Title

An Empirical Analysis of Arbitrage and Insider Trading on Polymarket NBA Markets

ISBN

9798277443699

Author

Yang, Jiaxin

Publication Date

2026-03-18

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

An Empirical Analysis of Arbitrage and Insider Trading on Polymarket NBA Markets

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Statistics

by

Jiaxin Yang

2026

ABSTRACT OF THE THESIS

An Empirical Analysis of Arbitrage and Insider Trading on Polymarket NBA Markets

by

Jiaxin Yang

Master of Statistics

University of California, Los Angeles, 2026

Professor Guang Cheng, Chair

This thesis leverages the transparent, on-chain architecture of Polymarket to empirically study algorithmic arbitrage and insider trading within NBA prediction markets. Analysis of high-frequency limit order book snapshots reveals extreme microstructural efficiency; single-market mispricings resolve in a median of 3.614 seconds, while combinatorial arbitrage is economically bounded by shallow liquidity. To detect insider trading without regulatory labels, an unsupervised Isolation Forest model is trained on pre-resolution behavioral features and calibrated against a quantitative Pseudo-Ground Truth (PGT) Oracle. The model's top 1.0% anomaly cohort captured a statistically significant 11.7% of aggregate market profits. These findings demonstrate that while structural inefficiencies on Polymarket are fleeting, distinct on-chain behavioral signatures effectively isolate insider trading.

The thesis of Jiaxin Yang is approved.

Mikhail Chernov

Nicolas Christou

Guang Cheng, Committee Chair

University of California, Los Angeles

2026

To my family and friends

Contents

List of Figures

List of Tables

Acknowledgements

This thesis would not have been possible without the help from Professor Guang Cheng, Anderson Zou, Minrui Gui, and Haoting Dong.

Chapter 1

Introduction

1.1 Background and Motivation

Prediction markets are financial platforms where participants trade on the outcomes of future events, with prices reflecting the market's aggregate probability estimates. Polymarket is one of the most prominent decentralized prediction market platforms, operating on the Polygon blockchain, where users speculate on future events by trading shares tied to binary or categorical outcomes. Since its launch in 2020, Polymarket has grown into an influential forecasting venue, attracting substantial liquidity across a wide range of topics, including politics, sports, cryptocurrency, technology, and macroeconomics.

Academic interest in Polymarket has grown alongside its prominence, though existing research remains narrowly focused. The bulk of studies examine its role in political forecasting, particularly electoral prediction and investigate the broader question of whether prediction market prices constitute efficient probability estimates of real-world outcomes [11]. Comparatively little attention has been paid to trading behavior and market microstructure within Polymarket. This gap is partly methodological: meaningful analysis requires integrating and decoding complex on-chain data, which demands careful data engineering before any market analysis can be conducted.

Yet this challenge is also an opportunity. Unlike traditional financial markets, Polymarket's on-chain architecture makes every trade, order placement, and wallet interaction publicly auditable, a level of data transparency that is essentially unavailable in conventional equity or derivatives markets. This thesis exploits that transparency to conduct a systematic study of two core aspects of trading in Polymarket: algorithmic arbitrage and insider trading. Using NBA game markets as the empirical setting, we analyze (1) single-market arbitrage opportunities arising from mispricings across outcome shares within a single market, (2) combinatorial arbitrage opportunities that span multiple related markets simultaneously, and (3) the detection of insider trading through machine learning methods applied to on-chain behavioral features. Together, these analyses provide a structured lens through which to examine market efficiency and information asymmetry in a decentralized prediction market.

1.2 The NBA Market as a Research Environment

While Polymarket hosts prediction markets across many domains, this thesis focuses specifically on NBA game markets, markets that resolve on the outcome of individual NBA games (win/loss, point spreads, player performance, etc.). This choice is motivated by several considerations.

First, NBA markets exhibit high trading activity and substantial liquidity on Polymarket. NBA markets were created on Polymarket since October 2024, and the trading volume (USDC) in NBA markets increased from \$51M in 2024 to \$0.89B in 2025, representing a 17.4-fold increase. For context, the total trading volume in sports markets on Polymarket reached \$3.11B in 2025, meaning NBA markets account for nearly 30% of all sports market activity.

Second, NBA markets have a structural regularity that is ideal for systematic analysis. The regular season runs from October through April, generating 1,230 games per season. Each game produces a fresh set of markets with consistent structure, the same outcome types,

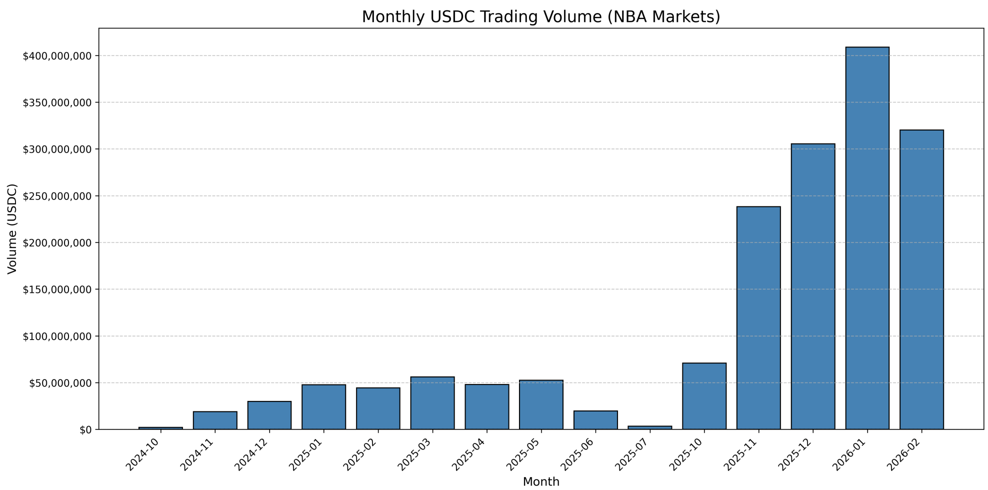

Figure 1.1: Monthly trading volume of NBA markets from Oct 2024 to Feb 2026

the same resolution logic, and comparable liquidity profiles, making cross-market comparisons tractable and findings generalizable.

Third, the high-frequency, recurrent nature of NBA games provides the statistical power needed to distinguish signal from noise in both arbitrage analysis and insider detection. Unlike one-off political events, where a single market may attract outsized attention, NBA games provide a large, relatively homogeneous sample.

For these reasons, NBA markets serve as the empirical playground for observing market inefficiencies and insider trading. By utilizing the NBA as a focused case study, this research establishes a data-driven framework that can be easily generalized to, more broadly, any class of Polymarket events with sufficient trading volume and structural regularity.

1.3 Arbitrage in Prediction Markets

Arbitrage, the practice of exploiting price discrepancies to earn risk-free profit, is a natural phenomenon in markets where prices are set by dispersed participants with heterogeneous information and reaction speeds. In Polymarket, where outcome probabilities must sum to one by definition, mispricing across complementary outcomes creates systematic arbitrage opportunities. Saguillo et al. provides the closest precedent, conducting an empirical arbitrage analysis on Polymarket data across market categories [14].

This thesis follows that line of inquiry but departs from it in one key methodological respect: rather than relying on executed trades data, we conduct our analysis using order book snapshots. Order book snapshots capture the state of the market at discrete points in time, enabling us to (i) detect arbitrage opportunities as they emerge in real time, (ii) measure their duration (i.e., how long a mispricing persists before it is corrected), and (iii) characterize whether opportunities are ever actually executable given prevailing liquidity. This real-time, duration-aware perspective is a methodological contribution beyond prior work.

1.4 Insider Trading

Insider trading, trading by parties with privileged, non-public information, is a well-studied problem in traditional financial markets [4], where legal frameworks and regulatory surveillance provide both a definition and an enforcement mechanism. In decentralized prediction markets like Polymarket, the situation is more complex. Each participant is represented solely by a pseudonymous wallet address; there are no disclosed identities, no regulated disclosures, and no verified insiders. However, this pseudonymity can make the decentralized platforms highly attractive to individuals possessing asymmetric information. Recent federal investigations revealing the direct involvement of NBA players and coaches in illicit sports betting networks highlight that insider wagering is an active, real-world issue within the league [17]. Given the proven existence of these bad actors, establishing a framework for detecting insider trading on platforms like Polymarket is necessary.

The theoretical foundation for detecting such insiders in sports markets is rooted in the work of Shin (1992). Shin demonstrated that prediction markets naturally exhibit a "favorite-longshot bias", where longshots are systematically overvalued and favorites are undervalued, as a rational pricing defense mechanism against insiders [15]. Market makers adjust odds because they know insiders will exploit non-public information to take concentrated positions on higher-probability outcomes, rather than gambling on extreme longshots. However, because traditional sportsbooks are anonymous, prior research has only been able to infer the presence of these insiders from aggregate price movements. Polymarket's transparent, on-chain architecture provides a unique opportunity to move beyond aggregate biases and actually identify the specific wallets acting as insiders.

To bridge this theoretical literature with empirical on-chain data, a verifiable baseline of insider trading is required. While several third-party tools and dashboards have emerged, such as Dune Analytics and Polysights, that offer heuristic metrics like fresh wallets or whale trades to flag potential insiders, the lack of a verified ground truth on Polymarket means such heuristics cannot be formally validated. To solve this, this thesis introduces a Pseudo-Ground Truth (PGT) Oracle. Following the unsupervised machine learning framework for insider identification proposed by Cheng et al. [4], this thesis leverages the PGT Oracle to construct features from uniquely available on-chain data, applying a robust anomaly detection and validation pipeline to isolate insiders in NBA markets.

1.5 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 introduces the problem setting: it covers the mechanics of Polymarket markets (order books, CLOB structure, outcome tokens, resolution), the specific structure of NBA markets on Polymarket, and a detailed walkthrough of the available data sources, both the off-chain REST APIs and the on-chain event data. Chapter 3 presents the single-market arbitrage analysis. Chapter 4 extends this to combinatorial arbitrage across related markets. Chapter 5 presents the insider trading detection methodology and results. Chapter 6 concludes with a discussion of limitations, broader implications, and directions for future research.

Chapter 2

Problem Setting and Data Architecture

2.1 Introduction to Polymarket and Core Mechanics

To contextualize the data and trading behaviors analyzed in this thesis, it is essential to first understand the operational structure of Polymarket. Polymarket is a decentralized prediction market platform built on the Polygon blockchain. Unlike traditional sportsbooks or betting platforms where users wager against a centralized "house" that sets the odds, Polymarket operates as a peer-to-peer exchange. Participants trade shares that represent the probability of future real-world occurrences [12].

2.1.1 Core Terminology: Events, Markets, and Outcome Tokens

The Polymarket ecosystem is structured around a specific hierarchy of concepts:

- **Events:** An event is a container that groups one or more related markets together. For example, each NBA game corresponds to an NBA event, which includes multiple markets.
- **Markets:** A "market" is the fundamental, tradable unit within an event. Each market is structured as a specific question with binary outcome slots. For example, within an NBA event, a specific market might be "Will the Los Angeles Lakers win Golden State

Warriors on Feb 28, 2026?" or "Will the Los Angeles Lakers win Golden State Warriors by 3 or more points on Feb 28, 2026?"

- **Outcome Tokens:** When users trade on a market, they are buying and selling "outcome tokens" (specifically, ERC-1155 tokens recorded on the blockchain). For every binary market, there is a "Yes" token and a "No" token. These tokens are fully collateralized; every Yes/No pair in existence is backed by exactly $1.00 of USDC (a stablecoin pegged to the US Dollar) locked in a smart contract.

2.1.2 Pricing as Probability

A defining feature of prediction markets is that the price of an outcome token directly translates to the market's collective estimation of that outcome's probability. Token prices on Polymarket strictly range between $0.00 and $1.00. If a "Yes" token for a specific NBA team winning a game is trading at $0.65, it indicates that the market's implied probability of that team winning is 65%.

Because the outcomes are mutually exclusive and fully collateralized, the combined price of the "Yes" token and the "No" token for a given market will naturally sum to $1.00 (discounting minor variations due to the bid-ask spread). Traders can buy tokens if they believe the true probability is higher than the market price, or sell them if they believe it is lower. Crucially, traders are not locked into their positions; they can buy and sell tokens at any time before the event resolves to realize profits or cut losses as market sentiment shifts.

2.1.3 The Trading Infrastructure: Hybrid-Decentralized CLOB

Trading on Polymarket is facilitated through a Central Limit Order Book (CLOB). To balance the speed required for efficient trading with the security of blockchain technology, Polymarket utilizes a hybrid-decentralized model:

- **Off-chain matching:** Orders are submitted, continuously monitored, and matched by an off-chain operator. This allows for low-latency trading and deep liquidity without requiring users to pay blockchain transaction fees (gas) every time they place or cancel a limit order.

- **On-chain settlement:** Once a buyer and seller are matched in the order book, the actual atomic swap of outcome tokens and USDC collateral is executed and settled transparently on the Polygon blockchain via smart contracts. This ensures the system remains non-custodial; the platform never holds user funds directly.

2.1.4 Resolution and Payouts

When the event outcome becomes known (after the market's specified closing time), the market enters the resolution phase. Polymarket relies on UMA's Optimistic Oracle for decentralized, permissionless settlement. For sports markets, resolution sources are typically official league data or reputable aggregators (e.g., NBA.com), minimizing ambiguity. The entire process is on-chain and publicly verifiable. Upon resolution, tokens representing the correct outcome become redeemable for exactly \$1.00, and tokens representing the incorrect outcome become worthless (\$0.00).

2.2 The NBA Markets on Polymarket

On Polymarket, every individual NBA game is categorized as a discrete "event." Within the platform's backend and on-chain data, each event is tracked using a unique identifier known as an "event slug," which serves as the primary key for querying and aggregating market data. Within a single game event, Polymarket hosts a cluster of related markets. Each of these NBA markets is strictly binary, featuring two mutually exclusive outcomes.

2.2.1 Typology of NBA Markets

NBA markets on Polymarket are not monolithic; they vary significantly in duration, liquidity, and complexity. They broadly fall into two main categories:

Game-Level Markets (this paper's focus): These are short-term, recurring markets tied to individual matchups throughout the NBA season, and they form the primary focus of this thesis. For a single game event, the available markets typically include:

- **Moneyline:** A binary market predicting the outright winner of the game.
- **Spreads:** Markets predicting the outcome based on a specified point differential between the two teams.
- **Totals (Over/Under):** Markets predicting whether the combined score of both teams will be over or under a specified number.

