Prediction Markets — Pricing and Information

Day 9 of 10

Prof. Jörg Osterrieder

MSc Seminar: Digital Finance

Spring 2026

Week 2: Advanced Topics in Digital Finance

Day 6 DeFi Derivatives → **Day 7** AI Agents LLMs → **Day 8** ZK Proofs Privacy → **Day 9** Prediction Markets → **Day 10** RWA Tokenization

Today's mission: How prediction markets turn opinions into prices, prices into probabilities, and crowds into oracles.

- How prediction markets work: price = probability
- The LMSR cost function: the Uniswap of probability
- CLOB vs. AMM: Polymarket vs. Augur
- Cross-platform arbitrage: free money or frictionful mirage?
- Information aggregation: the 2024 election case study

The Crowd That Beat the Polls

2024 US Presidential Election

Polymarket predicted the outcome more accurately than every major polling firm, 538, and The Economist's election model.

Source	**Trump Win Probability**	**Correct?**
Polymarket (election morning)	60%	**Closest**
538 model	48% (Harris favored)	Missed
The Economist	46% (Harris favored)	Missed
RealClearPolitics	50% (toss-up)	Weak signal
Actual result	Trump won decisively	

But Also: The Manipulation Question

The Other Side

$40M in arbitrage was extracted from Polymarket in 2024–2025. An estimated 25% of volume may be wash trading.

- "Fredi": a single whale who bet $30M on Trump. Informed trader or manipulator?
- If informed: validates information aggregation
- If manipulation: challenges market efficiency
- Hard to distinguish ex ante—this is the fundamental problem

Today's question: Prediction markets are powerful information aggregation tools. But are they accurate because they are *efficient*, or manipulable because they are *small*?

Prediction Markets: A Growing Ecosystem

Platform	Type	Volume (2024)	Markets
Polymarket	CLOB (Polygon)	\$9B+	Politics, sports, crypto
Kalshi	CLOB (regulated US)	\$1B+	Economics, weather, events
Augur (v2)	AMM (Ethereum)	\$200M	General events
Azuro	AMM (Polygon)	\$500M	Sports betting
PredictIt	CLOB (academic)	\$100M	US politics (capped \$850)

Total prediction market volume in 2024: \$10B+
Compare: Global derivatives market: \$715T notional outstanding.

Prediction markets are tiny but growing fast, and their information value far exceeds their dollar volume.

Today: Pricing, Microstructure, and Information

1. How Prediction Markets Work
2. LMSR: The Automated Market Maker for Predictions
3. Market Structure: CLOB vs. AMM
4. Cross-Platform Arbitrage
5. Information Aggregation
6. Hands-On: LMSR Simulator and Polymarket Analysis

The Core Idea: Price = Probability [ ]

Definition 1 (Prediction Market)

A market where contracts pay \$1 if an event occurs and \$0 otherwise. Under risk-neutrality, the market price equals the crowd's probability estimate:

$$P_{\text{Yes}} = \mathbb{E}^{Q}\left[\mathbf{1}_{\{\text{event occurs}\}}\right]$$

Mechanics:

- "Yes" share price $= \$0.72 \implies$ 72% implied probability
- "No" share price $= \$0.28 \implies$ 28% implied probability
- Always: $P_{\text{Yes}} + P_{\text{No}} = \$1.00$ (by construction)
- As new information arrives, traders buy/sell, and price updates in real time

Key insight: It is a continuous, incentivized poll where participants put money behind their beliefs.

Worked Example: Fed Rate Cut Market

Market: "Will the Fed cut rates in June 2026?"

Current state:

- Yes share = \$0.72 (72%), No share = \$0.28 (28%)

You buy 100 "Yes" shares at \$0.72 each. Cost: \$72.

