The Longshot Bias in Prediction Markets: Evidence from 68 Million Trades

Jonathan Becker
jonathan@jbecker.dev

November 28, 2025

Abstract

We analyze 67.8 million trades worth \$8.6 billion from Kalshi, a CFTC-regulated prediction market, to investigate market calibration and the longshot bias. Contracts priced below 10 cents resolve favorably 30–40% less often than their price-implied probability, confirming a systematic longshot bias at unprecedented scale. This bias diminishes monotonically with price and reverses slightly for near-certain outcomes above 95 cents. We document substantial heterogeneity in calibration: large trades (above \$5,000) achieve positive excess returns, while small trades consistently underperform. Contrarian traders who buy after price declines significantly outperform momentum traders. Market efficiency improves as resolution approaches, with the excess win rate improving from -1.8% at 3–7 days to -0.4% within the final hour. Bid-ask spreads follow a U-shaped pattern across probabilities, widest at 50%. These findings support probability weighting theories from behavioral finance, demonstrate limits to arbitrage in prediction markets, and suggest calibration adjustments for practitioners using market prices as probability forecasts.

Keywords: longshot bias, prediction markets, market efficiency, behavioral finance, probability weighting, Kalshi

Contents

1 Introduction

Prediction markets aggregate dispersed information through trading, theoretically producing prices that reflect collective beliefs about uncertain future events [1, 19]. Under the efficient market hypothesis, contract prices should equal the true probability of the underlying event occurring. However, decades of research have documented systematic deviations from this ideal, most notably the *longshot bias*—the tendency for low-probability outcomes to be overpriced relative to their true likelihood [8, 16, 17].

First documented in horse-race betting by Griffith [8], the longshot bias has been replicated across sports betting markets [21], laboratory experiments [3], and prediction markets [20]. Multiple theoretical explanations have been proposed, including risk-seeking preferences in the loss domain [10], probability weighting that overweights small probabilities [12], heterogeneous beliefs with limited arbitrage [15], and the skewness preference of gamblers seeking high-variance payoffs [7].

Despite extensive theoretical work, empirical studies of prediction market calibration have been limited by data availability. Prior research has relied on relatively small samples—hundreds to thousands of markets—often from platforms with thin liquidity and significant trading frictions [2, 13]. The recent growth of regulated prediction markets, particularly in the United States following the expansion of CFTC-authorized platforms, provides an unprecedented opportunity to study these phenomena at scale.

In this paper, we analyze 67.8 million trades worth \$8.6 billion from Kalshi, a CFTC-regulated prediction market operating since 2021. This dataset represents, to our knowledge, the largest systematic analysis of individual trades in prediction markets to date. Our granular transaction-level data allows us to investigate not only aggregate market calibration but also the behavior of individual traders, the dynamics of price formation over market lifetimes, and the relationship between trade characteristics and forecasting accuracy.

Our analysis yields several novel findings:

1. **Confirmation of longshot bias at scale:** We document systematic overpricing of low-probability contracts across nearly 68 million trades. Contracts priced at 5 cents (implying 5% probability) resolve favorably only 3.5% of the time—a 30% underperformance relative to their price-implied probability. This bias diminishes monotonically with price, with contracts above 85 cents slightly outperforming their implied probabilities.

2. **Trade size predicts calibration:** Larger trades exhibit significantly better calibration than smaller trades. Trades exceeding \$5,000 achieve positive excess returns, while trades under \$100 consistently underperform. This pattern suggests sophisticated traders systematically exploit the mispricing introduced by smaller, less-informed participants.

3. **Temporal dynamics of information aggregation:** Markets become more efficient as resolution approaches. The excess win rate (actual minus expected) improves from -1.8% in trades placed 3–7 days before resolution to -0.4% in trades placed within the final hour. This pattern reflects both information revelation and the mechanical convergence of prices to binary outcomes.

4. **Contrarian traders outperform:** Traders who buy contracts that have recently declined in price significantly outperform those who chase momentum. Trades following price drops exceeding 10% yield excess returns of +5.4%, while trades following similar increases yield -2.7%.

5. **Liquidity-driven spread dynamics:** Bid-ask spreads follow a U-shaped pattern across the probability spectrum, widest at intermediate probabilities (around 50%) where uncertainty is maximal. This pattern deviates from theoretical predictions and suggests market maker behavior responsive to volatility rather than edge.

These findings contribute to several literatures. For prediction market design, our results suggest that market calibration varies systematically with trade size and timing, implying that liquidity provision and trader composition affect informational efficiency. For behavioral finance, we provide large-sample evidence consistent with prospect theory's probability weighting function while documenting heterogeneity across trader sophistication levels. For forecasters using prediction market prices, our findings suggest that raw prices systematically overstate the likelihood of low-probability events and that adjustments based on trade-weighted prices may improve calibration.

The remainder of this paper proceeds as follows. Section 2 describes our dataset and methodology. Section 3 presents our main results on market calibration and the longshot bias. Section 4 analyzes heterogeneity across traders and trade characteristics. Section 5 examines temporal dynamics in price formation. Section 6 discusses implications and limitations, and Section 7 concludes.