Each of these markets is frequently offered in both "Full Game" and "1st Half" (1H) variations. Furthermore, starting in December 2025, Polymarket expanded these game-level events to include player proposition (prop) markets, allowing trading on individual player points, assists, and rebounds. Consequently, now a single NBA game event can encompass around 40 distinct markets, with about 14 Moneyline/Spread/Totals markets and over 20 player proposition markets. The summary of trading volume across different types of NBA markets is in Table 2.1.

Futures and Prop Markets: These are longer-term markets that resolve at the end of the season or a specific tournament phase. Examples include the outright NBA Championship Winner, Conference and Division Winners, and individual accolades such as the MVP or Rookie of the Year.

2.2.2 Structural Dependencies

While every NBA market trades independently on Polymarket's order books, their real-world outcomes are fundamentally correlated. A defining feature of the NBA prediction landscape

Market Type	No. of Markets	Mean No. of Trades	Total Volume	Mean Volume
Full Moneyline	176	10,412.13	251,207,897.58	1,427,317.60
Full Spread	769	366.78	41,057,412.35	53,390.65
Full Total	1,162	185.60	21,804,680.06	18,764.78
Player	5,074	13.32	1,285,140.84	253.28
1H	172	16.01	161,774.96	940.55
1H Total	457	6.12	48,107.17	105.27
1H Spread	341	5.94	26,928.93	78.97

Table 2.1: Summary of trading volume (USDC) across different NBA market types.

is the implicit mathematical hierarchy among these isolated markets. For instance, within a single game event, the moneyline and spread markets are highly interdependent. Because Polymarket's underlying smart contracts do not natively cross-margin or dynamically link the pricing of these separate markets, temporary pricing misalignments frequently occur across the ecosystem. These structural dependencies form the mathematical basis for the combinatorial arbitrage strategies explored in Chapter 4.

2.3 Polymarket Data Architecture and Retrieval

Polymarket's hybrid-decentralized infrastructure, where order matching occurs off-chain for low latency while final settlement occurs on-chain, generates a highly complex data ecosystem. To summarize the existing knowledge regarding Polymarket's data landscape, this section first categorizes all publicly available data sources before detailing the specific pipelines utilized for the empirical analysis in this thesis.

2.3.1 Overview of Available Data Sources

The Polymarket platform exposes market activity and user behavior through three primary data channels [13]:

1. **REST APIs:** A suite of off-chain interfaces that index and serve structured market data.

- Gamma API: A read-only service providing market metadata, discovery features, categorization tags, and aggregated statistics (e.g., volume, liquidity).
- CLOB API (Central Limit Order Book) The trading interface used to fetch live price quotes and order book depth for the arbitrage analyses in Chapters 3 and 4. It is important to note a structural data constraint: fetching the complete state across all active NBA markets requires 3.6 to 5.5 seconds per cycle. Consequently, the dataset is limited to this discrete polling frequency, restricting our ability to observe transient anomalies that resolve faster than this multi-second window.
- Data API: A specialized service focused on user-specific historical data, including open positions, past trades, and wallet activity.

2. **WebSocket Feeds:** Real-time data streams used primarily by automated market makers. These include public channels for live order book and price updates, private channels for user order execution notifications, and external data feeds (RTDS).

3. **On-Chain Data and Indexers:** The raw, immutable blockchain layer.

    - Direct RPC Endpoints: Allows developers to query the Polygon blockchain directly for raw smart contract event logs and transaction input data.
    - The Graph / GraphQL: Decentralized subgraphs that index Polymarket's on-chain events into a more queryable format, acting as a trustless alternative to the REST APIs.

2.3.2 Primary Data Sources Utilized

While the broader ecosystem offers multiple ways to ingest data, the methodology of this thesis relies heavily on specific REST API endpoints to capture the high-dimensional data required for analyzing arbitrage and insider trading in NBA markets:

- **Order Book Snapshots via the CLOB API:** To evaluate both single and combinatorial arbitrage opportunities, this study captures real-time market depth rather than relying on executed trade history. By querying the CLOB API's `GET /book` endpoint, this research constructs point-in-time snapshots of the limit order book. This endpoint retrieves aggregated bids and asks at various price levels. Capturing these snapshots enables the detection of transient, unexploited arbitrage windows and the calculation of true market liquidity.

- **User Portfolios via the Data API:** The insider trading analysis relies on tracking the behaviors of pseudonymous wallet addresses. This study utilizes the Data API to reconstruct trading timelines. Specifically, the `GET /positions` endpoint is used to retrieve a given user address's current and closed holdings, identifying the specific outcome tokens and share amounts held. Furthermore, the `GET /trades` endpoint provides a chronological ledger of a user's historical executions, including sizes, prices, and whether the wallet acted as a maker or taker.

2.3.3 On-Chain Data Validation and the Double-Counting Issue

To ensure the integrity of the trading volume reported in this thesis, we cross-validate data from the REST APIs against raw on-chain `OrderFilled` events. A common pitfall when computing volume from `OrderFilled` events is double-counting, which arises from how the smart contract handles order matching [16].

When a trade is executed, the contract follows a fixed sequence. It first transfers the taker's tokens to the exchange contract, then loops through each maker involved in the fill, computing fees, performing any necessary token splits or merges, and emitting a separate `OrderFilled` event for each maker. Once all makers are processed, the contract transfers the aggregated tokens to the taker, collects taker fees, emits a final taker-side `OrderFilled` event summarizing the full transaction, and closes with an `OrdersMatched` event.

This means every matched trade produces two types of `OrderFilled` events: per-maker events that describe each individual fill, and a single taker-side summary event that aggregates the entire transaction. These are two representations of the same economic exchange, not two separate trades. Summing all `OrderFilled` events without distinction therefore double-counts actual trading activity.

It is worth noting that the exchange contract is purely a routing engine: it holds no positions, provides no capital, and bears no risk. The true economic transfer occurs only between the taker and the makers.

Volume Calculation Methodology. Following [16], volume can be computed in three equivalent ways: taker-side volume (summing the taker-focused events), maker-side volume (summing the per-maker events), or via `OrdersMatched` events. As noted in [16], taker-side and maker-side volume converge over longer time horizons.

This thesis uses maker-side volume. The taker-side summary event, where the `taker` field is populated by the Exchange Contract address, blurs microstructural detail by reporting a single averaged execution price for the entire sweep. Retaining only the per-maker events preserves the exact price levels consumed at each step of the order book. All `OrderFilled` rows where the taker is the Exchange Contract are therefore removed. The remaining maker-focused events form the dataset used throughout this thesis, fully eliminating the double-counting anomaly while preserving the granularity of each fill.

Chapter 3

Single Market Arbitrage

3.1 The Dataset

This study uses orderbook snapshots data as introduced in Section 2.3.2. The high-frequency Limit Order Book (LOB) data are collected from the Polygon blockchain for 173 NBA games, with 3,042 markets, between Feb 4, 2026 and Mar 4, 2026. The dataset comprises 75,088,497 LOB snapshots. The example of orderbook snapshots data can be found in Appendix A.

3.1.1 Data Collection

To accurately identify arbitrage opportunities in a continuous limit order book (CLOB), the discrete, asynchronous updates of individual tokens must be reconstructed into a synchronized, holistic view of the market state. In a binary NBA game market (with Team A and Team B), an arbitrage opportunity exists only when the sum of the best available ask prices across all mutually exclusive outcomes is strictly less than the payout:

$$Ask_A + Ask_B < \$1.00$$

Because the data was collected via a REST API polling mechanism spanning a methodology update, a preprocessing pipeline was developed to reconstruct accurate market states and measure duration while strictly controlling for network latency and exchange outages.

3.1.2 Data Collection Architecture and Polling Latency

Orderbook snapshots were collected via the Polymarket REST API (`/books` endpoint). To ensure the highest possible synchronization between the two sides of a binary market (e.g., "Team A" and "Team B"), requests were batched by `event_slug`. This architectural choice yielded near-perfect intra-market synchronization: network round-trip time (RTT) was actively monitored and averaged 171 milliseconds, and server-side timestamps (`ts_utc`) confirmed that competing outcomes within the same market were recorded with a discrepancy of roughly between 2 milliseconds and 50 miliseconds. Furthermore, to minimize redundant storage, updates were persisted only when the orderbook's cryptographic hash changed, efficiently capturing true shifts in price or liquidity depth.

However, iterating sequentially through all active NBA markets introduced a mechanical polling loop. Analysis of unique chronological timestamps revealed a median interval of approximately 3.6 to 5.5 seconds between data requests for any specific market. Consequently, while intra-market prices are perfectly aligned, the absolute measurement of an arbitrage event's start and end times carries a systematic margin of error constrained by this polling frequency.

In algorithmic trading research, discrete polling inherently leads to the systematic underestimation of market inefficiencies. Because the script observes the market at roughly 3 to 5 second intervals rather than through a continuous websocket stream, the true onset of an arbitrage opportunity likely precedes the first snapshot that records it. Similarly, the opportunity may persist after the final confirming snapshot, resolving only moments before the subsequent poll detects its absence. Furthermore, highly transient opportunities that open and close entirely within a single polling interval evade detection completely. Because the pipeline only credits duration to actively confirmed snapshots, the frequencies and durations reported in this study represent a conservative lower bound of the true arbitrage available to market participants.

3.1.3 State Alignment

Database inserts were restricted strictly to actual changes in the orderbook state, verified by tracking the API's cryptographic hash of the market. Furthermore, the timestamp recorded (`ts_utc`) was transitioned to the Polymarket matching engine's server-side timestamp.

In this sparse paradigm, a timestamp gap indicates market stability, not a collection failure. If the hash for Team A changes but Team B remains identical, Team B is omitted from the API response to save bandwidth. To reconstruct the continuous market state, a forward-fill (`ffill`) operation is applied.

Micro-Desynchronization Mitigation: While the off-chain matching engine updates mirrored tokens synchronously, the aforementioned 2 to 50 millisecond serialization delay in the REST API introduces a significant vulnerability when reconstructing the limit order book. If left uncorrected, applying a forward-fill to staggered timestamps creates a transient market state. The algorithm momentarily crosses the new price of one outcome against the stale price of its complement, falsely registering a sub-second phantom arbitrage.

To prevent this microstructure artifact, a dynamic time-based clustering algorithm was applied prior to state alignment. Sequential updates occurring within 500 milliseconds of each other were classified as originating from the exact same API polling batch. These clustered events were algorithmically synchronized to share a single discrete timestamp. Only after this synchronization was the forward-fill operation executed. This mathematically projects the last-known valid limit orders forward in time until a new synchronized batch update overrides them, allowing for the accurate calculation of $Ask_A + Ask_B$ at any subsequent state change while completely eliminating API serialization noise.

3.1.4 Forward Duration Measurement

In a sparse dataset, counting the number of recorded rows is mathematically invalid for measuring time, as a static orderbook will only generate one row regardless of whether it persists for five seconds or five minutes. To measure the duration of any given market state (and consequently, any arbitrage episode), a forward-looking difference calculation was utilized:

$$Duration = Timestamp_{n+1} - Timestamp_n$$

The Trust Ceiling (Outage Protection): To ensure that actual exchange outages or halted markets were not mistakenly classified as hours-long arbitrage opportunities, a dynamic "Trust Ceiling" was applied to the forward duration calculation. The ceiling was defined by the natural cadence of the game phase:

- Pre-Game & Post-Game: 30 minutes (1800 seconds).
- In-Game: 5 minutes (300 seconds, accounting for extended television timeouts or official reviews).

If the forward difference between two rows exceeded this ceiling, the duration credited to that state was strictly capped at the ceiling limit.

Dynamic Polling Fallback:

For terminal states (the final row of an episode, where $Timestamp_{n+1}$ does not exist), hardcoding an arbitrary duration (e.g., 1.0 seconds) would distort the data. Instead, the preprocessing script calculates the exact median gap between all unique snapshots for that specific market. This dynamically calculated latency (typically ~ 4.0s) serves as the mathematical baseline, ensuring terminal single-row events are credited precisely with the natural mechanical heartbeat of the data collector on that specific day.

3.2 Methodology: Arbitrage Detection

This section outlines the mathematical framework and procedural pipeline used to identify pricing inefficiencies within the Polymarket orderbook. It details the theoretical conditions for arbitrage, the treatment of the mirrored liquidity pool, and the strict filtering criteria applied to ensure identified opportunities are actionable.

3.2.1 Mathematical Definition of Arbitrage in Binary Markets

In a strictly binary prediction market, the underlying asset resolves to either 1 (True) or 0 (False), guaranteeing a combined payout of exactly \$1.00 for holding one share of each outcome. Therefore, a risk-free arbitrage opportunity exists whenever the cost to acquire exposure to both outcomes is misaligned with this guaranteed payout. This study evaluates two distinct mechanical paths for arbitrage execution as introduced in [14]:

1. Long Arbitrage (The Buy Path): An opportunity exists when a trader can simultaneously buy both mutually exclusive outcomes from the limit order book for a combined cost strictly less than \$1.00. This requires executing against the best available Ask prices.

$$Cost = Ask_A + Ask_B$$

$$Profit_{Long} = 1.00 - Cost \quad (\text{where } Cost < 1.00)$$

2. Short Arbitrage (The Mint-and-Sell Path): A trader can lock \$1.00 of USDC collateral into the smart contract to "mint" one share of Team A and one share of Team B. An opportunity exists if the trader can immediately sell both shares to the orderbook for a combined revenue strictly greater than the \$1.00 collateral cost. This requires executing against the best available Bid prices.

$$Revenue = Bid_A + Bid_B$$

$$Profit_{Short} = Revenue - 1.00 \quad \text{(where } Revenue > 1.00\text{)}$$

3.2.2 Polymarket Microstructure and the Mirrored Orderbook

As stated in Section 2.1.3, Polymarket utilizes a Central Limit Order Book (CLOB) architecture. To maximize execution speed, the matching engine operates off-chain, while trade settlement and escrow are handled on-chain via smart contracts deployed on the Polygon blockchain.

A unique feature of this CLOB is its mirrored liquidity design. There is only one shared pool of underlying liquidity for any binary market. The matching engine automatically reflects orders across the complementary tokens. For example, if a market maker places a limit order to buy 100 shares of Outcome A (YES) at \$0.40, the off-chain engine simultaneously generates a synthetic limit order to sell 100 shares of Outcome B (NO) at \$0.60. The inversion is a purely mathematical reflection ($\$1 - P_{original}$). Consequently, querying the REST API for Outcome B yields an effective orderbook that combines direct orders placed on Outcome B with the synthetic, mirrored orders originating from Outcome A.