Scenario A: Fed cuts

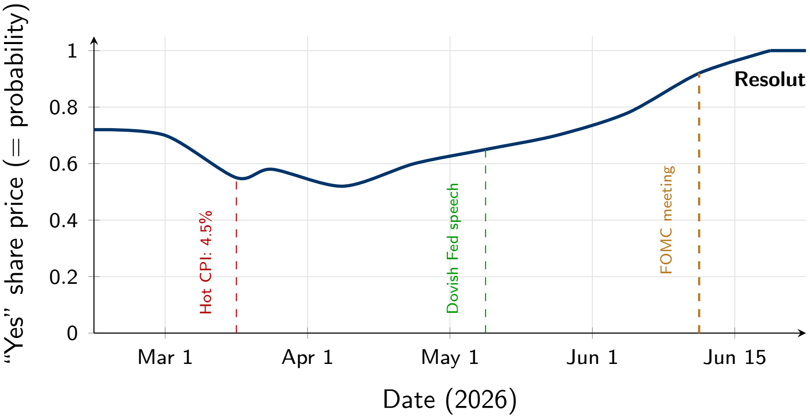
- Each share pays \$1
- Revenue: $100 \times \$1 = \$100$
- Profit: $\$100 - \$72 = \$28$
- Return: $+38.9\%$

Scenario B: No cut

- Each share pays \$0
- Revenue: \$0
- Loss: $-\$72$
- Return: -100%

New info: Hot CPI at 4.5% $\Rightarrow$ traders sell Yes $\Rightarrow$ new price \$0.55. Your 100 shares now worth \$55. Paper loss: \$17.

Price Evolution: Information Drives Probability



Each price movement reflects new information being incorporated by traders. The price path *is* the market's evolving probability estimate.

Multi-Outcome Markets

Markets can have $n > 2$ outcomes:

Example: "Which party wins the 2028 election?"

Outcome	**Share Price**	**Implied Prob.**
Republican	\$0.48	48%
Democrat	\$0.42	42%
Independent	\$0.07	7%
Other	\$0.03	3%
Total	\$1.00	100%

Constraint: $\sum_{i=1}^{n} P_i = 1$ always holds. If prices temporarily deviate (e.g., sum to \$1.03), an arbitrageur can sell all outcomes for \$1.03 and cover for \$1.00 at resolution, pocketing \$0.03.

Checkpoint: Prediction Market Basics

Quick Question

A prediction market for "BTC above \$100K by Dec 2026" shows Yes $=$ \$0.40, No $=$ \$0.58. Is there an arbitrage opportunity?

Checkpoint: Prediction Market Basics

Quick Question

A prediction market for "BTC above \$100K by Dec 2026" shows Yes = \$0.40, No = \$0.58. Is there an arbitrage opportunity?

Answer

Yes. The prices sum to \$0.98, not \$1.00. Buy one Yes (\$0.40) and one No (\$0.58) for a total of \$0.98. At resolution, exactly one pays \$1.00. **Guaranteed profit: \$0.02 per pair** (2.0% return).

This is a *Dutch book*—the market is mispriced and an arbitrageur can extract risk-free profit.

Logarithmic Market Scoring Rule (LMSR) [ ]

Definition 2 (LMSR Cost Function (Hanson, 2003))

$$\boxed{C(\mathbf{q}) = b \cdot \ln\left(\sum_{i=1}^{n} e^{q_i/b}\right)}$$

- q_i: total shares of outcome i purchased so far
- b: **liquidity parameter** (controls price impact)
- n: number of outcomes

Instantaneous price for outcome i:

$$p_i = \frac{\partial C}{\partial q_i} = \frac{e^{q_i/b}}{\sum_{j=1}^{n} e^{q_j/b}} \qquad \text{(softmax of share quantities)}$$

Prices always sum to 1: $\sum_i p_i = 1$ (probability distribution).

LMSR Intuition: The Uniswap of Prediction Markets

Just as Uniswap provides continuous liquidity for token swaps using $x \cdot y = k$, LMSR provides continuous liquidity for probability shares using log-sum-exp.

- The protocol *is* the market maker
- Always willing to buy/sell at the current price
- No human market makers needed
- Parameter b controls the depth of the "book"
- Higher b = more liquidity = less price impact per trade

Market maker's risk:

- Maximum loss is *bounded*:

$$L_{\max} = b \cdot \ln(n)$$

- Binary market ($n = 2$): $L_{\max} = b \cdot \ln 2 \approx 0.693 \cdot b$
- If $b = 100$: max loss = \$69.31
- This is the cost of providing infinite liquidity

LMSR: Worked Example ($b = 100$, Binary Market)

