.0170	(0.5851)	(0.0047)	0.0051	(0.9235)	0.0187	0.0295	0.6302

Notes: The sample includes all trades across markets.

Regression Estimates

The table reports estimated coefficients on the whale indicator across multiple time horizons and whale thresholds, ranging from 12 hours to 3 days and 2.5% to 20%. Across all specifications, estimated coefficients are statistically insignificant, with standard errors generally exceeding estimates. At the baseline 1-day horizon, the coefficient is close to zero; similar patterns extend to both shorter and longer horizons. Analysis across different whale thresholds yields comparable results, with estimated coefficients being imprecise and lacking consistent directional patterns. There is no clear trend in magnitude or sign as either time horizons or thresholds change.

Magnitude of Results

While statistically insignificant, the estimated magnitudes suggest that extremely large trades may in some cases have the ability to generate non-trivial shifts in prices, particularly at the 20% whale threshold, where coefficients imply changes of approximately 1.4-1.8 percentage points. However, these effects are highly variable and lack consistent directional patterns. Given the substantial market volume required to reach this threshold, it is plausible that observed changes in price reflect short-term liquidity effects rather than persistent informational impact. It should also be noted that these results are based on only four observations exceeding the 20% threshold and thus should be interpreted with caution.

Interpretation of Results

Economically, this near-zero estimation suggests that prediction markets exhibit high degrees of information efficiency. If market participants interpreted large trades as signals of superior information, one would expect prices to subsequently move in the direction the trade is executed in, as often observed in cryptocurrency and other financial asset markets. Instead, there is a clear absence of systematic post-trade price movements, implying the market absorbs whale trades without persistent distortions to prices and thus odds. Indicating that at this market size, even large traders are, on average, unable to meaningfully influence prices.

Resilience to disruption likely stems from the nature of the assets on the exchange. Unlike traditional financial assets, where valuations are ambiguous and clouded, prediction markets are centered on specific, real-world outcomes. This likely means participants form relatively concrete and informed beliefs when trading, resulting in market participants that are not easily swayed from their pre-formed expectations on the market and resolution. These findings suggest that the strength of belief aggregations in prediction markets limits the ability of concentrated capital to influence prices on average, reinforcing the market's role as an efficient mechanism to aggregate dispersed information, fulfilling their intended role.

Limitations

These results have several limitations. Trade size may simply not be a sufficient proxy for informed trading. Given the open nature of these markets, anyone with sufficient funds could make trades large enough to pass our threshold for whale without any particular domain knowledge. Additionally, the anonymousness that Polymarket's platform provides would provide any trader with additional knowledge to easily go untracked by the market, simply by having their volume dispersed over multiple trades or wallets. Thus, one is unable to guarantee that traders are working under one proxyWallet or attempting to influence the market all at once.

Our empirical specification may be subject to omitted variable bias. External information shocks, news events, or other changes in the situation a market observes could simultaneously influence trading behaviour and prices without being controlled for. Trying to capture the influence of very large whales also has its limits due to a relatively small number of observations. This leads to noisy estimates and reduced statistical power.

Finally, due to Polymarkets API limiting downloadable trades per market to 3500, we have only been able to observe "small" markets (roughly 500k or under in total volume). In these markets, the inconsistency in trades means we are unable to run a regression with enough data points for any time horizon smaller than 12 hours. Additionally, in large markets, a response to a "whale" may be different simply because of the value of the necessary volume traded to be considered so. For example, although 30,000 and 3,000,000 may be the same in terms of market share, given the right markets, simply trading 3,000,000 may be perceived by traders as a signal of information asymmetry, and thus a market would respond differently.

Thus, while the results suggest limited price impact from whale trades, they should not be interpreted as evidence that they have no influence on prediction markets. A lack of statistically significant findings may be due to the limited data set, and does not rule out economic impact in specific instances. The analysis also fails to capture longer-term effects or strategic behaviours, such as gradual positions or accounting for revealed information over time. These findings can also not be taken to apply across all prediction market platforms, as liquidity levels, event types, and other core aspects of markets may vary.

Robustness

To ensure the robustness of our results, I tested the regression across alternative Whale definitions and time horizons. Using stricter thresholds, 7.5% and 10% of market volume, yields similar results, with coefficients remaining small and imprecisely estimated. At very high thresholds, such as 20%, a small effect is observed; the limited observations at this size caused volatility in estimates. Thus, at no point did a consistent statistically significant relationship emerge. Across shorter and longer time horizons, ranging from 12 hours to 3 days, results also remained consistent, with all coefficient values consistently near zero. Overall, the results provide no evidence that whale trades systematically predict subsequent price movements.