3.2.3 Signal Consolidation and the Double-Counting Problem

The mirrored nature of the liquidity pool introduces a systemic risk of double-counting arbitrage signals. If a severe mispricing occurs, an anomaly where $Ask_A + Ask_B < 1.00$ will automatically manifest as a crossed synthetic book where $Bid_A + Bid_B > 1.00$. Naively summing the profit from both the Long and Short signals would falsely double the executable profitability of the market.

To resolve this, the analysis pipeline evaluates both paths independently at every timestamp available in the data to capture all possible price discrepancies, but enforces a strict deduplication protocol. The analysis consolidates the arbitrages by isolating the simultaneous signals and retains only one path in the results.

3.2.4 Signal Validation and Execution Constraints

To ensure the identified inefficiencies represent actionable trading opportunities rather than phantom signals, two strict filters were applied to the data:

1. **Top-of-Book Restriction (Level 1 Data):** The algorithm exclusively evaluates the Best Bid and Best Ask prices, ignoring deeper orderbook liquidity (Level 2 data). In high-frequency cryptocurrency markets, liquidity beyond the top-of-book is subject to high cancellation rates . By the time a trader walks the book to execute deeper orders, the liquidity profile has often changed. Restricting the analysis to Level 1 data represents the most conservative, highly executable liquidity available. Furthermore, because arbitrage is defined by the crossing of the spread, if the top-of-book fails to meet the profitability threshold, deeper levels mathematically cannot.

2. **Minimum Liquidity Constraint:** Opportunities were only recorded if the available size at the bottleneck price (the minimum executable shares across Outcome A and B) exceeded $10 USDC. This aligns with [14] to filter out microscopic dust orders that are economically unviable due to gas fees or slippage.

3. **Profit Threshold:** Because Polymarket currently imposes zero trading fees on NBA markets, the raw mathematical threshold of $Profit > 0.00$ was maintained, capturing the pure theoretical inefficiency of the orderbook prior to external network gas considerations.

4. **Exclusion of Post-Game Resolution Noise:** The analysis strictly filters out orderbook snapshots recorded after the physical conclusion of the sporting event. Once a game ends, the market enters a transitional state awaiting official third-party oracle resolution. During this period, active market makers withdraw their liquidity, and the resting limit orders that remain visible via the REST API often represent frozen or “ghost” liquidity that cannot be actively exploited. Excluding the Post-Game

phase ensures that the analysis measures only true, exploitable arbitrage opportunities operating under future uncertainty, preventing frozen liquidity from artificially inflating episode durations.

3.2.5 The Transition from Snapshots to Episodes

To measure the duration of arbitrages, we aggregate discrete orderbook snapshots into continuous market events, or "episodes." An arbitrage episode is defined as a contiguous block of time during which the theoretical profitability condition ($Ask_A + Ask_B < \$1.00$ or $Bid_A + Bid_B > \$1.00$) remains strictly satisfied.

As explained in Section 3.1.4, we used game-phase dependent ceilings to cap the arbitrage durations. Once every snapshot is mapped to a validated duration, consecutive arbitrage snapshots are grouped into distinct episodes. The total duration of an episode is the scalar sum of the validated forward durations of its constituent snapshots.

Crucially, the profitability of an episode is calculated using a "One-Shot" execution paradigm. In a high-frequency environment, a trader can typically only execute against a specific limit order once before consuming that liquidity. Therefore, summing the profit of multiple snapshots within a single episode would falsely assume regenerating liquidity and drastically inflate theoretical returns. Instead, the analysis script evaluates all snapshots within an episode's time boundaries and extracts the single maximum realizable profit ($Profit_{capped}$). This ensures that the reported monetary values represent the precise, executable ceiling of what an algorithmic participant could have extracted from that specific continuous market anomaly.

3.3 Results

This section presents the results of the single-market arbitrage analysis across 3,042 NBA markets on Polymarket, spanning February 4 to March 4, 2026. Player prop markets are excluded from this analysis due to insufficient baseline liquidity.

The raw detection pipeline initially identified 37 theoretical arbitrage episodes. However, 30 of these episodes (81.1%) occurred strictly in the post-game phase. To ensure the analysis focuses solely on genuine, executable market inefficiencies, these post-game episodes were excluded.

3.3.1 Validation of Post-Game Episode Exclusion

The median bid-ask spread (bps) of the limit order book was compared across the different phases of the event lifecycle (Table 3.1).

Table 3.1: Order book liquidity comparison by game phase.

Market Phase	**Median Bid-Ask Spread (bps)**
Pre-Game	392.20
In-Game	1030.90
Post-Game	7532.65

The data clearly demonstrates that post-game episodes are structural artifacts rather than actionable trading opportunities . On Polymarket, a mechanical delay exists between the physical conclusion of a game and the final Oracle resolution. During this window, rational market makers withdraw their active quotes because the underlying uncertainty has been resolved. This mass withdrawal severely hollows out the order book, causing the median bid-ask spread to explode to 7,532.65 bps, more than seven times wider than the in-game phase.

Consequently, although stale limit orders deep in the order book may appear to cross when prices move to extreme levels (0.99 or 0.01), there is no active liquidity available to execute those trades. This severe illiquidity is further confirmed by episode duration: post-game episodes persist for a median of 15.90 seconds and a mean of 147.8 seconds, compared to just a median of 3.61 seconds and a mean of 4.80 seconds for in-game episodes.

Therefore, excluding these post-game artifacts is mathematically necessary. All subsequent figures and statistics in this section reflect only the 7 valid, executable in-game arbitrage episodes identified after these microstructure filters are applied.

3.3.2 Overall Market Results

The data collection pipeline processed 75,088,497 individual orderbook snapshots across 173 unique NBA games. Despite this massive scale, the analysis identified only 7 valid arbitrage episodes across the 3,042 markets evaluated. The mean number of episodes per market was just 0.01, with a median of exactly 0.00. This heavily right-skewed distribution indicates that the vast majority of Polymarket NBA markets exhibit perfect microstructural efficiency, with pricing anomalies being exceedingly rare and concentrated within a microscopic subset of events.

Table 3.2 summarizes episodes and profitability by game phase. True arbitrage activity is exclusively an In-Game phenomenon, accounting for 100% of all valid episodes. Across the entire observation period, the aggregate capped profit, assuming a maximum capital deployment of \$100 per episode, totaled just \$210.19, with a median profit per episode of \$11.01. Assuming unlimited liquidity, the theoretical uncapped profit was \$4,418.44, illustrating the gap between theoretical market inefficiencies and practically extractable value.

Table 3.2: Summary of valid arbitrage episodes and profitability by game phase.

Game Phase	Episodes	% Time in Arb	Median Dur. (s)	Capped Profit	Uncapped Profit
Pre-Game	0	0.0000%	—	\$0.00	\$0.00
In-Game	7	0.0001%	3.614	\$210.19	\$4,418.44
Total	**7**	—	**3.614**	**\$210.19**	**\$4,418.44**

3.3.3 Speed of Arbitrage and Market Resilience

When true pricing inefficiencies do occur, they are corrected swiftly, though they persist just long enough to be recorded by discrete polling architectures. The median duration of an arbitrage episode was 3.614 seconds, with a mean of 4.80 seconds. The shortest validated episode lasted 2.276 seconds, and the longest observed episode lasted 10.996 seconds.

As Figure 3.1 illustrates, the duration distribution aligns closely with the mechanical reality of the 3-to-5 second API polling cycle. The absence of sub-second artifacts confirms that these 7 episodes represent genuine mispricings that survived at least one full network polling sweep. While human exploitation remains infeasible, these brief multi-second durations indicate that automated market makers occasionally require a few seconds to fully reprice all legs of an orderbook following a sudden live-game shock.

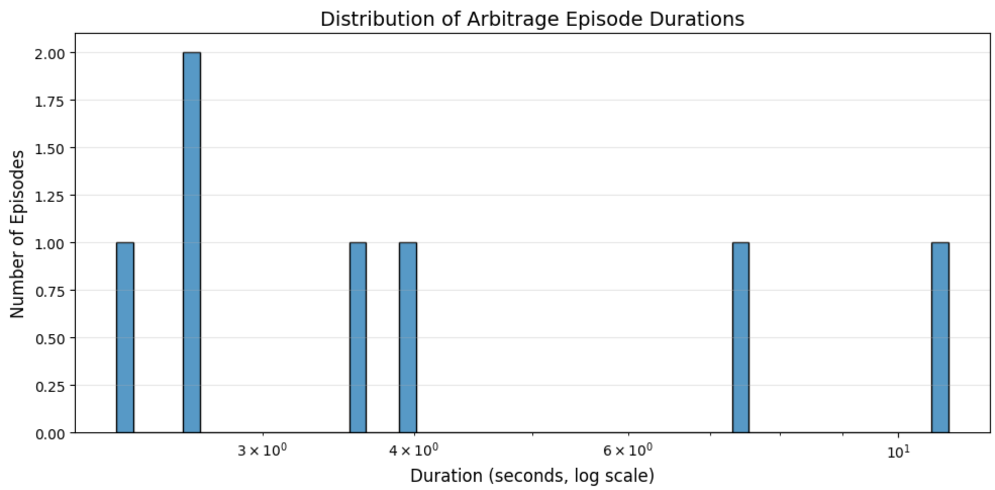

Figure 3.1: True inefficiencies generally persist for a single API polling cycle (roughly 2.5 to 11.0 seconds) before automated market makers restore efficiency.

3.3.4 Profitability by Market Type

Examining profitability by market type reveals which segments experience the deepest mispricings when live-game shocks occur. Given that the analysis identified only 7 valid episodes in total, 3 Moneyline, 3 Spread, and 1 Total, the breakdown by market type is reported as descriptive and exploratory only. The sample size is insufficient to support statistically meaningful comparisons across market types, and the differences in profitability figures should not be interpreted as generalizable structural findings. Table 3.3 displays the breakdown across Moneyline, Spread, and Total (Points) markets.

While occurrences are equally rare (3 episodes each for Moneyline and Spread), Spread markets generate disproportionately higher theoretical profit. The 3 Spread episodes produced $194.08 in capped profit and $4,368.23 in uncapped profit, compared to just $5.10 and $39.20 respectively across the 3 Moneyline episodes. This suggests that while Moneyline markets may break briefly, their liquidity remains relatively dense, keeping price deviations shallow. Conversely, Spread markets experience much deeper, albeit brief, liquidity vacuums when mispricings occur.

Table 3.3: Arbitrage episodes and profitability by market type.

Market Type	**Episodes**	**Median Duration (s)**	**Capped Profit**	**Uncapped Profit**
Moneyline	3	3.964	$5.10	$39.20
Spread	3	2.652	$194.08	$4,368.23
Total (Points)	1	2.276	$11.01	$11.01
Overall	**7**	**3.614**	**$210.19**	**$4,418.44**

Interestingly, because these 7 episodes represent the deepest, most sustained mispricings in the dataset, liquidity constraints were less prohibitive than initially theorized. In only 14.3% of episodes (1 out of 7) was the $100 capital budget bottlenecked due to insufficient orderbook depth at the mispriced level. For the remaining 85.7% of episodes, the resting liquidity was sufficient to fully absorb a retail-sized arbitrage execution.

3.3.5 Temporal Distribution Relative to Game Start

The distribution of arbitrage episodes relative to the game start relies entirely on live play. As established, 100% of the 7 valid episodes occurred In-Game. However, due to the limited arbitrages, there is no significant pattern we can conclude on the arbitrage relative to the game time.

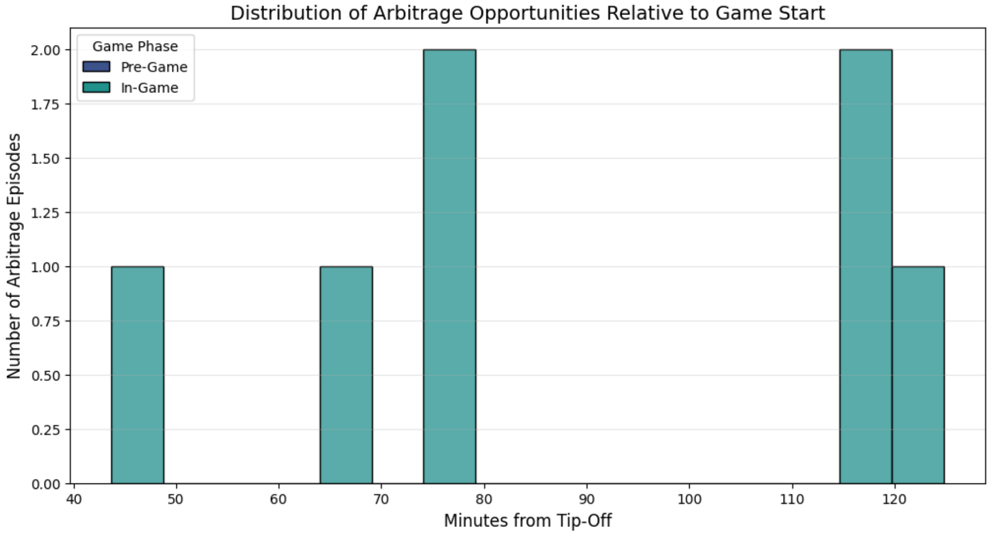

Figure 3.2: Distribution of arbitrage opportunities relative to game start.

Chapter 4

Combinatorial Arbitrage

4.1 Problem Setting

This section formalizes the combinatorial arbitrage conditions between the NBA Moneyline (outright winner) and Point Spread (winner adjusted by a handicap) markets and details the synthetic execution mechanics required on Polymarket where short selling is not available.

4.1.1 Formal Definition of Market Outcomes

Let Δ represent the final point differential between the Favorite team (Team A) and the Underdog team (Team B), defined as $\Delta = S_A - S_B$, where S_A and S_B are their respective final scores. In NBA, games cannot end in a tie, implying $\Delta \in \mathbb{Z} \setminus \{0\}$.

We define two binary derivative contracts for Team A:

- **Moneyline (ML_A):** Pays off if Team A wins the game outright. The payoff indicator is $\mathbb{I}_{\{\Delta \geq 1\}}$.