Initial state: $q_{\text{Yes}} = 0,\ q_{\text{No}} = 0.$

Step 1: Initial prices

$$p_{\text{Yes}} = \frac{e^{0/100}}{e^{0/100} + e^{0/100}} = \frac{1}{2} = 0.50 \quad (50\%)$$

Step 2: Trader buys 10 "Yes" shares

$$\begin{aligned}
\text{Cost} &= C(10,0) - C(0,0) \\
&= 100 \ln\left(e^{0.1} + e^{0}\right) - 100 \ln\left(e^{0} + e^{0}\right) \\
&= 100 \ln(1.10517 + 1) - 100 \ln(2) \\
&= 100 \times 0.74440 - 100 \times 0.69315 = \boxed{\$5.13}
\end{aligned}$$

Step 3: New prices

$$p_{\text{Yes}} = \frac{e^{0.1}}{e^{0.1} + 1} = \frac{1.10517}{2.10517} = 0.5250 \quad (52.5\%)$$

10 shares cost \$5.13 and moved probability from 50.0% to 52.5%. Average price per share:

The Liquidity Parameter b: Controlling Price Impact

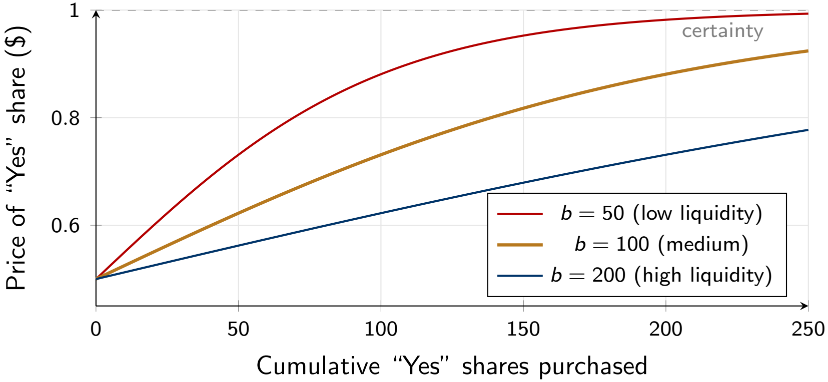
Same 10 "Yes" shares purchased, different b values:

b	Cost	New p_{Yes}	Price Impact	Max Loss
50	\$5.27	55.0%	5.0 pp	\$34.66
100	\$5.13	52.5%	2.5 pp	\$69.31
200	\$5.06	51.2%	1.2 pp	\$138.63
500	\$5.03	50.5%	0.5 pp	\$346.57

Trade-off:

- Higher b: deeper liquidity, less price impact, but higher maximum loss
- Lower b: cheaper to subsidize, but prices move aggressively
- Optimal b depends on expected information quality and budget

LMSR Price Curves for Different b



Higher b produces a flatter curve: the same number of shares moves the price less. This is the "depth" of the automated market maker.

Why LMSR Works: Key Properties

1. **Prices always sum to 1:**

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \frac{e^{q_i/b}}{\sum_j e^{q_j/b}} = 1 \qquad \text{(valid probability distribution)}$$

2. **Truth-telling is incentive-compatible:** For a myopic trader with belief $\hat{p}_i$, buying until $p_i = \hat{p}_i$ maximizes expected profit (proper scoring rule)
3. **Bounded loss:** $L_{\max} = b \ln(n) < \infty$ (the market maker's subsidy is finite and known in advance)
4. **Path independence:** The cost to move from state $\mathbf{q}^A$ to $\mathbf{q}^B$ depends only on the endpoints, not the order of trades

Compare to Uniswap: LMSR has bounded loss; Uniswap's CPMM does not (impermanent loss is unbounded). LMSR is the theoretically optimal AMM for discrete outcome markets.

Checkpoint: LMSR Calculation

Quick Exercise

A binary LMSR market has $b = 200$ and current state $q_{\text{Yes}} = 50$, $q_{\text{No}} = 0$.

1. What is the current price of a "Yes" share?
2. How much does it cost to buy 20 more "Yes" shares?

Checkpoint: LMSR Calculation

Quick Exercise

A binary LMSR market has $b = 200$ and current state $q_{\text{Yes}} = 50$, $q_{\text{No}} = 0$.