Conclusion

This paper set out to find how high-volume traders, ‘Whales’, influence prices in prediction markets. The results provided limited evidence on a statistically significant impact of such activity on prices. Even in terms of magnitude, effects remain relatively limited, with the most noticeable shifts occurring only at the limited 20% whale threshold, where prices move by approximately 1.4–1.8 percentage points. Suggesting that, on average, the nature of prediction markets creates markets robust to concentrated trading behaviour and market power. This resilience to market disruptions highlights how prediction markets draw on concrete and informed user beliefs, allowing prices to be formed by educated expectations about specific outcomes rather than the uncertain estimations that often cloud traditional financial markets. However, this analysis is limited by its small data set of relatively small to medium-sized markets. Although it would be tempting to extrapolate the effects discovered, or lack thereof, to large volume markets, there is something to be said about responding trader behaviour relative to absolute value. Although a whale trading $5,000 or $5,000,000 may be equal in relativity to market volume, whales in markets with larger volumes may elicit different reactions simply due to the sheer size of capital they are willing to bet. Future research could expand on this by using richer datasets, with both more observations and markets larger than $500k in volume, to provide a more comprehensive understanding of how concentrated capital interacts with prediction market dynamics, particularly at larger absolute values.

Given my results, further exploration could attempt to incorporate the core mechanics of prediction markets, particularly the direct link between prices, probabilistic beliefs, and aggregation of dispersed information, into traditional financial asset markets. Such integration could enhance informational efficiency and price setting in markets, especially in environments that are currently characterized and uncertain and volatility. Finding a way to leverage the structures that make prediction markets resilient to disruptions and reflective of collective expectations is a promising next step towards creating financial markets that best aggregate information and accurately reflect collective beliefs.

Citations

Chernoff, Alan, and Julapa Jagtiani. 2024. Beneath the Crypto Currents: The Hidden Effect of Crypto "Whales." Federal Reserve Bank of Philadelphia Working Paper No. 24-14. https://www.philadelphiafed.org/-/media/frbp/assets/working-papers/2024/wp24-14.pdf.

Clinton, Joshua, and Ping Huang. 2025. Prediction Markets? The Accuracy and Efficiency of $2.4 Billion in the 2024 Presidential Election. SocArXiv. https://ideas.repec.org/p/osf/socarx/d5yx2_v1.html.

Cutting, Bryce, Ethan Harris, and Stephen Rao. 2025. Are Betting Markets Better than Polling in Predicting Political Elections? Preprint, Vanderbilt University. https://facultyprofiles.vanderbilt.edu/esploro/outputs/preprint/Are-Betting-Markets-Better-than-Polling/991044693428503276.

Goodell, McGee, & McGroarty. 2020. Election uncertainty, economic policy uncertainty and financial market uncertainty: A prediction market analysis. https://www.sciencedirect.com/science/article/abs/pii/S0378426619302584?via%3Dihub

Graefe, Andreas. 2017. Prediction Market Performance in the 2016 U.S. Presidential Election. International Journal of Forecasting 45 (1): 38–42. https://ideas.repec.org/a/for/ijafaa/y2017i45p38-42.html.

Saguillo, Julian, Tushar Lakshmanan, Laura Hartmann, and Michael Kearns. 2025. Unravelling the Probabilistic Forest: Arbitrage in Prediction Markets. arXiv:2508.03474. https://arxiv.org/abs/2508.03474.

Shen, Ying, and Wei Shi. 2025. The Role of Whale Investors in the Bitcoin Market. Journal of International Financial Markets, Institutions & Money. https://www.sciencedirect.com/science/article/abs/pii/S0275531925002648.

Reichenbach, Felix and Walther, Martin. 2026. Exploring Decentralized Prediction Markets: Accuracy, Skill, and Bias on Polymarket. SSRN Scholarly Paper ID 5910522. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5910522.

Appendix

Figure 1. Example Observation

proxyWallet	side	asset	conditionId	size	price	timestamp	title	slug	icon	eventSlug	outcome	outcomeIndex	...
0xea1cf96f16d	BUY	115423132994	0xd5893d2696	2487.89	1.00	1772348349.0	Fact Check: is	fact-check-is-la	https://polymar	fact-check-is-la	No	1.00	...

Figure 2. Market Visualizations

Figure 3. Extra Market Visualizations (for an interactive visualization of all the markets, refer to the code base.)