- **Spread (Sp_A):** Pays off if Team A wins by a margin strictly greater than the handicap h (where $h > 0$). The payoff indicator is $\mathbb{I}_{\{\Delta > h\}}$.

Because $\{\Delta > h\} \subset \{\Delta \geq 1\}$, the Spread outcome is a strict mathematical subset of the Moneyline outcome. By the axioms of probability, the probability of the subset cannot exceed the probability of the superset:

$$P(\Delta > h) \leq P(\Delta \geq 1)$$

Consequently, in an efficient, no-arbitrage market, the price of the Spread contract must be strictly less than or equal to the price of the Moneyline contract. A theoretical arbitrage opportunity exists if the market misprices this relationship such that the bid price of the harder outcome exceeds the ask price of the easier outcome:

$$\text{Bid}(Sp_A) > \text{Ask}(ML_A)$$

This analysis is applied to two same-period Moneyline–Spread pairs, corresponding to the two game periods for which Polymarket lists markets:

1. **Full Game:** Full Game ML_A and Full Game Sp_A
2. **First Half:** First Half ML_A and First Half Sp_A

In both cases, the logical subset relationship holds within the same period, ensuring the arbitrage condition is well-defined.

4.1.2 Execution Constraints and the Synthetic Short

If the theoretical arbitrage condition is met, the standard financial action is to sell the overpriced asset (Sp_A) and buy the underpriced asset (ML_A). However, Polymarket does not support direct short selling.

To circumvent this constraint, an algorithmic participant must construct a *synthetic short* by purchasing the complementary token. The complement to the Favorite covering the spread

(Sp_A) is the Underdog covering the spread (Sp_B):

$$Sp_B = (\Delta < h)$$

Thus, instead of selling Sp_A, the trader buys Sp_B at the ask price. The execution pair becomes:

1. Buy the Superset: $\text{Ask}(ML_A)$

2. Buy the Complement of the Subset: $\text{Ask}(Sp_B)$

The execution condition for a risk-free synthetic arbitrage is that the combined cost of acquiring both contracts must be strictly less than the minimum guaranteed payout of \$1.00:

$$\text{Ask}(ML_A) + \text{Ask}(Sp_B) < 1.00$$

4.1.3 Payoff Matrix and The "Middle" Jackpot

This combinatorial strategy does not merely guarantee a risk-free return; it creates an asymmetric upside scenario commonly referred to as the "Middle."

To illustrate, consider a matchup between the Cleveland Cavaliers (Favorite) and the Orlando Magic (Underdog) with a handicap of $h = 1.5$. The executed strategy is to buy Cavaliers ML $(\Delta \geq 1)$ and buy Magic +1.5 Spread $(\Delta \leq 1)$. Let C be the total combinatorial cost, where $C = \text{Ask}(ML_{\text{Cavs}}) + \text{Ask}(Sp_{\text{Magic}}) < 1.00$.

Table 4.1: Payoff Matrix for Synthetic Combinatorial Arbitrage (Cavs ML + Magic +1.5)

Game Scenario	$\Delta_{\textbf{Cavs}}$	**Cavs ML**	**Magic +1.5**	**Total**
Cavaliers Win Big	≥ 2	\$1.00	\$0.00	\$1.00
Cavaliers Win Narrow	$= 1$	\$1.00	\$1.00	**\$2.00**
Magic Win	≤ 0	\$0.00	\$1.00	\$1.00

As shown in Table 4.1, the portfolio guarantees a minimum payout of \$1.00 across all possible game states, ensuring a strictly positive profit of $1.00 - C$. However, if the final margin falls precisely in the gap between the Moneyline and the handicap ($\Delta = 1$), both contracts resolve to "Yes," yielding a payout of \$2.00 and producing an outsized return on the initial risk-free capital.

4.2 Methodology

This study uses the same orderbook snapshot dataset from Chapter 3. The detection of combinatorial arbitrage between two dependent prediction markets (e.g., Moneyline and Spread) introduces a significant synchronization challenge. Because Polymarket's orderbooks for distinct markets update asynchronously, Market A and Market B rarely register state changes at the exact same millisecond. A naive temporal join (such as a backward-looking "as-of" join) is insufficient, as it introduces look-ahead bias or fails to properly account for the varying collection architectures (Dense vs. Sparse regimes) utilized during the sampling period. To accurately reconstruct the combinatorial market state at any given microsecond, a Unified State Machine approach was implemented.

4.2.1 Generation of the Unified Event Timeline

Rather than aligning one market onto the timestamps of another, the methodology constructs a unified, chronological timeline of all discrete events across both markets. By performing a full outer join on the UTC timestamps of the Moneyline (ML) and Spread (S) updates, we create a master sequence of events. Let T_{ML} be the set of timestamps where the Moneyline market updated, and T_S be the set of timestamps where the Spread market updated. The unified timeline T_U is defined as:

$$T_U = T_{ML} \cup T_S$$

At any index n in this timeline ($t_n \in T_U$), an update occurs in at least one of the two markets, defining a new potential combinatorial state.

4.2.2 State Imputation via Forward-Fill

Because an update at t_n typically only involves one market, the state of the other market must be imputed to evaluate the combinatorial price, $Ask_{ML} + Ask_S$. To avoid look-ahead bias, a strict forward-fill ($ffill$) operation is applied. Mathematically, the imputed price for a market M at time t_n is the price from its most recent recorded update:

$$P_M(t_n) = P_M(t_i) \quad \text{where} \quad t_i = \max\{t \in T_M \mid t \leq t_n\}$$

This ensures that at any timestamp t_n, the algorithm is only utilizing the limit orders that were verifiably resting on the exchange at that exact moment.

4.2.3 Combinatorial Duration and Episode Measurement

Once the unified chronological states are filtered for validity, the true duration of the combinatorial arbitrage opportunity is calculated using a forward-looking difference:

$$Duration_n = t_{n+1} - t_n$$

To prevent unrecorded market halts or exchange outages from appearing as infinite arbitrage windows, this duration is strictly capped by the phase-dependent Trust Ceiling (C_{phase}) as stated in Section 3.1.4:

$$Valid\_Duration_n = \min(t_{n+1} - t_n, C_{phase})$$

For terminal states where t_{n+1} does not exist, the mathematical baseline of the collector's median gap (typically ≈ 4.0 seconds) is dynamically applied. This Unified State Machine ensures that combinatorial profit calculations represent strictly executable temporal windows, honoring the distinct mechanical reality of both data collection regimes.

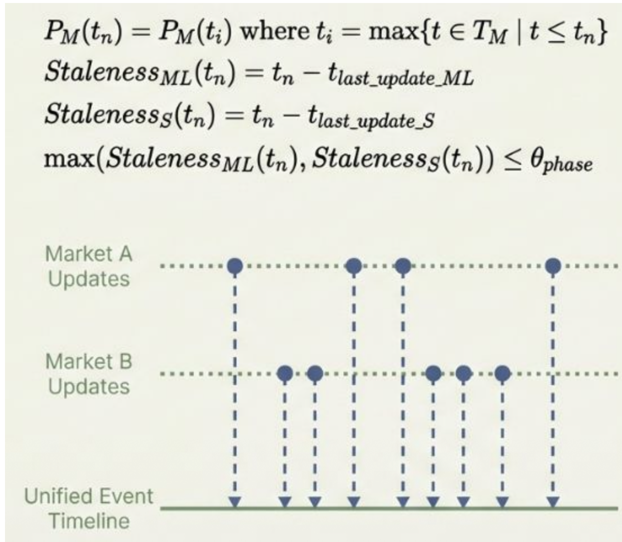

Figure 4.1: Generation of the unified event timeline

4.3 Results

4.3.1 Dataset Overview and Market Coverage

The analysis pipeline processed order book data from 170 NBA games spanning Feb 4, 2026 to March 4, 2026. After aligning the asynchronous Moneyline and Point Spread data streams into a Unified State Machine as described in Section 4.2.1, the pipeline evaluated 8.59 million discrete combinatorial market states, detecting 114,845 raw arbitrage snapshots that condensed into 523 candidate Arbitrage Episodes.

Following post-resolution filtering as stated in Section 3.3.1, we excluded 233 arbitrage episodes in post-game stage, yielding a final set of **290 active, executable Arbitrage Episodes**. The median number of arbitrage episodes per game was 2.00. Full-Game markets exhibited a median of 0.00 episodes per game, while 1H markets saw a median of 1.00, suggesting that shorter-duration markets may be subject to less efficient liquidity provisioning.

As shown in Table 4.2, market inefficiencies are overwhelmingly concentrated during live play. The Pre-Game phase, encompassing over 6.76 million evaluated states, yielded only 11 episodes, a trivial 0.0034% time-in-arbitrage rate. The In-Game phase generated 279 active episodes at a 0.18% time-in-arbitrage rate, underscoring that while rapid shifts in win probability driven by live scoring do create cross-market dislocations, the order book corrects them quickly enough that the market spends the vast majority of its time in an efficient state.

Table 4.2: Market coverage and frequency of arbitrage by game phase

Game Phase	Evaluated States	Total Arb Episodes	% of Time in Arb State	Median Duration (s)
Pre-Game	6,756,454	11	0.0034%	11.0
In-Game	1,834,885	279	0.1762%	16.0

Figure 4.2 plots the temporal distribution of episodes relative to tip-off, filtering pre-game occurrences to a window of 60 minutes before to 180 minutes after game start to remove sparse outliers that would otherwise compress the scale. The distribution is right-skewed, with the majority of arbitrage activity concentrated in the final stretch of the game. This is consistent with the hypothesis that end-game scenarios—where winning probabilities approach their terminal values and small scoring events cause large, abrupt shifts in implied odds—are the primary driver of cross-market dislocations.

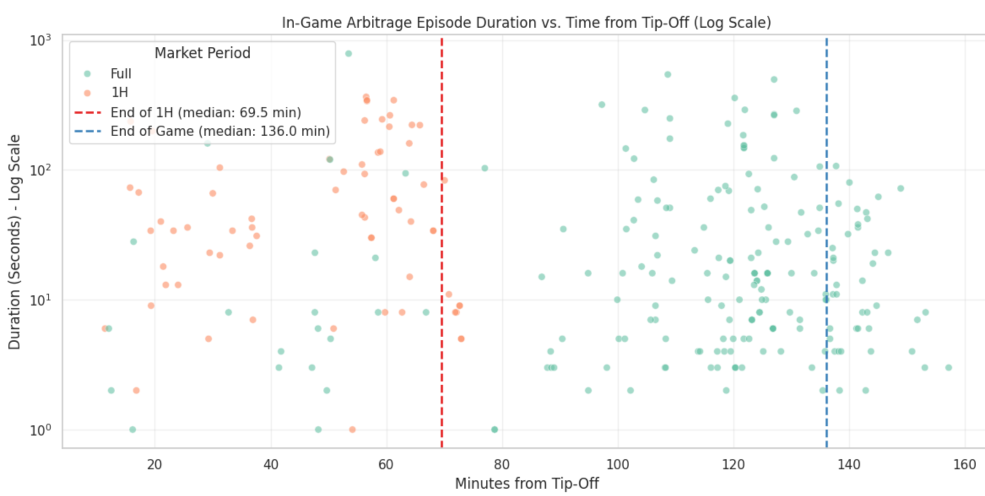

Figure 4.2: Distribution of arbitrage episodes (Pre-game episodes outside the $[-60, +180]$ minute window are excluded for visual clarity).

Figure 4.3 further illustrates the transient nature of these opportunities. The duration distribution is heavily concentrated below 10 seconds, with a long but thin right tail. The median episode lasted 16 seconds, the 95th percentile was 263.1 seconds, and the maximum observed duration was 786 seconds. Notably, 17.2% of all active episodes lasted ≤ 4.0 seconds, near the polling frequency of the data collection pipeline. These "flash" arbitrages, resolved before the next polling cycle, would be entirely undetectable at lower sampling rates and represent a lower bound on the true frequency of short-lived inefficiency.

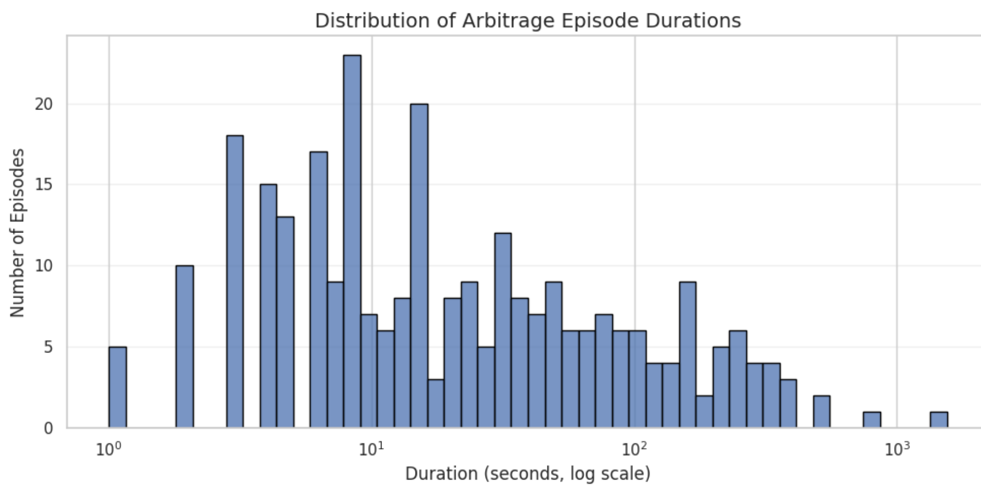

Figure 4.3: Distribution of arbitrage episode durations.

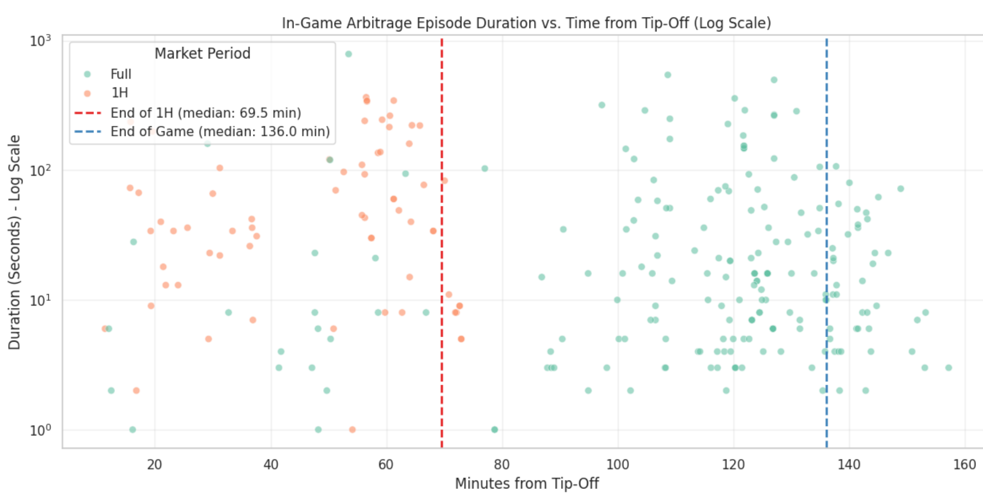

Figure 4.4: In-Game arbitrage duration vs. time from tip-off.

