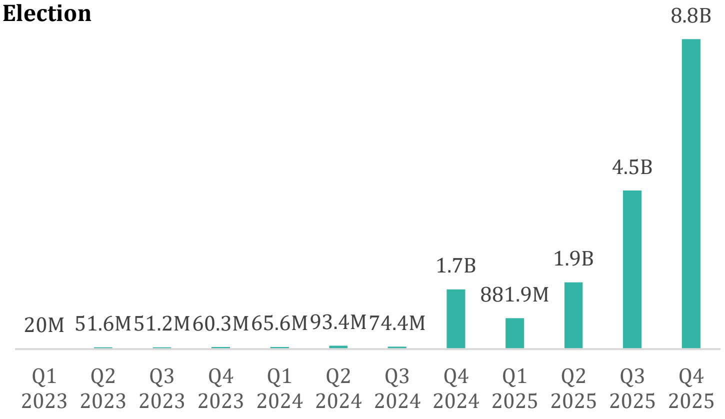
Arbitraging Prediction Markets

Overview of Prediction Markets

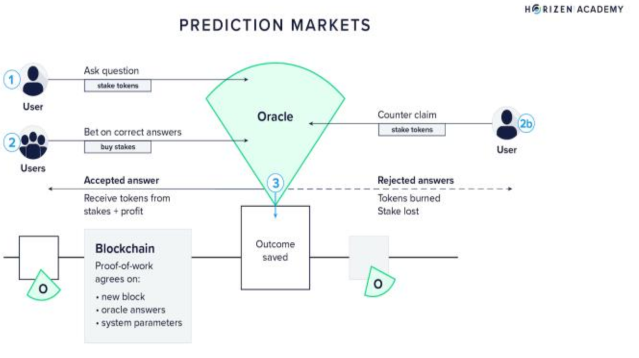
Prediction markets represent one of the most compelling alpha opportunities to emerge in the last decade. They are regulated exchanges where participants trade binary event contracts, which are instruments that pay $1 if a specified outcome occurs and $0 otherwise. The contract price at any time represents the market's implied probability of that event occurring. For instance, a contract priced at $0.65 on "Will the Fed cut rates in June?" implies the market assigns a 65% probability to that outcome. The prediction market landscape is dominated by two platforms (Kalshi & Polymarket) that collectively control approximately 97.5% of total volume. The former is more compliance-focused and operates a quote-driven orderbook where "makers" post offers and "takers" accept them, and the latter is a crypto-native, blockchain-based prediction market.

The sector's growth trajectory has been remarkable, with trading volume expanding from $74.4 million in Q3 2024 to approximately $9 billion in 2024, with the U.S. presidential election being its biggest driver. Total notional volume surged from under $100 million per month to over $13 billion per month by late 2025. Concurrently, monthly active users grew from approximately 4,000 to over 600,000 in the same timeframe. This growth has been accompanied by significant capital formation: Polymarket raised over $250 million across multiple funding round. Kalshi, secured $300 million in a funding round co-led by Andreessen Horowitz (a16z) and Sequoia Capital on October 10, 2025, reaching a $5 billion valuation.s between 2024 and 2025, culminating in the $2 billion ICE strategic investment. Besides regulatory uncertainty in certain regions as a major industry risk, we foresee continued acceleration with major events including the World Cup and the World Baseball Classic which drives sustained engagement.

Surge in prediction markets, fueled by the U.S. Presidential Election



The process of how decentralized prediction markets work



Source: Horizen Academy

Understanding Logarithmic Market Scoring Rule (LMSR)

Polymarket primarily uses an Automated Market Makers (AMMs) –based system rather than a traditional order book. Simply, instead of waiting for a human counterparty, traders buy from and sell to a smart contract that algorithmically sets prices. The Logarithmic Market Scoring Rule (LMSR) is undoubtedly the most important algorithm in prediction markets. LMSR is an automated market maker mechanism, invented by Robin Hanson.