1. What is the current price of a "Yes" share?
2. How much does it cost to buy 20 more "Yes" shares?

Answer

1. $p_{\text{Yes}} = e^{50/200}/(e^{50/200} + e^{0/200}) = e^{0.25}/(e^{0.25} + 1) = 1.284/2.284 = 0.5622$ (56.2%)
2. Cost $= C(70, 0) - C(50, 0) = 200 \ln(e^{0.35} + 1) - 200 \ln(e^{0.25} + 1)$
   $= 200 \times 0.8544 - 200 \times 0.8259 = 170.88 - 165.17 = \$5.71$
   New $p_{\text{Yes}} = e^{0.35}/(e^{0.35} + 1) = 0.5866$ (58.7%)

Two Market Architectures for Prediction Markets

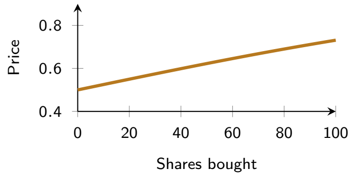
CLOB (Polymarket):

Spread: $0.02

Bids	Asks
$0.71 500 shares	$0.73 200 shares
$0.70 1,200	$0.74 800
$0.69 3,000	$0.75 1,500

Professional market makers quote bids/asks. Tight spreads.

AMM / LMSR (Augur):



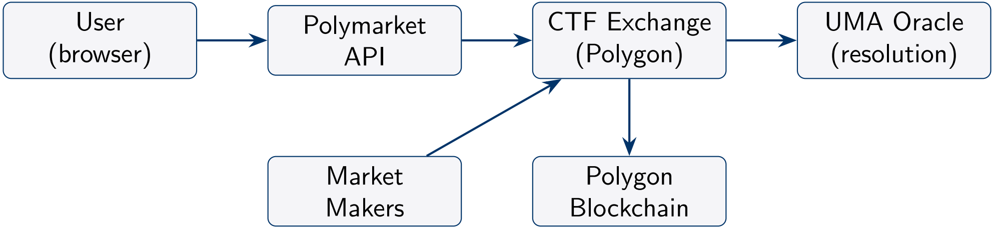
Algorithm provides liquidity. No order book. Always available.

CLOB vs. AMM: Detailed Comparison

Feature	**CLOB (Polymarket)**	**AMM (LMSR)**
Liquidity source	Active market makers	Algorithm (subsidized)
Spread	Tight (1–3 cents)	Wider (depends on b)
Capital needed	Market makers supply	Protocol subsidizes
Best for	Popular, high-volume	Niche, new markets
Price discovery	Excellent	Good
Manipulation cost	High (deep book)	Lower (bounded loss b)
Composability	Moderate	High (on-chain native)
User experience	Familiar (like stocks)	Simple (buy/sell button)
Example	Polymarket, Kalshi	Augur, Gnosis, Azuro

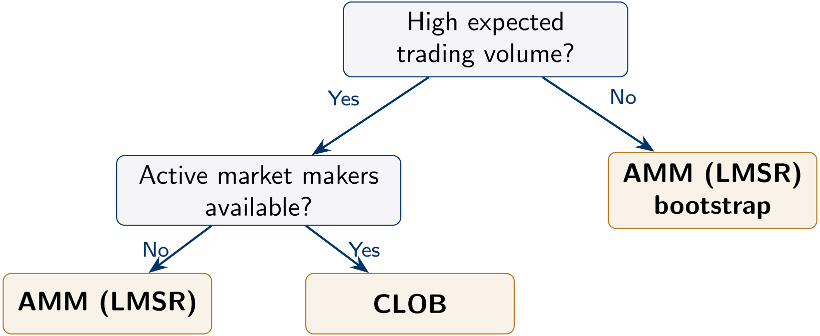
Trend: CLOBs dominate high-volume markets. AMMs serve as bootstrapping mechanisms for long-tail markets with uncertain liquidity.