Figure 4.4 verifies that most arbitrage opportunities occur near the end of the game, while there is no clear pattern between the arbitrage duration and the time from tip-off.

4.3.2 The Theoretical "Middle": Absence of Jackpot Realizations

As outlined in Section 4.1, the synthetic short strategy carries an inherent asymmetric upside. Because the portfolio pairs the Moneyline Favorite with the Spread Underdog, a specific terminal outcome, where the Favorite wins by a margin falling exactly within the handicap gap, causes both contracts to resolve to "Yes," doubling the risk-free payout from \$1.00 to \$2.00 and dramatically inflating the realized yield of the episode.

To determine whether this theoretical upside materializes in practice, an ex-post resolution analysis was conducted across all 523 candidate Arbitrage Episodes using their final Polymarket CLOB settlement prices. Empirically, zero episodes resulted in a "Jackpot" realization; every observed combinatorial pair resolved to the baseline \$1.00 payout.

This null result is consistent with the joint probability structure of the event. Capturing the double payout requires the simultaneous occurrence of two independently rare conditions: a transient cross-market pricing dislocation that permits execution of the synthetic short below \$1.00, and a terminal point differential that falls precisely within the handicap gap (e.g., the Favorite winning by exactly 1 point on a -1.5 spread). Given the wide variance of final margins in professional basketball, this intersection is exceedingly unlikely over a one-month sample.

The empirical absence of any jackpot realization therefore has a direct modelling implication: algorithmic participants must underwrite cross-market arbitrage strictly on the guaranteed \$1.00 execution floor; the incremental upside from the "Middle" cannot be assigned meaningful probability weight in a rigorous expected-value calculation.

4.3.3 Profitability, Yield, and Execution Bottlenecks

We assess economic viability using *Yield (ROI)*, applying a "One-Shot" execution paradigm capped at a \$100 theoretical budget to prevent the artificial inflation of returns that would arise from assuming regenerating liquidity.

Table 4.3: Aggregate and Median Profitability by Game Phase

Phase	Episodes	Median Yield (bps)	Capped Profit	Uncapped Profit
In-Game	279	101.01	\$527.17	\$1,924.79
Pre-Game	11	309.28	\$32.42	\$107.96
Total	**290**	**101.01***	**\$559.59**	**\$2,032.75**

**Note:* The overall median yield is identical to the In-Game median because In-Game episodes comprise 96% of the total dataset, anchoring the midpoint of the distribution.

Across the 290 active episodes, the aggregate capped profit was \$559.59 with a median of \$0.40 per episode. The median yield of 101.01 basis points per execution is economically meaningful in percentage terms, yet absolute returns remain heavily constrained by market depth.

The discrepancy between the uncapped aggregate profit (\$2,032.75) and the capped figure (\$559.59) directly quantifies this liquidity bottleneck: in 76.9% of all episodes, the \$100 budget could not be fully deployed, with the average executable size bottlenecked at approximately 14.79 synthetic shares. This confirms that while exploitable combinatorial mispricings exist, they are strictly "retail-sized", the lack of order book depth prevents large-scale, risk-free capital deployment.

These findings confirm that executable combinatorial mispricings exist across NBA prediction markets, yet their exploitability is structurally bounded by liquidity. The strategy yields a consistent and statistically meaningful return in basis-point terms, but the shallow order book confines it to retail scale, precluding any meaningful capital deployment beyond trivial position sizes.

Chapter 5

Insider Trading

While Chapters 3 and 4 demonstrated the structural mispricings through arbitrage, prediction markets are equally susceptible to inefficiencies driven by asymmetric information.

5.1 Context

The presence of insider trading on Polymarket has been previously documented in geopolitical and macroeconomic contexts, such as preemptive trading ahead of unannounced military actions [5]. In these binary events, the pool of potential insiders is theoretically limited to those with direct knowledge of the event occurring, making an event-driven analytical approach tractable.

NBA markets, however, present a fundamentally different landscape for informational asymmetry. Insider information in sports betting typically manifests as pre-release knowledge of critical roster changes, undisclosed player injuries, late lineup scratches, or sudden tactical shifts [3]. Unlike a geopolitical event, this type of information does not guarantee a specific outcome; it merely shifts the true probability of that outcome before the broader market can reprice. This distinction has two important methodological consequences. First, an event-driven approach, identifying a specific news event and testing whether certain wallets traded ahead of it, is unreliable here, since the informational edge is probabilistic rather than deterministic. Second, while Polymarket does list player-specific markets that would be the most direct vehicle for this type of information, these markets suffer from insufficient trading volume to support meaningful analysis.

Instead, this study adopts a purely data-driven approach. By leveraging Polymarket's transparent on-chain history alongside its off-chain APIs, we construct behavioral profiles of pseudonymous wallets to identify trading patterns that are structurally consistent with insider participation, regardless of the specific information source. This approach is agnostic to the nature of the insider edge and relies instead on the behavioral footprint that insider trading leaves in the order book.

5.2 Data

To conduct this data-driven analysis, we use wallet-level transaction histories sourced from the Polymarket Data API [13], which enables the reconstruction of individual wallet portfolios and execution timelines. The data collection pipeline is structured into two phases, described below.

5.2.1 Target Market and Cohort Identification

The analysis targets Moneyline markets exclusively, as they carry the highest trading volume and the largest number of active wallets among all NBA market types on Polymarket as shown in Table 2.1. The first step is to identify the full population of participants in each target market. Using the `get-positions-for-a-market` API endpoint, we retrieve the wallet addresses of all traders who held a position in the market. To focus the analysis on participants whose behavior is consistent with the theoretical profile of an insiders, one who takes large, aggressive positions to maximize the value of a short-lived informational edge, we apply a volumetric filter, retaining only wallets with a total notional exposure exceeding 1,000 USDC. This forms the baseline cohort for subsequent feature construction and anomaly detection.

5.2.2 Portfolio Reconstruction and the Resolution Quirk

Once the cohort of wallets is identified, their historical trading behaviors must be evaluated to separate consistently insiders from lucky participants. This requires a comprehensive reconstruction of each user's lifetime portfolio.

A critical methodological challenge in reconstructing user portfolios stems from Polymarket's position classification architecture within its Data API. Relying exclusively on the `/closed-positions` endpoint to calculate a wallet's historical performance yields an incomplete and structurally biased dataset due to a platform-specific settlement behavior. Specifically, when a market resolves, positions do not automatically migrate to the closed ledger. Instead, they require an explicit on-chain action by the user—such as selling prior to resolution or manually redeeming winning shares post-resolution—to be officially classified as closed. Consequently, unresolved and unredeemed outcomes persist indefinitely within the active `/positions` endpoint.

This architectural quirk manifests in two distinct ways that severely distort raw performance metrics:

- **Delayed PnL Recognition:** Unredeemed winning positions remain classified as active, denoted within the API by the boolean flag `redeemable: true`, a `curPrice` resting at 1.00, and a `currentValue` reflecting the total claimable USDC. Until manually redeemed via smart contract, these functionally realized gains are entirely absent from the user's closed Profit and Loss (PnL) ledger.

- **"Zombie" Holdings and Inflated Win Rates:** Conversely, losing positions for resolved markets also persist indefinitely as active holdings (denoted by `redeemable: false`, `curPrice: 0`, and `currentValue: 0`). Known within the trading community as "zombie" holdings, these lingering losses are frequently left unclosed by high-volume traders to avoid unnecessary transaction friction. Analytically, this creates a severe survivorship bias. Because the `/closed-positions` endpoint predominantly reflects realized wins and strategically cut losses, evaluating a wallet based solely on closed data artificially inflates its historical win rate and perceived profitability.

This reconciliation produces a comprehensive and adjusted portfolio view through two concrete corrections: (1) unredeemed winning positions, identified by the flag redeemable: true in the active positions endpoint, are recognized as realized gains at their full claimable USDC value rather than being omitted from the PnL ledger entirely; and (2) zero-value zombie holdings, identified by `redeemable`: false and `curPrice`: 0, are explicitly counted as realized losses rather than being silently excluded from the closed positions ledger. Together, these two adjustments ensure that the performance metrics used to construct the PGT Oracle and to evaluate the flagged cohort reflect each wallet's true economic outcome, preventing the misclassification of standard high-volume wallets, whose winning redemptions happen to be pending, as anomalously profitable pseudo-insiders.

5.3 Methodology

This section presents a retrospective analysis of 176 NBA Moneyline markets on Polymarket from Feb 4, 2026 to March 4, 2026. The methodology consists of four components:

1. **Define the Oracle (PGT):** We heuristically define the undeniable insiders using strictly post-trade data (High Profit + High ROI + Single Direction). Purpose: To create a target for model calibration.

2. **Train the Unsupervised Learner:** We train an Isolation Forest [8] on strictly pre-resolution data (Wallet Age, Sizing Ratios), completely blind to who actually won or lost. Purpose: To find structural anomalies.

3. **Calibrate the Thresholds (Discrete Sensitivity Analysis):** To evaluate the

model's performance while avoiding optimization bias, we tested exactly four discrete anomaly cutoffs (Top 0.5%, 1.0%, 2.0%, and 5.0%) rather than conducting a continuous grid search. Consequently, multiple comparison corrections are not required. The final 1.0% threshold was selected because it structurally aligns with the empirical base rate of PGT insiders in the dataset ($\sim$1.3%), rather than by maximizing the F_2-Score. The economic validity of this selected threshold was then formally validated out-of-sample using a 1,000-iteration bootstrap of random baseline portfolios, yielding extreme statistical significance ($p < 0.001$, see Table 5.1).

4. **Out-of-Sample Validation:** We take the 2% of traders the model flagged, and we compare their actual financial returns against bootstrapped random portfolios. Purpose: The ultimate economic proof that pre-resolution structural anomalies predict post-resolution information advantages.

The Oracle serves purely as a calibration anchor for evaluating the model after training. Because it is constructed from post-resolution outcomes (realized PnL), it is strictly excluded from the training phase to prevent data leakage.

5.3.1 Defining the Oracle (Pseudo-Ground Truth)

Because Polymarket lacks regulatory labels for insider trading, this study constructs a quantitative proxy for inside information, termed the Pseudo-Ground Truth (PGT) Oracle. This Oracle is defined using strictly post-trade outcomes and is completely isolated from the pre-resolution features used to train the unsupervised Isolation Forest.

To separate genuine insider edge from idiosyncratic luck or structural market making, the PGT heuristic relies on three intersecting behavioral pillars: Profitability, Return on Investment (ROI), and Directional Conviction.

The Rationale for Heuristic Criteria

While absolute profitability is a necessary prerequisite for identifying an insider, it is an insufficient standalone filter. In zero-sum prediction markets, casual bettors frequently achieve high absolute returns simply through sheer capital size or standard variance. To distinguish genuine insider edge from capital-heavy gambling, a Return on Investment (ROI) threshold is required, ensuring the profits were generated with extreme capital efficiency.

Furthermore, high profit and efficiency can also be achieved by structural market participants. The literature identifies that a small cohort of actors consistently extracts profits from betting exchanges by acting as "large market makers" or arbitrageurs [3]. However, these professionals survive by hedging risk, resulting in a balanced portfolio. Insiders, conversely, possess deterministic private information and therefore have no incentive to hedge. To account for this, the PGT requires a strict Directional Concentration threshold.

Ultimately, an actor who achieves high absolute profit, demonstrates extreme capital efficiency (ROI), and systematically takes unhedged directional risk embodies the textbook microstructural definition of an insider, decisively filtering out both lucky gamblers and risk-free arbitrageurs and constructing a clean, high-confidence seed set that the unsupervised model can use as a calibration reference.

Calibration via Microstructural Audit

To determine the optimal mathematical thresholds for these three heuristics (Profit Percentile, ROI, and Directional Concentration), a grid search was conducted. Rather than selecting thresholds arbitrarily, the grid search evaluated each parameter combination against two strict microstructural requirements:

1. **Base Rate Alignment:** The chosen threshold must restrict the flagged population to a realistic base rate. Mathematical models presented by Whelan (2023) demonstrate that sports betting markets collapse when insider volume exceeds "a relatively low percentage" (around 2.5% to 5%) [18].

2. **Microstructural and Behavioral Validity:** The cohort isolated by the thresholds must exhibit the theoretical footprints of insider trading when compared to normal wallets. To evaluate this, we examine three specific microstructural and behavioral signatures. Importantly, these signatures are used exclusively for PGT threshold calibration and validation, they are not included as features in the Isolation Forest training. This separation is deliberate. These metrics, execution urgency and price insensitivity, are environmental in nature: they describe the market conditions surrounding a trade rather than the wallet's own structural trading behavior. Including them in the model was explored empirically, but doing so marginally reduced detection performance, likely for two reasons. First, these features introduce noise that is partly driven by prevailing market microstructure rather than by the trader's own choices; a wide spread at the time of execution, for instance, may reflect temporary illiquidity rather than an insiderss willingness to pay a premium. Second, the Isolation Forest is most effective when its feature space captures stable, wallet-level behavioral patterns that persist across markets; execution environment features vary significantly from trade to trade and dilute the signal from the more structurally informative wallet-level features. Retaining them solely as a post-hoc calibration check, where their signal is interpretively clear, therefore produces a cleaner model while still leveraging their theoretical validity to validate the PGT threshold selection. The three signatures evaluated are:

   - **Execution Urgency (`sweep_ratio`):** Insiders operate within a fleeting window of time. To capitalize on their information before it is absorbed by the broader public, speed becomes the primary objective, requiring aggressive market orders that consume outstanding limit orders [2]. This urgency is quantified by the `sweep_ratio`, defined as the trader's order size divided by the best available liquidity at the top of the book (Ask 1 or Bid 1 size). The chosen PGT cohort must exhibit statistically higher sweep ratios, confirming they prioritize immediate execution over price optimization.