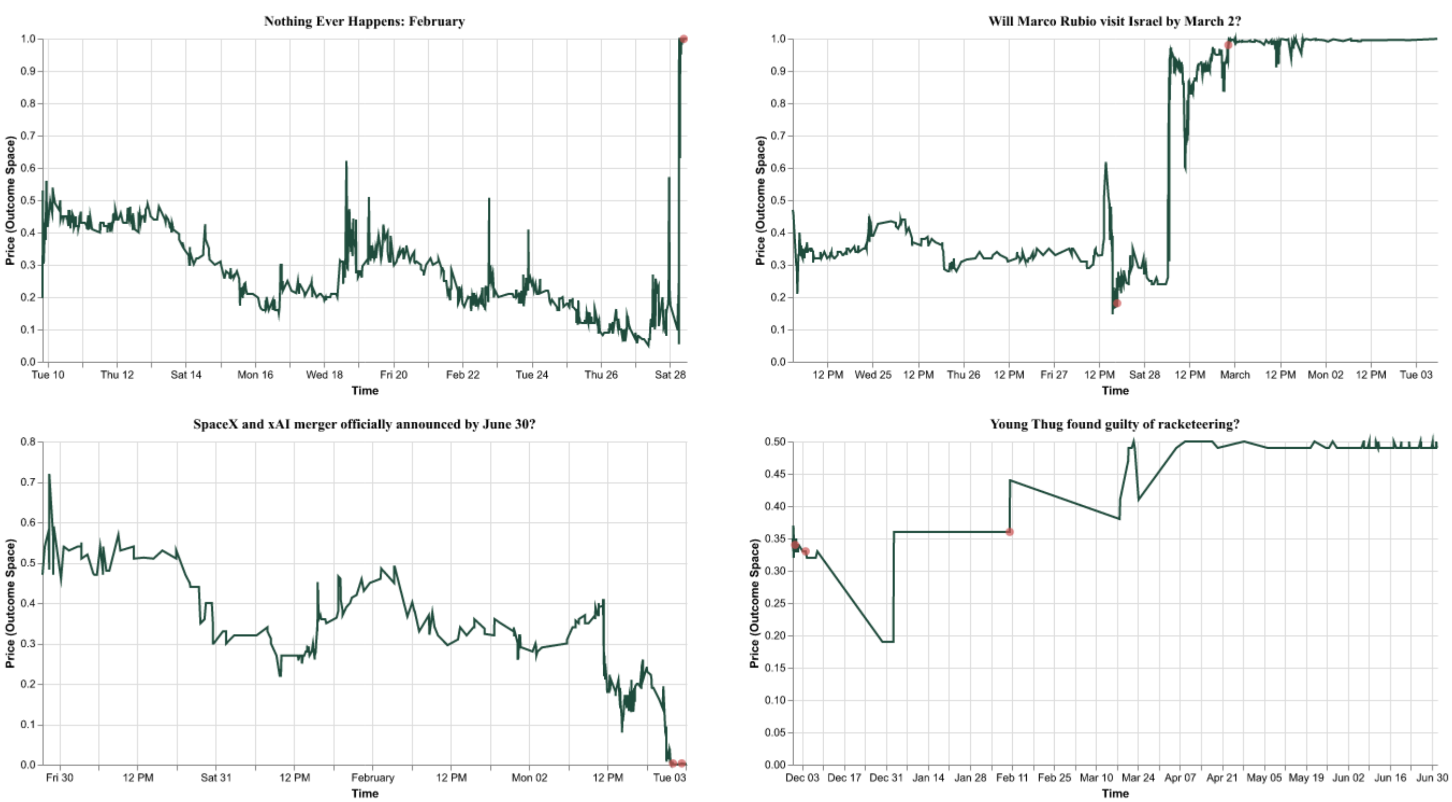

Figure 4. All trades correlation matrix

Figure 5. Whale trades correlation matrix

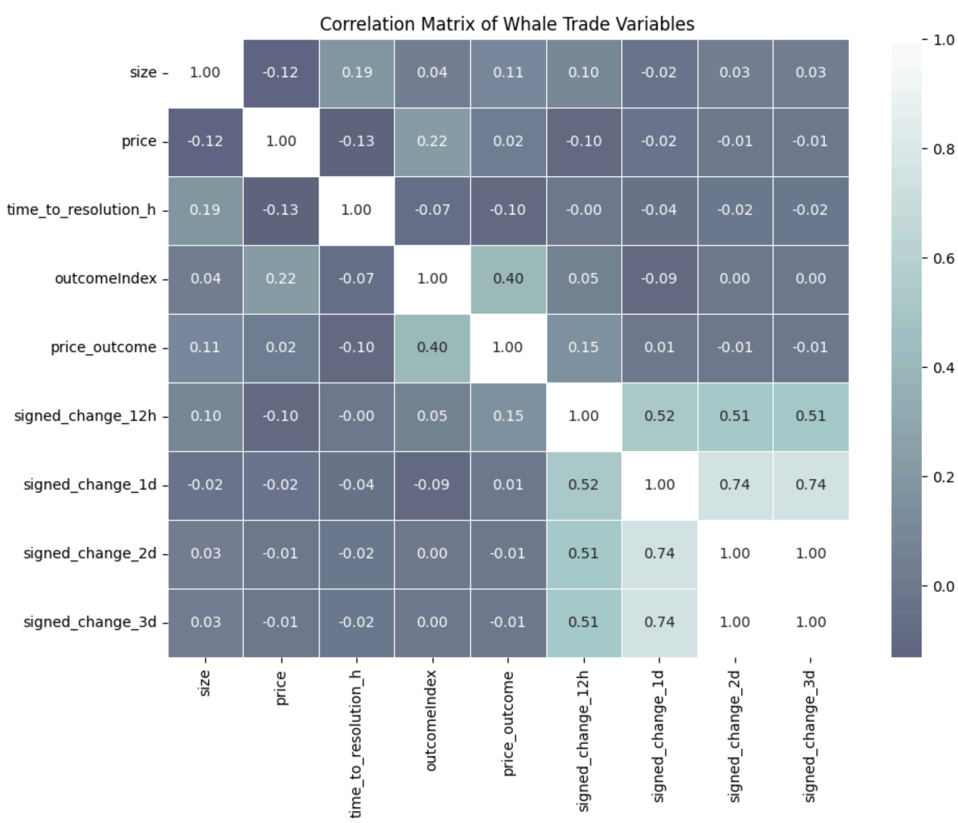

Table 1. Complete data summary

	Size	Price	Time to Resolution	Outcome Index	Price of Outcome	Whale Dummy	Signed Change 12h	Signed Change 1d	Signed Change 2d	Signed Change 3d
count	29167	29167	29167	29167	29167	29167	29167	29167	29167	29167
mean	139.426	0.597	983.228	0.470	0.443	0.003	0.003	0.008	0.011	0.011
min	0.010	0.001	0.000	0.000	0.001	0.000	-0.956	-0.956	-0.956	-0.956
25%	3.940	0.300	104.993	0.000	0.041	0.000	-0.010	-0.012	-0.020	-0.020
50%	16.930	0.640	491.133	0.000	0.390	0.000	0.000	0.000	0.000	0.000
75%	81.979	0.970	1398.662	1.000	0.830	0.000	0.013	0.020	0.029	0.029
max	55871	0.999	7664.198	1.000	0.999	1.000	0.969	0.950	0.960	0.960

Table 2. Whales (5% Threshold) data summary

	Size	Price	Time to Resolution	Outcome Index	Price of Outcome	Whale Dummy	Signed Change 12h	Signed Change 1d	Signed Change 2d	Signed Change 3d
count	77	77	77	77	77	77	77	77	77	77
mean	11493.202	0.677	665.032	0.519	0.596	1.000	-0.001	0.008	0.008	0.008