Polymarket: Architecture of the Largest Prediction Market



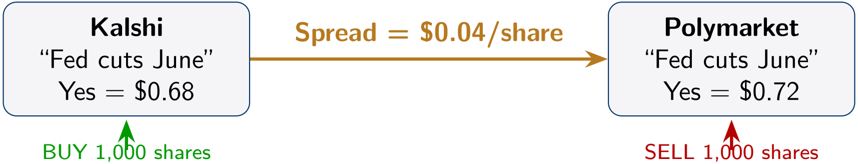
- **Conditional Token Framework (CTF):** ERC-1155 tokens on Polygon
- **Settlement:** USDC on Polygon
- **Resolution:** UMA optimistic oracle (human-verified with dispute mechanism)
- **Fees:** 0% maker, 2% on winnings
- **2024 volume:** \$9B+, 200K+ active traders

Decision Framework: CLOB vs. AMM



Polymarket started with AMM liquidity and migrated to CLOB as volume grew. The AMM bootstraps; the CLOB scales.

Cross-Platform Arbitrage: \$0.04/Share Profit



	If Fed Cuts	**If No Cut**
Kalshi Yes pays	\$1,000	\$0
Polymarket Yes obligation	−\$1,000	\$0
Net from event	\$0	\$0
Spread captured	**\$40**	**\$40**

Guaranteed profit: \$40 on \$1,400 capital = 2.9% return.

Why Does the Arbitrage Persist?

\$40M extracted from Polymarket in 2024–2025. If it is free money, why does it not disappear?

Friction	**Cost / Impact**
Capital lockup	\$1,400 tied up for weeks–months
Settlement mismatch	Kalshi: USD via bank wire; Polymarket: USDC on Polygon
Withdrawal timing	Moving money between platforms: days
Regulatory access	Kalshi is US-regulated; Polymarket is offshore
Counterparty risk	Platform insolvency (remember FTX)
Resolution disputes	Different oracles may resolve differently

Formal model:

$$\Pi = (P_A - P_B) \cdot N - c_{\text{tx}} - c_{\text{capital}} \cdot T$$

where c_{capital} is the opportunity cost of locked capital and T is time to resolution. The spread reflects these frictions.

Intra-Market Arbitrage: Dutch Books

If prices for mutually exclusive outcomes do not sum to $1:

Example: Sum $>$ \$1 (overround)

- Republican = \$0.52, Democrat = \$0.45, Other = \$0.06
- Sum = \$1.03 (the market charges a 3% "vig")
- **Sell** one share of each: receive \$1.03 now, pay \$1.00 at resolution. Profit: \$0.03 per set.

Example: Sum $<$ \$1 (underround)

- Republican = \$0.48, Democrat = \$0.42, Other = \$0.07
- Sum = \$0.97
- **Buy** one share of each: pay \$0.97 now, receive \$1.00 at resolution. Profit: \$0.03 per set.

In efficient markets, the sum stays very close to \$1.00. Any deviation is quickly arbitraged away.

Scale of Prediction Market Arbitrage (2024–2025)

Arbitrage Type	Estimated Volume	Avg. Spread
Cross-platform (Poly $\leftrightarrow$ Kalshi)	\$25M+	3–5%
Dutch book (intra-market)	\$8M+	1–3%
Related-market inconsistencies	\$7M+	2–4%
Total extracted	**\$40M+**	

Compare to traditional markets:

- Equity cross-listing arbitrage: spreads of 0.01–0.05%
- FX triangular arbitrage: spreads of $<$0.001%
- Prediction markets: spreads of 1–5% (100–1,000$\times$ wider)

Implication: Prediction markets are still *very* inefficient by traditional finance standards. This represents both opportunity and a maturity gap.

Information Aggregation: Markets as Oracles [ ]

Hayek's insight (1945): Prices aggregate dispersed private information that no single agent possesses.

Applied to prediction markets:

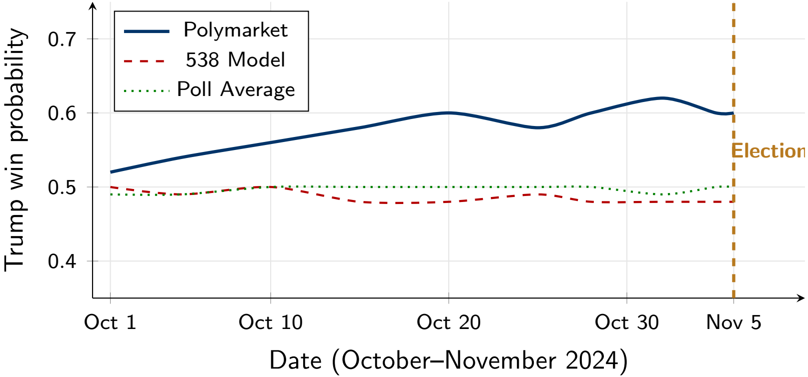
- Trader A follows CPI data
- Trader B monitors Fed speeches
- Trader C has contacts in the bond market
- Trader D analyzes social media sentiment