- **Price Insensitivity (`relative_spread`):** Normal traders systematically avoid crossing wide bid-ask spreads because doing so destroys their expected returns. Insiders, however, possess deterministic knowledge of the true probability and are therefore price-insensitive. We capture this environment using the `relative_spread`, defined as the cost of immediate execution (the bid-ask spread normalized by the mid-price) at the exact time of the trade. The PGT cohort should therefore systematically execute trades in higher `relative_spread` environments, signaling their willingness to pay a premium to bypass adverse selection defenses.
- **Performance Persistence (Recurrence):** Bongart and Conradsson (2006) found that normal bettors who experience a large win generally fall victim to overconfidence and self-attribution bias, subsequently increasing their risk and deteriorating their returns [3]. Therefore, if a wallet recurrently meets the strict PGT criteria across multiple independent markets, it defies this psychological trap. High wallet recurrence serves as empirical proof of persistent skill rather than a lucky streak.

Optimal PGT Threshold Selection

Evaluating the grid search across these dimensions revealed that setting the thresholds at the **95th Percentile of Profit**, **0.60 Minimum ROI**, and **0.90 Directional Concentration** produced the most structurally sound oracle.

This specific combination is empirically robust for three reasons:

- **Strict Base Rate Adherence:** The chosen threshold identifies approximately 1.30% of the market population as pseudo-insiders. This highly conservative base rate perfectly aligns with the theoretical efficiency bounds proposed by Whelan (2023), which dictate that prediction markets cease to function when insider prevalence exceeds a ceiling of 2.5% to 5% [18]. Because the NBA markets in this dataset are active and liquid, capturing an empirical insider rate of 1.0%, safely below the collapse threshold, confirms that the Oracle is modeling a realistic, sustainable level of insider trading.

- **Execution Urgency:** This threshold created the widest behavioral gap in execution urgency across the grid search (Δ`Sweep_Ratio` $= +0.0889$). Mann-Whiteney U tests confirm that the PGT insiders, compared to the normal wallets, exhibit a statistically significant increase in sweep ratios ($p < 0.01$), indicating they actively consume liquidity across multiple price levels.

- **Price Insensitivity:** Mann-Whiteney U tests also confirm that the PGT insiders exhibit a statistically significant increase in relative spread ($p < 0.01$), indicating PGT trades systematically occur during periods of elevated relative spread.

- **Massive Performance Persistence:** Under these parameters, 46.9% of the identified insider wallets appeared multiple times within the dataset. The binomial probability of a random noise trader hitting this exact, extreme threshold in multiple independent markets is practically zero, firmly rejecting the null hypothesis of idiosyncratic luck.

With the Oracle mathematically calibrated and validated against microstructural theory, it serves as the definitive target for evaluating the unsupervised Isolation Forest.

5.3.2 Training the Unsupervised Model

With no reliable regulatory labels available for the full prediction market population, we employ an Isolation Forest. This unsupervised anomaly detection algorithm has been demonstrated to be highly effective in financial settings at scoring traders based on how structurally anomalous their trade behavior is relative to the broader population [4].

Rather than fitting a separate model for each event, the Isolation Forest is trained globally across a concatenated dataset of 176 Polymarket NBA Moneyline markets. Each observation represents a unique wallet's activity within a single target market, combined with its historical trading record prior to that event.

This design is motivated by two complementary considerations. First, Liu et al. [8] demonstrate that Isolation Forests achieve optimal detection performance using random sub-samples of 256 instances per tree. Filtering to wallets with total notional exposure exceeding 1,000 USDC restricts the population to approximately 200 wallets per market, below this threshold. Training locally on a single market would therefore eliminate ensemble diversity and cause the algorithm to overfit to localized variance. Concatenating 176 markets yields a global dataset of over 35,000 observations, allowing the model to draw sufficiently diverse, optimal sub-samples for each tree.

Second, global training mitigates swamping, a failure mode in which a model trained locally on an anomaly-free market is forced to partition normal data, arbitrarily flagging standard traders simply because they occupy the tails of a localized distribution [8]. By drawing sub-samples from a pooled dataset spanning the full platform, the model establishes a stable baseline of normal trading behavior that is not distorted by the idiosyncratic composition of any single market.

Feature Selection and Engineering

To prevent target leakage, the feature matrix (X) is strictly constrained to pre-resolution information, ensuring the model has no access to post-resolution outcomes (e.g., realized PnL). Furthermore, all features are designed to be scale-invariant. By describing a trader's behavior relative to the immediate market environment, the model is able to generalize across markets with vastly different levels of absolute liquidity and volume.

Following the feature selection principles outlined in [4], we focus exclusively on individual trading and wallet maturity behaviors. The final feature set was refined and validated using SHapley Additive exPlanations (SHAP) [10] to ensure the model isolated genuine behavioral outliers. The finalized feature space consists of:

- `wallet_age_days`: The time elapsed between the wallet's creation on the blockchain and its first trade in the target market, capturing the divergence between established professionals and ephemeral “burner” accounts.

- `unique_markets`: The total number of distinct Polymarket events the wallet has traded in historically, serving as a proxy for platform experience.

- `historical_roi`: The wallet's historical profitability prior to the target event, calculated as historical PnL divided by historical traded notional.

- `market_notional_ratio`: The user's total traded notional in the specific target market divided by the total aggregate volume of that market, capturing outsized market participation.

- `position_anomaly_ratio`: The user's maximum position size in the target market divided by their own historical average position size, identifying severe, uncharacteristic deviations in bet sizing.

- `avg_buy_price`: The average execution price obtained by the trader, implicitly capturing the probability tier at which they are willing to assume risk.

Data Transformation

Empirical analyses of financial market microstructure have long established that behavioral trading metrics, particularly trade and position sizes, exhibit massive right-skewness and heavy power-law tails rather than normal distributions [6]. To prevent extreme magnitude outliers from dominating the Isolation Forest's distance metrics, a log-transformation is applied to the heavily skewed feature, `position_anomaly_ratio`:

$$X_{scaled} = \ln(1 + X_{raw}) \tag{5.1}$$

By compressing the extreme right tail, this transformation ensures the ensemble trees isolate wallets based on the confluence of anomalous behaviors rather than purely on absolute dollar magnitude.

The ensemble is configured with `n_estimators` = 2000 to produce a smooth, continuous distribution of anomaly scores.

5.3.3 Threshold Calibration (Anomaly Score Cutoff)

Having established a structurally validated PGT Oracle comprising exactly 1.0% of the market population (462 pseudo-insiders), the final step requires calibrating the classification cutoff for the unsupervised Isolation Forest. The model scores every wallet on a continuous spectrum of structural anomaly; the threshold determines which percentile of scores constitutes a flagged anomaly.

One natural approach to setting a classification threshold is to maximize machine learning metrics like the F_2-Score or Precision-Recall Area Under the Curve (PR-AUC). In practice, however, this often leads to "threshold collapse" in financial settings: the optimizer selects an overly liberal cutoff that inflates recall at the extreme cost of precision and economic viability.

As demonstrated in the sensitivity analysis (Table 5.1), optimizing strictly for the highest F_2-Score (0.1297) pushes the model toward a 5.0% cutoff. This liberal threshold flags $1,777$ traders, nearly quadruple the empirical prevalence of insiders identified by the PGT Oracle (462). More problematically, expanding the cutoff to 5.0% entirely destroys the economic edge: the flagged cohort generates a negative expected out-of-sample return ($-\$48.01$) but the empirical p-value rises to 0.486, completely failing to achieve statistical significance. This empirically proves that standard classification metrics fail to capture financial edge without structural constraints.

To prevent threshold collapse, we apply the principle of base rate alignment: the model's flagging rate should strictly mirror the empirical prevalence of the Oracle. A top **1.0% cutoff** was selected on this basis. This highly restrictive threshold flags exactly 356 traders, keeping the cohort operationally meaningful and mathematically synchronized with the 1.0% true base rate of insiders. Crucially, the 1.0% cutoff maximizes the out-of-sample economic edge, generating an expected alpha of $\$6,638.62$ per market with absolute statistical certainty

($p < 0.001$), while still capturing 11.7% of all market profits.

Model Cutoff Percentile	**Flagged Anomalies**	**True Positives**	**Precision**	F_2**-Score**	**Profit Capture Ratio (%)**	**Empirical** P**-Value**
Top 0.5%	178	18	0.1011	0.0444	7.54%	< 0.001
Top 1.0%	**356**	**33**	**0.0927**	**0.0749**	**11.70%**	$<$ **0.001**
Top 2.0%	711	51	0.0717	0.0996	18.97%	0.002
Top 5.0%	1,777	94	0.0529	0.1297	31.02%	0.486

Table 5.1: Sensitivity Analysis of Anomaly Cutoff Percentiles against the Validated PGT Oracle ($N = 462$). The 1.0% cutoff is selected to perfectly mirror the empirical base rate, ensuring extreme out-of-sample statistical significance ($p < 0.001$) and protecting against threshold collapse.

5.3.4 Validation Results of the Unsupervised Pipeline

Because the Isolation Forest is an unsupervised algorithm, it cannot be evaluated using standard held-out test sets. Instead, the final model (calibrated to the 1.0% anomaly threshold) is validated across three dimensions: algorithmic sanity and out-of-sample economic performance.

- **Algorithmic Sanity Checks:** To evaluate if the Isolation Forest is able to detect insider trading behavior, we evaluate the raw, continuous anomaly scores before any classification threshold is applied. A Welch's t-test confirms that the model assigns statistically higher anomaly scores to PGT Oracle wallets than to the general wallet population ($p < 0.001$). Furthermore, a binomial exact test confirms that the strict 1.0% cutoff captures a statistically significant share of the Oracle ($N = 33$), far exceeding the expected capture rate under a random distribution ($p < 0.001$).

- **Market Dominance and Expected Alpha:** As established during threshold calibration, the flagged cohort exhibits massive market dominance, capturing 11.7% of all aggregate profits ($88.26 million) despite representing only 1.0% of the population. To rigorously validate that this extraction represents a genuine insider edge rather than high-volume luck, we conducted an out-of-sample bootstrapping simulation. We generated $1,000$ random portfolios of equal size ($N = 356$) drawn from the full trader population. The actual mean PnL of the flagged anomalies (**$6,638.62**) completely eclipsed the expected baseline PnL of the bootstrapped null distribution (-$59.61), producing an empirical p-value of < 0.001.

Validation Metric	**Result**	**Statistical Significance**
Algorithmic Sanity (Raw Scores)		
PGT vs. Normal Anomaly Score (T-Test)	Statistically Higher	$p < 0.001$
Oracle Capture Rate (Binomial Test)	33 / 462	$p < 0.001$
Economic Performance (1.0% Cutoff)		
Profit Capture Ratio	11.70%	–
Flagged Cohort Mean PnL	$6,638.62	Empirical $p < 0.001$
Bootstrapped Baseline Expected Mean PnL	-$59.61	–
Flagged Cohort Median PnL	$150.00	–
Bootstrapped Baseline Expected Median PnL	$85.86	–

Table 5.2: Validation summary of the unsupervised Isolation Forest. The model demonstrates both structural accuracy in isolating the Oracle and statistically significant out-of-sample economic extraction.

A key limitation of the model's performance, however, is its recall against the PGT Oracle. At the 1.0% threshold, the Isolation Forest captures 33 of 462 pseudo-insiders identified by the Oracle, corresponding to a recall of 7.1%. This low recall is a direct consequence of the threshold calibration strategy: by anchoring the flagging rate to the empirical base rate of insider trading to prevent threshold collapse, the model necessarily prioritizes precision and economic significance over coverage. The practical implication is that the Isolation Forest should not be interpreted as a comprehensive detector of all insider trading activity on the platform. Rather, it identifies one specific behavioral archetype of insider participation, wallets that are structurally anomalous in terms of platform footprint and position scaling, and does so with statistically significant economic validity. The majority of the Oracle's pseudo-insiders who are not flagged by the model likely exhibit insider behavior through different structural signatures that the current feature set does not fully capture, representing a direction for future work.

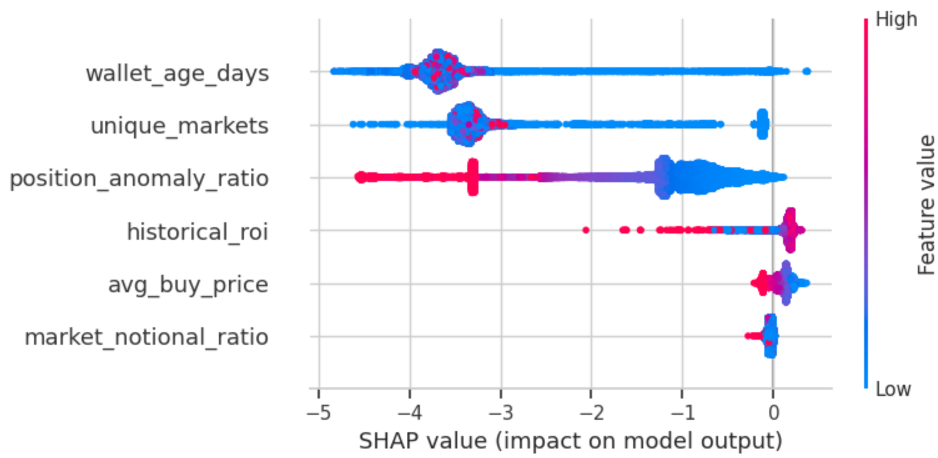

Figure 5.1: SHAP summary.

To understand the structural drivers of the Isolation Forest's anomaly classification, we analyze the global SHAP (SHapley Additive exPlanations) summary (Figure 5.1). The model is primarily driven by platform footprint features, `wallet_age_days` and `unique_markets`, followed closely by extreme bet scaling `position_anomaly_ratio`. Notably, execution mechanics such as `avg_buy_price` and `market_notional_ratio` had minimal impact on the model's unsupervised flagging decisions, residing near the bottom of the SHAP ranking with values tightly clustered around zero.

5.4 Post-Hoc Behavioral Analysis of Flagged Wallets

The Isolation Forest pipeline identifies structural anomalies in behavioral features, but its output is a binary flag rather than a behavioral characterization. This section investigates what the flagged wallets actually represent, why the model missed the majority of PGT pseudo-insiders, and what distinguishes the minority it successfully detected. The audit proceeds in four stages: feature distribution analysis, an explanation of the masking phenomenon, a behavioral comparison of the 33 flagged PGT insiders against the 429 the model missed, and a methodological conclusion for future work.