min	498.166	0.001	0.000	0.000	0.001	1.000	-0.201	-0.201	-0.489	-0.489
25%	1256.540	0.327	2.101	0.000	0.020	1.000	-0.002	-0.005	-0.008	-0.008
50%	3644.250	0.949	79.361	1.000	0.874	1.000	0.000	0.000	0.000	0.000
75%	10186.390	0.994	400.095	1.000	0.991	1.000	0.007	0.007	0.009	0.009
max	55871.000	0.999	6025.655	1.000	0.999	1.000	0.125	0.719	0.815	0.815
std	16481.234	0.406	1542.937	0.503	0.433	0.000	0.044	0.094	0.121	0.121

Table 3. Effect of Whale Trades on Signed Price Changes

	2.5%			5%			10%			15%			20%		
Horizon	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**	**β**	**Std. Error**	**z-Stat**
12 hours	0.0028	0.0066	0.4176	0.0012	(0.0081)	0.1500	(0.0040)	0.0129	(0.3116)	(0.0027)	0.0048	(0.5542)	0.0148	0.0243	0.6090
1 day	0.0041	0.0088	0.4677	0.0060	(0.0127)	0.4700	(0.0124)	0.0129	(0.9613)	(0.0034)	0.0036	(0.9552)	0.0144	0.0213	0.6794
2 days	0.0026	0.0097	0.2661	0.0012	(0.0133)	0.0900	(0.0100)	0.0170	(0.5851)	(0.0047)	0.0051	(0.9235)	0.0187	0.0295	0.6302
3 days	0.0026	0.0097	0.2661	0.0012	(0.0133)	0.0900	(0.0100)	0.0170	(0.5851)	(0.0047)	0.0051	(0.9235)	0.0187	0.0295	0.6302