When all trade on their private signals, the market price *synthesizes* every signal:

$$p^* = \mathbb{E}^Q[\mathbf{1}_{\text{event}} \mid s_1, s_2, \ldots, s_N]$$

No single trader knows everything, but the price reflects *as if* someone did.

Case Study: 2024 US Election Forecast



Polymarket consistently showed Trump at 55–62%, while polls and models called it a toss-up. The market was right.

Why Markets Beat Polls

Dimension	Polls	Prediction Markets
Incentive to be honest	None	Financial (skin in the game)
Information sources	Survey responses	All available information
Update frequency	Weekly	Continuous (24/7)
Non-response bias	Yes (systematic)	No (self-selecting traders)
Social desirability bias	Yes	No (anonymous trading)
Late-breaking info	Missed (lag)	Incorporated instantly
Sample size	1,000–5,000	Market depth (unlimited)

The Grossman–Stiglitz paradox:

- If prices are fully efficient, there is no incentive to gather information
- Resolution: *noise traders* provide profit opportunities for informed traders
- Information aggregation works *because* markets are not perfectly efficient

Limitations and Manipulation Risks

Limitation	Example
Thin liquidity	Most markets have <\$100K volume
Wash trading	$\sim$25% of Polymarket volume (estimated)
Whale manipulation	"Fredi" bet \$30M on Trump (2024)
Regulatory uncertainty	Elections banned on some platforms
Moral hazard	Betting on negative events
Risk premium	$P_{\text{Yes}} \neq \mathbb{P}(\text{event})$ if risk-averse

Manipulation cost in LMSR ($b = 1{,}000$):

- Move price from 50% to 80%: $q = 1{,}000 \ln 4 = 1{,}386$ shares. Cost: \$916
- Move to 95%: $q = 1{,}000 \ln 19 = 2{,}944$ shares. Cost: \$2,303
- Key: manipulation is expensive and *temporary*—informed traders push the price back

Checkpoint: Information Aggregation

Discussion Question

A pharmaceutical company's CEO buys \$500K of "FDA approval" shares on a prediction market the day before the FDA decision, moving the price from 60% to 85%. Is this:

(a) Illegal insider trading

(b) Valuable information aggregation

(c) Both

Checkpoint: Information Aggregation

Discussion Question

A pharmaceutical company's CEO buys $500K of "FDA approval" shares on a prediction market the day before the FDA decision, moving the price from 60% to 85%. Is this:

a. Illegal insider trading
b. Valuable information aggregation
c. Both

Key Insight

This is the fundamental tension. In equity markets, this is clearly insider trading. In prediction markets, regulation is unclear. The information *is* valuable to the market—but acquired through privileged access. Current CFTC rules for Kalshi prohibit insider trading; Polymarket (offshore) has no such rule.

Hands-On Session

LMSR Simulator and Polymarket Data Analysis

Python notebook on course platform

Step 1: Build an LMSR Simulator

LMSR implementation

```
import numpy as np

def lmsr_cost(q, b):
    """Cost function C(q) = b * ln(sum(exp(q_i/b)))"""
    return b * np.log(np.sum(np.exp(q / b)))

def lmsr_prices(q, b):
    """Prices = softmax(q/b)"""
    e = np.exp(q / b)
    return e / np.sum(e)

def buy_shares(q, b, outcome, n_shares):
    """Buy n_shares of outcome. Returns cost."""
    q_new = q.copy()
```

Step 2: LMSR Exercises

Using your simulator with $b = 100$, binary market:

1. Starting from $q = [0, 0]$, buy 10, 20, 50, 100 "Yes" shares. Record cost and new price for each.
2. Plot the **marginal cost curve** (cost per additional share as a function of cumulative shares purchased).
3. Repeat for $b = 50$ and $b = 200$. Overlay the three curves.
4. **Market maker loss:** Simulate a market where the outcome resolves to "Yes." Calculate the market maker's total loss as final payout − total revenue from trades.
5. Verify: is the loss $\leq b \ln 2$?