Throughout this section, four cohorts are referenced consistently:

- **Baseline**: The unflagged general wallet population.
- **Profitable Flags**: Wallets flagged by the Isolation Forest whose realized PnL is positive.
- **Unprofitable Flags**: Wallets flagged by the Isolation Forest whose realized PnL is zero or negative.
- **PGT Insiders**: Wallets independently identified as pseudo-insiders by the Oracle (Section 5.3.1), regardless of whether the model flagged them.

5.4.1 Feature Distribution Analysis: The Model Found the Extremes

To characterize the behavioral differences across all four cohorts, Kernel Density Estimates were computed for each training feature (Figure 5.2).

Before interpreting the distributions, Figure 5.1 establishes which features actually drove the model's flagging decisions. The Isolation Forest weighted `wallet_age_days` and `unique_markets` most heavily, followed by `position_anomaly_ratio` and `historical_roi`. Execution price `avg_buy_price` and `market_notional_ratio` had relatively low influence.

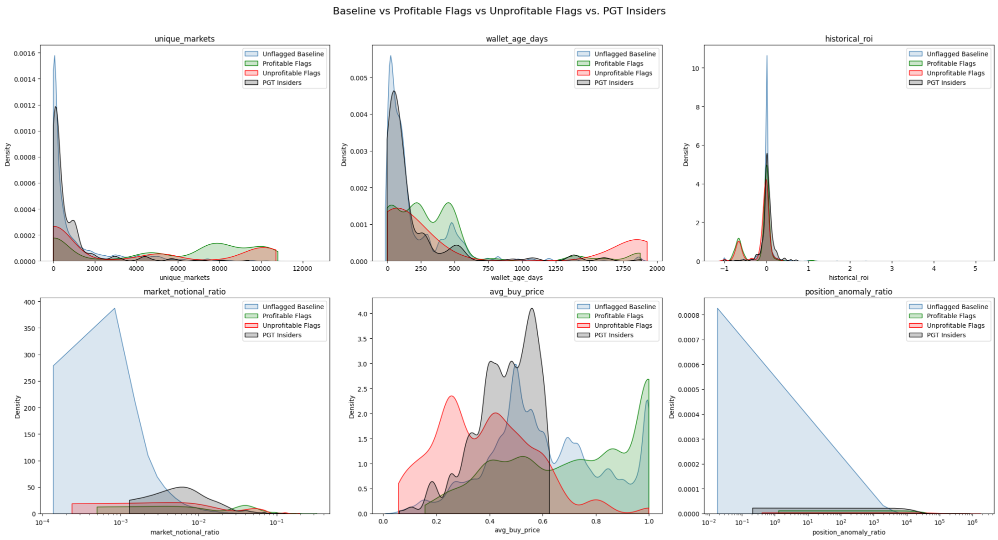

Figure 5.2: Feature distribution comparison: baseline, profitable flags, unprofitable flags, and PGT insiders.

This distinction is critical: the model flagged wallets for who they structurally are, not for how or when they traded.

Platform Footprint (`unique_markets`, `wallet_age_days`). The flagged cohorts separate sharply from the baseline on both footprint features. They are heavily concentrated in an extreme high-frequency tier (e.g., 8,000 to 11,000 unique markets) with older wallet ages. While we lack execution logs to definitively classify these wallets as automated bots, this extreme volume strongly suggests algorithmic or API-driven execution.

Crucially, the PGT insiders do not share this extreme footprint. Instead, their distributions for (`unique_markets` and `wallet_age_days`) almost perfectly overlap with the unflagged baseline, clustering near zero markets and newer accounts. This aligns with intuition: genuine insiders trade selectively, deploying capital only when they possess a specific informational edge, often using fresh "burner" accounts.

However, this overlap introduces a critical analytical tension. If the PGT insiders look structurally identical to the standard baseline on the model's most heavily weighted features, how and why did the model manage to successfully flag 33 of them?

Market-Specific Behavior (`position_anomaly_ratio`, `market_notional_ratio`,

avg_buy_price). The answer begins to emerge in the market-specific execution features. While PGT insiders blend into the baseline's platform footprint, they diverge significantly in their capital deployment. For both market_notional_ratio and position_anomaly_ratio, PGT insiders and flagged group shift substantially rightward from the near-zero baseline, demonstrating highly concentrated capital deployment.

Furthermore, their execution pricing is highly specific. The PGT insiders cluster tightly between 0.3 and 0.6 on avg_buy_price, the mid-probability range associated with true pre-game directional conviction. In contrast, the profitable flagged cohorts are heavily concentrated near 1.0, indicating they are primarily scraping late-game or post-game certainties and behave unlike insiders.

Historical ROI (historical_roi). All cohorts exhibit a primary mode near zero. This confirms that even the highly active flagged wallets do not hold a persistent, universal trading edge. Their outsized returns, much like those of the PGT insiders, are event-specific rather than universal.

The feature distributions reveal a straightforward reality: the structurally "weirdest" wallets on Polymarket, the ones that trigger the vast majority of the model's anomaly flags—exhibit high-frequency, bot-like characteristics. They are completely distinct from the focused, selective behavior expected of human insiders. The Isolation Forest detected exactly the extreme global outliers it was designed to find. This raises the central question for the following sections: given that the true insiders mimic the baseline's footprint, what masking dynamics hid the majority of them, and what extreme behaviors allowed exactly 33 to break through?

5.4.2 The Masking Phenomenon in Unsupervised Detection

This dynamic explains why the Isolation Forest flagged only 33 of the 462 total PGT insiders identified by the Oracle. This discrepancy exposes a known theoretical limitation of the standard Isolation Forest: while it excels at identifying global outliers, it struggles to detect local anomalies [9].

The algorithmic bots in this dataset act as extreme global outliers. Their massive platform footprints isolate them in sparse feature space, requiring very few algorithmic partitions to detect. In contrast, the genuine PGT insiders operate as local anomalies. On macro-level features like `unique_markets` and `wallet_age_days`, they reside deep within the dense cluster of standard baseline traders. Their anomalous behavior, extreme capital concentration, is only structurally apparent within that localized context.

Because the Isolation Forest relies on random partitioning, it easily isolates the global algorithmic outliers but requires deep, extended path lengths to partition the dense baseline cluster where the insiders reside. Consequently, the true insiders are assigned longer path lengths, which the model interprets as "normal" behavior. In effect, the extreme structural presence of the algorithmic bots "masks" the localized insiders.

5.4.3 Breaking the Mask: The Anatomy of the Flagged Insider

To understand how the 33 flagged PGT insiders overcame the masking effect, their feature distributions were compared directly against the 429 PGT insiders the model did not flag (unflagged PGT insiders) (Figure 5.3).

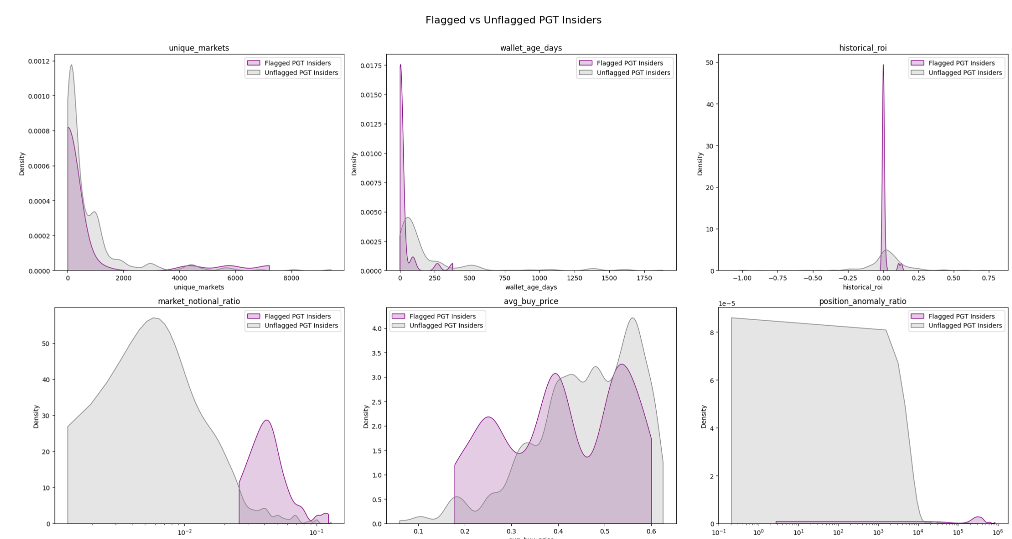

Figure 5.3: Feature distribution comparison: flagged vs. unflagged PGT Insiders.

An assumption might be that these 33 wallets overcame the masking effect by exhibiting even more extreme bot-like behavior to be flagged by the model. However, the data rejects this, showing the 33 flagged PGT insiders have significantly *fewer* unique markets and *shorter* wallet ages than the unflagged PGT group.

Mechanically, this proves that the 33 wallets were not flagged because of their platform footprint (`unique_markets`, `wallet_age_days`. Because the Isolation Forest computes a composite anomaly score across all features simultaneously, a sufficiently extreme value on any subset of features can push a wallet over the detection threshold even when other features are unremarkable [8]. For these 33 wallets, `market_notional_ratio` and `position_anomaly_ratio` were extreme enough to accomplish this on their own. Table 5.3 presents the key comparisons.

Table 5.3: Median Feature Values Comparison

Feature	**33 Flagged PGT**	**429 Unflagged PGT**	**General Flagged**
`position_anomaly_ratio`	336,538.5	2.9	16.0
`market_notional_ratio`	0.044	0.010	0.038
`avg_buy_price` range	0.2 – 0.6	0.1 – 0.6	0.4 – 0.9

The position scaling contrast is particularly striking. The 33 flagged PGT insiders show a median `position_anomaly_ratio` of 336,538, more than 100,000 times larger than the 429 unflagged PGT insiders (median: 2.9) and more than 20,000 times larger than the general flagged population (median: 16.0). These wallets were not simply large traders; they deployed capital at a scale so disproportionate to their own historical baseline that the anomaly signal broke through the masking effect entirely on its own.

Beyond the scaling dimensions, the 33 wallets exhibit three additional characteristics that collectively paint a coherent behavioral profile distinct from both the bot-dominated flagged population and the broader Oracle cohort.

First, their `avg_buy_price` is concentrated between 0.2 and 0.6, the mid-probability range associated with pre-game directional conviction, compared to a range below 0.1 to above

0.6 for the 429 unflagged PGT insiders. They are not chasing late-game certainty; they are taking targeted directional positions at meaningful probability levels before the game begins. This aligns with what Shin (1992) states that true insiders exploit higher-probability pricing rather than gambling on low-probability longshots [15].

Second, their platform participation is focused rather than broad. Fewer unique markets and younger wallet ages indicate that these are not seasoned platform veterans accumulating positions across thousands of events. They are concentrated participants whose anomalous behavior is localized to specific high-conviction opportunities.

Third, their entry timing is tightly aligned with the game-day information cycle. All 33 wallets placed their initial trades exclusively in the pre-game phase, clustering between 0 and 550 minutes before tip-off. The 429 unflagged PGT insiders, by contrast, exhibit a wide right-skewed distribution extending up to 4,000 minutes before the game, consistent with early speculative positioning rather than acting on late-breaking information. The 33 flagged insiders are not taking positions days in advance; their capital deployment aligns precisely with the window when undisclosed injury reports and roster changes become available.

Taken together, the 33 flagged PGT insiders are distinguished from the 429 unflagged PGT insiders on four behavioral dimensions: greater market dominance, more aggressive position scaling, narrower mid-probability execution prices, and tighter pre-game entry timing. Importantly, this behavioral extremity translates directly into outsized economic outcomes. The flagged cohort achieved an average realized profit percentile of 0.99 within their respective markets, compared to 0.97 for the unflagged PGT insiders.

Beyond individual performance, this localized extremity results in massive aggregate economic dominance. Table 5.4 illustrates the disproportionate financial impact of these 33 wallets across multiple structural contexts.

This concentration proves that while high-frequency bots may trigger the model's structural thresholds through sheer volume, the true insiders dominate the market economically. Representing less than 15% of the profitable flagged population, these 33 wallets extracted

Table 5.4: Economic concentration of the 33 flagged PGT insiders across different cohorts.

Comparative Cohort	Wallet Share (%)	Profit Share (%)
Within All PGT Insiders ($N = 462$)	7.1%	23.2%
Within Profitable Flags ($N = 224$)	14.7%	78.6%
Global Aggregate Market ($N = 35,488$)	0.09%	9.2%

nearly 80% of its profits. Ultimately, the unsupervised architecture successfully bypassed the low-impact noise traders to isolate the absolute most financially severe actors on the platform.

Critically, none of these distinguishing characteristics correspond to the platform footprint features the model weighted most heavily globally. Their detection therefore reflects high-conviction, insider-like concentration rather than algorithmic extremity. Ultimately, while the model's overall detection rate was masked by high-frequency participants, the market-specific aggression of these 33 wallets was severe enough to overcome that distributional noise. This demonstrates that under the current unsupervised architecture, the Isolation Forest acts as a strict filter for severity, successfully isolating the most economically impactful insiders on the platform.

5.4.4 Methodological Conclusion

The post-hoc analysis yields a clear methodological lesson for future research applying unsupervised anomaly detection to decentralized prediction markets. The standard Isolation Forest, applied directly to a heterogeneous participant population, detects algorithmic bots rather than human insiders. This is not a failure of the algorithm; it is a predictable consequence of applying a global outlier detector to a dataset where the most structurally extreme observations are not the ones of interest. The bots mask the insiders by dominating the anomaly score distribution.

The practical implication is that a pre-filtering step is required before the Isolation Forest is applied. For example, a threshold-based filter on `unique_markets` and `wallet_age_days` would remove the algorithmic participant tier from the feature space prior to model training. With the bot population removed, the Isolation Forest's partitioning would operate within a more homogeneous baseline of human traders, allowing the market-specific behavioral signals that distinguish genuine insiders, concentrated position sizing, targeted market dominance, and pre-game entry timing, to emerge as the primary drivers of anomaly scores rather than being suppressed by global distributional extremes.

This two-stage pipeline, bot filter followed by Isolation Forest, represents the most direct architectural improvement to the methodology developed in this thesis and is the recommended starting point for future work on insider trading detection in decentralized prediction markets.