The current price for a stock (assuming a market with 2 stocks) using LMSR is determined by the binary outcome pricing formula as given below:

$$Price = \frac{e^{q_1/b}}{e^{q_1/b} + e^{q_2/b}}$$

where b is an arbitrary constant, q1 is the number of outstanding shares in the stock that you're calculating the price for, q2 is the number of outstanding shares in the other stock

In the context of prediction markets like Polymarket & Kalshi, q1 and q2 represents total shares of YES and NO of an event bought so far. This formula is mathematically equivalent to a softmax function where the price of YES and NO sums up to exactly $1.00. However, many prediction markets feature more than two mutually exclusive outcomes. For instance, a market on "Which company will have the top-ranked AI model this year" might present contracts for OpenAI, Google DeepMind, xAI, Meta and others. In such cases, we use the following equation:

$$Price = \frac{e^{q_1/b}}{e^{q_1/b} + e^{q_2/b} + \cdots + e^{q_n/b}}$$

A critical operational requirement for any market participant is to compute the cost of executing a given trade. Under LMSR, the cost function is derived from the logarithm of the partition function:

$$C(q) = b \ln\left(e^{q_1/b} + e^{q_2/b} + \cdots + e^{q_n/b}\right)$$

To determine the cost of a specific trade, we find the difference between the cost function calculated before and after the trade. The significance of finding the cost of a specific trade is to ensure that the trader pays an amount that reflects the market impact of their position as the more your trade pushes the price, the more expensive each extra share becomes as LMSR is designed so that traders do not pay one flat price for every share. Instead, as you keep buying, the market updates and the next shares purchased get pricier. This is why we believe pure arbitrage is possible, but highly unscalable, especially in small markets, leading to us to venture into other alpha-seeking opportunities.

Structural Inefficiencies in Prediction Markets Create Alpha Opportunities

Despite their informational efficiency at an aggregate level, structural inefficiencies persists in prediction markets, creating a repeatable edge. Such inefficiencies are well-documented in academic literature and confirmed by our proprietary analysis. This section presents an in-depth analysis of a few primary sources of alpha we identified in prediction markets.

1. Favourite-Longshot Bias

The favourite-longshot bias (FLB) is the most extensively studied pricing anomaly in wagering and prediction markets, which essentially describes the empirical pattern in which contracts on unlikely outcomes ("Longshots") are systematically overpriced relative to their true probability of occurring, while contracts on likely outcomes ("Favourites") are systematically underpriced. The most compelling evidence comes from Bürgi, Deng, and Whelan (2026) with uses transaction-level data on 46,282 distinct contracts from 12,403 individual events, spanning from Kalshi's launch in 2021 through April 2025. The dataset encompasses over 300,000 price observations with key findings on how contracts priced below $0.10 win far less often than 10% of the time, with buyers of these contracts losing over 60% of their capital on average. Conversely, contracts priced above $0.50 win more often than their price implies.

2. Slower Incorporation of Information

Traditional financial markets incorporate new information within seconds via algorithmic trading infrastructure, co-located servers and institutional market-making operations. Prediction markets, on the other hand, operate in a fundamentally different regime – their participant base is overwhelmingly retail and a thinner liquidity depth. As such, prediction market prices adjust to new information over a longer period, instead of milliseconds. This creates a persistent window for participants with faster information processing capabilities.