Key finding: As b increases, the cost curve flattens (more liquidity) but the maximum loss increases proportionally.

Step 3: Polymarket Historical Data Analysis

Data loading

```
import pandas as pd
import matplotlib.pyplot as plt

# Load Polymarket election market data
data = pd.read_csv('polymarket_election_2024.csv',
                     parse_dates=['timestamp'])
data['midpoint'] = (data['best_bid'] + data['best_ask']) / 2

# Plot probability trajectory
plt.plot(data['timestamp'], data['midpoint'])
plt.ylabel('Trump Win Probability')
plt.title('Polymarket: 2024 Election')
```

Tasks:

1. Plot the full probability trajectory (Jul–Nov 2024)

Step 4: Forecast Accuracy — Brier Score

Definition 3 (Brier Score)

$$\text{BS} = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2$$

where p_t is the forecast probability and $o_t \in \{0, 1\}$ is the actual outcome. Lower = better.

Calculate for the 2024 election:

Source	Final Forecast (p)	Brier Score
Polymarket	0.60	$(0.60 - 1)^2 = 0.16$
538	0.48	$(0.48 - 1)^2 = 0.27$
Polls	0.50	$(0.50 - 1)^2 = 0.25$

Polymarket's Brier score was 41% better than 538.

Step 5: Spread and Liquidity Analysis

Using Polymarket order book data:

1. Calculate the average bid-ask spread over time. Does the spread tighten as the event approaches?
2. Plot daily trading volume alongside the spread. What is the relationship?
3. Estimate the **effective spread** (difference between trade price and midpoint) for different trade sizes.
4. Compare market depth across 5 different Polymarket markets. Do political markets have tighter spreads than sports or crypto markets?

Expected finding: Spreads narrow as resolution approaches and volume increases. This is consistent with declining uncertainty and increasing information flow.

Step 6: Cross-Platform Arbitrage Detection

Compare simultaneous prices on Polymarket and Kalshi:

1. For 5 overlapping markets, plot the price difference over time
2. Calculate the **average absolute spread** between platforms
3. Estimate the **net arbitrage profit** accounting for:
   - Trading fees (2% on Polymarket, 1% on Kalshi)
   - Capital lockup cost (assume 5% annual opportunity cost)
   - Expected time to resolution
4. Is the net profit still positive after frictions?
5. At what capital lockup duration does the arbitrage become unprofitable?

Discussion: How does this compare to cross-exchange crypto arbitrage (Day 1)?

Key Takeaways: Day 9

1. **Prediction market prices are probabilities:** a \$0.72 "Yes" share implies 72% probability (under risk-neutrality)
2. **LMSR** is the canonical AMM for prediction markets: $C(\mathbf{q}) = b \ln\left(\sum e^{q_i/b}\right)$, with bounded market-maker loss $b \ln n$
3. **CLOB beats AMM** for high-volume markets (Polymarket); AMM bootstraps new, illiquid markets
4. **Cross-platform arbitrage** persists due to settlement frictions, capital lockup, and regulatory barriers
5. **Markets aggregate information** better than polls: Polymarket beat 538 and polling averages in the 2024 election

Tomorrow: Real-world asset tokenization—where crypto meets the \$600 trillion traditional asset market.

Further Reading

Core references:

- Hanson (2003): LMSR design [1]
- Wolfers & Zitzewitz (2004): Prediction markets survey [2]

Recommended reading:

- Manski (2006): "Interpreting the Predictions of Prediction Markets"
- Othman et al. (2013): "A Practical Liquidity-Sensitive AMM"
- Polymarket documentation: https://docs.polymarket.com

Questions? • Office hours: by appointment • joerg.osterrieder@usi.ch

References I

[1] Robin Hanson. "Combinatorial Information Market Design". In: *Information Systems Frontiers* 5.1 (2003), pp. 107–119.

[2] Justin Wolfers and Eric Zitzewitz. "Prediction Markets". In: *Journal of Economic Perspectives* 18.2 (2004), pp. 107–126.