Chapter 6

Limitations and Conclusion

6.1 Limitations

6.1.1 Sample Period and Generalizability

All empirical analyses in this thesis are drawn from a single calendar month, February 4 to March 4, 2026. This period corresponds to the late regular season, a phase of the NBA calendar characterized by elevated stakes, heightened lineup management, and potentially distinct liquidity conditions relative to earlier months. It is therefore not possible to determine whether the observed levels of microstructural efficiency, the prevalence of combinatorial mispricings, or the 1.0% insider trading base rate generalize to other phases of the season. In particular, the early season, when markets are new, liquidity is thinner, and participant composition is less settled, may exhibit materially different dynamics. Future work should replicate this analysis across a full season to assess robustness.

6.1.2 Temporal Resolution and Latency

The primary limitation of the single-market arbitrage analysis is the reliance on the REST API for order book snapshots. While intra-market updates were batched to ensure synchronization, the mechanical polling loop resulted in a median interval of 3.6 to 5.5 seconds between requests, meaning that any arbitrage opportunity resolving within this window is invisible to the pipeline. The frequency and cumulative duration of sub-polling-frequency episodes cannot be measured from this dataset, meaning the results reported in Chapter 3 understate both the count and the brevity of microstructural inefficiencies. Future work should implement a continuous WebSocket feed to achieve millisecond-level resolution and enable precise duration measurement.

6.1.3 Model Scope and Real-Time Adaptation

The current anomaly detection framework is inherently retrospective and global in scope. Due to time constraints in this thesis, the Isolation Forest was trained exclusively on a historically concatenated dataset of all games. While this global approach successfully establishes a broad baseline of normal trading behavior, future work should explore applying the Isolation Forest to individual games independently and including not only the wallets with notional value over 1,000 USDC. Fitting models at the single-market level could potentially allow the algorithm to capture distinct microstructural nuances and baseline liquidity conditions unique to specific matchups.

Furthermore, because the model is trained on this static, historical dataset, it is well-suited for post-hoc analysis but cannot dynamically adjust to live market conditions without full periodic retraining. Deploying this framework in a real-time environment would require a fundamental architectural shift toward streaming anomaly detection, for instance via a continuous WebSocket feed. Future research should explore transitioning to an Online Isolation Forest [7], which incrementally updates its isolation trees as new observations arrive rather than requiring batch retraining, making it naturally compatible with live order flow data.

A second and equally promising direction is the generalization of this methodology to other Polymarket categories, particularly political markets. Relative to NBA markets, political markets offer a structurally cleaner setting for insider detection for two reasons. First, several political markets on Polymarket have documented cases of anomalous pre-event trading with verified informational context [1], providing a rare opportunity to benchmark the model against events where the existence of an informational edge is externally confirmed rather than inferred. Second, political outcomes are typically binary and decisive, an event either occurs or it does not, which removes the probabilistic ambiguity inherent in sports markets, where injury news and lineup changes shift win probabilities gradually rather than discretely. Applying the same detection pipeline to political markets would therefore serve both as a robustness check on the methodology and as a substantively richer environment in which to study the behavioral signatures of insider trading.

6.2 Conclusion

This thesis provided a systematic empirical investigation into the market microstructure and informational efficiency of Polymarket's NBA prediction markets. By integrating high-frequency order book snapshots with verified on-chain transaction data, the research successfully bridged the gap between decentralized finance (DeFi) architecture and traditional market efficiency theories.

- **Microstructural Efficiency:** The analysis of single-market arbitrage reveals an environment of extreme resilience; median mispricing windows resolve in 3.614 seconds. This correction rate suggests that these markets are highly efficient at the Level-1 order book level.

- **The Liquidity Ceiling:** While combinatorial arbitrage opportunities (spanning Moneyline and Spread markets) persist for longer durations, their economic exploitability is strictly bottlenecked by a lack of order book depth. These anomalies are primarily "retail-sized," as the sheer lack of liquidity prevents large-scale, risk-free capital deployment.

- **Detection of Economically Dominant Insiders and Policy Implications:** The unsupervised anomaly detection pipeline successfully isolated a concentrated cohort of 33 highly aggressive insider traders. While these insiders represent a fraction of the participant base, they exhibit massive economic severity, capturing 11.7% of aggregate market profits and accounting for 78.6% of the total profits made by the profitable flagged population. This extreme profit concentration carries significant practical implications. For decentralized platforms like Polymarket, it demonstrates that basic heuristics are insufficient for market integrity; platforms must adopt real-time algorithmic surveillance, such as the two-stage bot-filtering and anomaly detection architecture proposed herein, to monitor for specific signatures of capital concentration. For regulators, these findings highlight a paradigm shift. While insider trading is demonstrably present and economically impactful, the inherent transparency of on-chain limit order books enables deterministic surveillance that is impossible in traditional, opaque sportsbooks.

Appendices

Appendix A

Sample Orderbook Data

Table A.1 presents a representative sample of the Level-1 orderbook snapshot data collected via the Polymarket REST API. This data illustrates the mirrored liquidity structure of binary markets and the high-frequency nature of the state updates analyzed in Chapters 3 and 4.

ts_utc	Outcome	Bid Px	Ask Px	Bid Size	Ask Size	RTT (ms)
2026-02-15 15:05:17.403	Suns	0.05	0.95	500.00	500.00	193.27
2026-02-15 15:05:17.403	Magic	0.05	0.95	500.00	500.00	193.27
2026-02-15 15:05:57.321	Suns	0.05	0.95	500.00	500.00	187.93
2026-02-15 15:05:57.321	Magic	0.05	0.95	500.00	500.00	187.93
2026-02-15 15:10:35.707	Suns	0.05	0.95	500.00	500.00	165.16
2026-02-15 15:10:35.707	Magic	0.05	0.95	500.00	500.00	165.16
2026-02-15 15:10:36.355	Magic	0.05	0.95	500.00	500.00	162.22
2026-02-15 15:10:36.355	Suns	0.05	0.95	500.00	500.00	162.22
2026-02-15 15:11:03.403	Magic	0.06	0.94	250.00	250.00	161.83
2026-02-15 15:11:03.403	Suns	0.06	0.94	250.00	250.00	161.83
2026-02-15 15:14:04.280	Suns	0.06	0.94	250.00	250.00	188.59
2026-02-15 15:14:04.280	Magic	0.06	0.94	250.00	250.00	188.59
2026-02-15 15:14:04.286	Suns	0.06	0.94	255.24	255.24	170.45
2026-02-15 15:14:04.286	Magic	0.06	0.94	255.24	255.24	170.45
2026-02-15 15:34:57.000	Suns	0.06	0.94	255.24	255.24	183.67
2026-02-15 15:34:57.000	Magic	0.06	0.94	255.24	255.24	183.67
2026-02-15 15:34:58.000	Magic	0.06	0.94	255.24	255.24	230.17
2026-02-15 15:34:58.000	Suns	0.06	0.94	255.24	255.24	230.17
2026-02-15 15:36:25.899	Magic	0.40	0.94	55.00	255.24	201.87
2026-02-15 15:36:25.899	Suns	0.06	0.60	255.24	55.00	201.87

Table A.1: Raw Level-1 limit order book snapshots for a binary NBA market. Displayed RTT values are rounded to two decimal places, and the UTC timezone offset is omitted for formatting constraints.

Column Definitions:

`ts_utc` **(Timestamp):** The exact Coordinated Universal Time (UTC) when the orderbook snapshot was recorded, captured down to the microsecond.

`outcome`**:** The specific entity or side of the binary market being traded (e.g., "Suns" or

"Magic").

`best_bid_price`: The highest price (in USDC) that a buyer is currently willing to pay for one share of this outcome.

`best_ask_price`: The lowest price (in USDC) that a seller is currently willing to accept for one share of this outcome.

`best_bid_size`: The total number of shares available to be sold to the market at the current best bid price (representing buy-side liquidity).

`best_ask_size`: The total number of shares available to be purchased from the market at the current best ask price (representing sell-side liquidity).

`rtt_ms` **(Round-Trip Time):** The network latency measured in milliseconds. It tracks exactly how long it took for the data request to travel to the Polymarket servers and return the snapshot, ensuring data synchronization.

Appendix B

The Isolation Forest Algorithm

B.1 Conceptual Overview

The Isolation Forest algorithm, developed by Liu, Ting, and Zhou (2008) [8], is an unsupervised, non-parametric method for anomaly detection. Unlike traditional distance-based or density-based anomaly detection models that build a profile of "normal" instances and evaluate deviations, the Isolation Forest explicitly isolates anomalous observations. The foundational intuition of the algorithm is that abnormal instances are "few and different" from the majority of the data, making them more susceptible to isolation.

The algorithm operates by constructing an ensemble of "Isolation Trees." For each tree, the algorithm recursively divides the dataset by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of that selected feature. Because anomalies reside in sparse regions of the feature space and possess values that deviate significantly from the mean, they are systematically isolated by these random partitions in much fewer steps than normal observations. Therefore, the path length from the root node to the terminating node serves as a natural anomaly score.

B.2 Mathematical Formulation

Consider a dataset $D = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$, where each observation x_i is a d-dimensional feature vector. The algorithm constructs an ensemble of T trees.

For each tree $t \in [1, T]$, a random subsample of size ψ is drawn from D, denoted as D_t. The tree is built recursively until every data point is isolated or the tree reaches a maximum height limit l, defined as:

$$l = \lceil \log_2 \psi \rceil$$

where $\lceil \cdot \rceil$ represents the ceiling function.

For a given instance x, let $h_{D_t}(x)$ represent the path length (the number of edges traversed from the root node to a terminating node) in a single isolation tree. The raw outlier score is defined as the expected path length averaged across the entire forest of T trees:

$$h_D(x) = \frac{1}{T} \sum_{t=1}^{T} h_{D_t}(x)$$

To provide a standardized anomaly score that is comparable across different sample sizes, the expected path length is normalized. Because the structure of an Isolation Tree is equivalent to a Binary Search Tree (BST), the average path length of an unsuccessful search is used as a normalizing constant c_n. For a subsample of size n, this is computed as:

$$c_n = 2H(n-1) - 2\frac{n-1}{n}$$

where $H(i)$ is the harmonic number, which can be approximated by $H(i) \approx \ln(i) + \gamma$, with $\gamma \approx 0.5772$ representing Euler's constant.

The final normalized anomaly score $s_D(x)$ is calculated as:

$$s_D(x) = 2^{-\frac{h_D(x)}{c_n}}$$

If an observation's score $s_D(x)$ approaches 1, it indicates a highly anomalous instance (shorter path length). Conversely, a score considerably smaller than 0.5 indicates a normal observation.

B.3 Theoretical Suitability for Extreme Value Detection

A critical property of the Isolation Forest algorithm is its theoretical propensity to isolate extreme-value observations with high efficiency. As demonstrated by Cheng et al. [4], the probability of early isolation in the uppermost layers of the tree increases monotonically with the number of dimensions in which a data point exhibits extreme values.

In the context of the training procedure, an extreme point is highly likely to be separated from the rest of the distribution by the very first random splits, whereas points that represent general boundary noise require much deeper paths to achieve isolation. This geometric bias toward extreme values makes the algorithm uniquely suited for detecting insider trading, as suspicious actors typically manifest extreme deviations in behavioral features—such as maximum position sizing and portfolio concentration—rather than general, multidimensional noise.

Bibliography

[1] Bobby Allyn. Israel accuses two of using military secrets to place polymarket bets. https://www.npr.org/2026/02/12/nx-s1-5712801/polymarket-bets-traders-israel-military, February 12 2026. NPR. Accessed: March 15, 2026.

[2] Paolo Bizzozero, Raphael Flepp, and Egon Franck. What is the effect of insider trading on price efficiency? evidence from a betting exchange. UZH Business Working Paper Series 368, University of Zurich, Department of Business Administration, 2017. ISSN: 2296-0422.

[3] Robert Bongart and Johan Conradsson. Information and prices on a betting exchange: A study of the market microstructure and the efficiency of a betting exchange. Master's thesis, Stockholm School of Economics, 2006.

[4] Guang Cheng, Christian T. Lundblad, Zhishu Yang, and Qi Zhang. Detecting insider trading in the era of big data and machine learning. *SSRN Electronic Journal*, 2025.

[5] Amy Fan. How anonymous bettors cashed in on the iran strike, just hours before it happened. *The New York Times*, March 3 2026. Accessed: March 10, 2026.

[6] Parameswaran Gopikrishnan, Vasiliki Plerou, Xavier Gabaix, and H Eugene Stanley. Statistical properties of share volume traded in financial markets. *Physical Review E*, 62(4):R4442, 2000.

[7] Filippo Leveni, Guilherme Weigert Cassales, Bernhard Pfahringer, Albert Bifet, and Giacomo Boracchi. Online isolation forest. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Burch, Simon Duan, Hao Jia, and Daniel Tarlow, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 27288–27298. PMLR, 21–27 Jul 2024.

[8] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, December 2008.

[9] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. On detecting clustered anomalies using sciforest. In *Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, Lecture Notes in Computer Science, pages 274–290. Springer Berlin Heidelberg, 2010.

[10] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[11] Hunter Ng, Lin Peng, Yubo Tao, and Dexin Zhou. Price discovery and trading in prediction markets. *Available at SSRN 5331995*, 2025.

[12] Polymarket. Polymarket 101. `https://docs.polymarket.com/polymarket-101`, 2026. Accessed: March 10, 2026.

[13] Polymarket. Polymarket api. `https://docs.polymarket.com/api-reference/introduction`, 2026. Accessed: March 10, 2026.

[14] Oriol Saguillo, Vahid Ghafouri, Lucianna Kiffer, and Guillermo Suarez-Tangil. Unravelling the probabilistic forest: Arbitrage in prediction markets. *arXiv preprint arXiv:2508.03474*, 2025.

[15] Hyun Song Shin. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *The Economic Journal*, 102(411):426–435, 1992.

[16] Storm Slivkoff. Polymarket volume is being double-counted. https://www.paradigm.xyz/2025/12/polymarket-volume-is-being-double-counted, December 2025. Paradigm. Accessed: March 10, 2026.

[17] Joe Vardon and Mike Vorkunov. The NBA players, coaches and gamblers at the center of a federal betting investigation. *The New York Times*, November 2025. Accessed: March 10, 2026.

[18] Karl Whelan. How does inside information affect sports betting odds? *Scottish Journal of Political Economy*, 72(5), 2023.