How Apeiron Effectively Captured Such Discrepancies

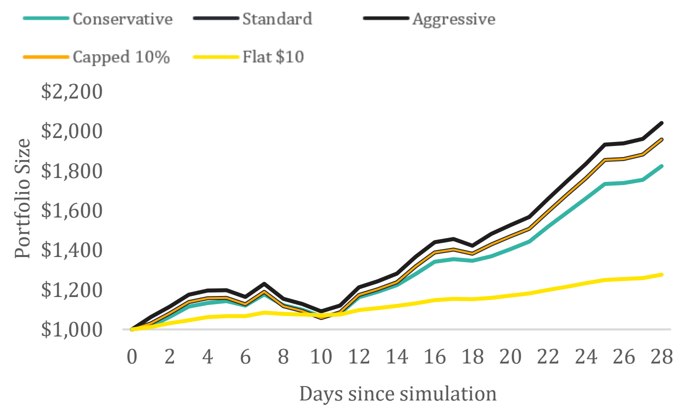
Our strategy deploys a systematic, quantitative framework to identify and capture mispricing across selected prediction markets. tapping onto selected sub-markets (which we will be naming it as Market X for confidentiality purposes) and implementing algorithms that identify the differences in the current bet prices in Market X and our own calculated probabilities by leveraging on quantitative models. After scanning through a myriad of potential bets in Market X for under-priced odds, we compare it with the data and probabilities that our model have generated. We utilise LLMs (specifically Kimi K2.5) to classify each market as BUY_YES/BUY_NO/HOLD alongside a few filters that we have coded it in.

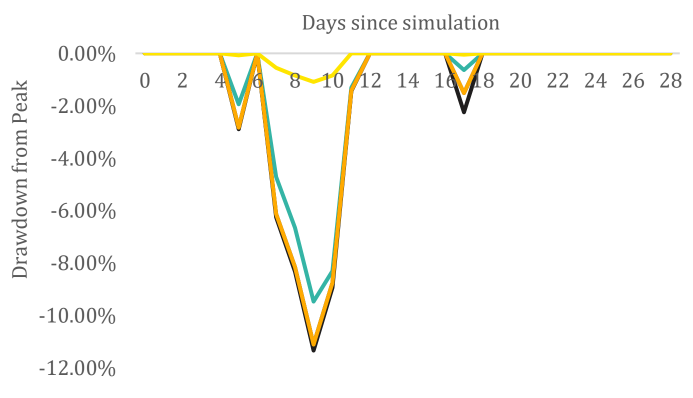
The core alpha engine of our strategy is as follows: For each identified position, we calculated an edge metric defined as the spread between our fair value estimate and the prevailing market price, expressed as a percentage deviation. Positions are only initiated when this deviation exceeds a predefined consensus threshold, ensuring that we act on statistically significant mispricing in the market rather than noise.

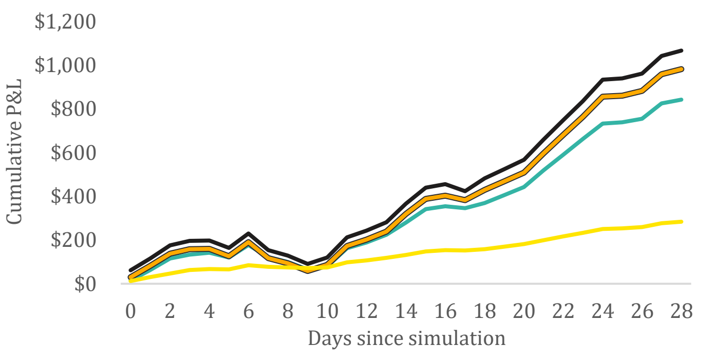
Capital deployment is governed by a Kelly Criterion sizing model that we have built, subject to a fixed position cap and a mandatory cash reserve floor. The key philosophy of our strategy is conservative, where we aim to capture small yet reliable gains on outcomes that are near-certain, avoiding speculative or lottery-ticket positions entirely as part of our risk management.

In summary, we have successfully built and operate a market-neutral alpha engine that profits from systematic inefficiencies in prediction markets, while ensuring that robust risk controls are embedded at every single layer of execution.

Model Performance: Simulated Portfolio Returns







Strategy	Sharpe Ratio (Daily)	Sharpe Ratio (Annualised)
Conservative	0.7625	14.57
Standard	0.7368	14.08
Aggressive	0.7437	14.21
Capped 10%	0.7368	14.08
Flat $10	1.2205	23.32

We assumed a risk-free rate of 4.35% with an annualization factor of $\sqrt{365}$ and a starting capital of $1000