2 Data and Methodology

2.1 The Kalshi Platform

Kalshi is a CFTC-regulated exchange for event contracts, operational since July 2021. Unlike earlier prediction markets that operated in legal gray areas (e.g., Intrade, PredictIt), Kalshi functions as a designated contract market under U.S. commodity trading law, subject to position limits, trading surveillance, and financial reporting requirements. This regulatory framework provides some assurance regarding data integrity and market manipulation controls.

The platform offers binary contracts on a diverse range of events spanning politics, economics, weather, sports, and current events. Each contract pays \$1 if its specified outcome occurs and \$0 otherwise. Contracts trade continuously at prices between \$0.01 and \$0.99, with the price interpretable as the market-implied probability of the event. The platform operates a central limit order book with price-time priority, and traders can submit both market and limit orders.

2.2 Dataset Construction

We obtained complete transaction records from Kalshi's public API, covering all trades executed on the platform from inception through November 2024. Our dataset comprises:

- **Trades:** 67,761,406 individual transactions
- **Total volume:** \$8.59 billion in notional value
- **Markets:** Approximately 50,000 unique event contracts
- **Price coverage:** All integer cent prices from 1 to 99

Each trade record includes the ticker symbol, execution timestamp, price, quantity (number of contracts), taker side (yes or no), and the prices for both sides at execution. We merge trade data with market metadata containing resolution outcomes, expiration times, and contract descriptions.

2.3 Key Variables

We construct several variables central to our analysis:

Win Rate. For each trade, we observe whether the taker's position ultimately resolved favorably. Letting $W_i \in \{0, 1\}$ indicate whether trade i won, the empirical win rate at price p is:

$$\text{WinRate}(p) = \frac{\sum_{i:p_i=p} W_i}{\sum_{i:p_i=p} 1} \tag{1}$$

where p_i is the execution price for trade i.

Excess Win Rate. Under perfect calibration, contracts purchased at price p should win p% of the time. We define the excess win rate as:

$$\text{ExcessWinRate}(p) = \text{WinRate}(p) - p \tag{2}$$

Positive values indicate underpricing (contracts win more often than their price implies), while negative values indicate overpricing.

Trade Size. We measure trade size in dollar terms as $\text{Size}_i = p_i \times \text{Contracts}_i$, representing the capital at risk for the taker.

Time to Resolution. For each trade, we compute the time remaining until market resolution: $\tau_i = t_{\text{resolve}} - t_i$, where t_i is the trade timestamp and t_{resolve} is the resolution time.

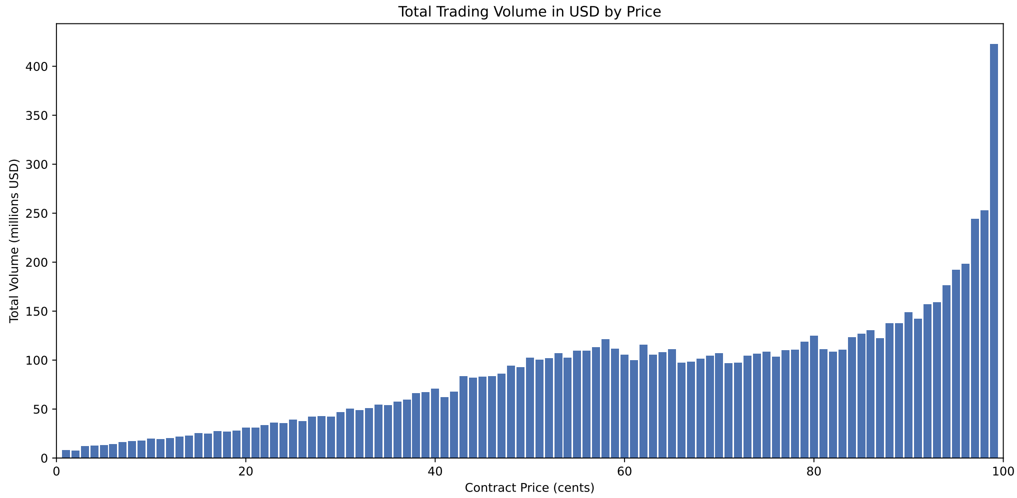

Figure 1: Total trading volume (USD) by contract price. Volume concentrates heavily at extreme prices, with contracts at 99 cents accounting for the highest single-price volume. The asymmetry reflects both contract availability and trader preferences for high-conviction positions.

2.4 Volume Distribution

Figure 1 displays the distribution of trading volume across prices. Volume exhibits a striking asymmetric pattern, with the highest concentration at extreme prices—particularly at 99 cents, which alone accounts for over $420 million in trading volume. This reflects both the natural supply of contracts at different probability levels and traders' strong demand for near-certain outcomes.

Figure 2 shows the number of contracts traded at each price level. While the dollar volume peaks at high prices (where each contract costs more), the raw contract count peaks at low prices, reflecting the lottery-like appeal of cheap longshots.

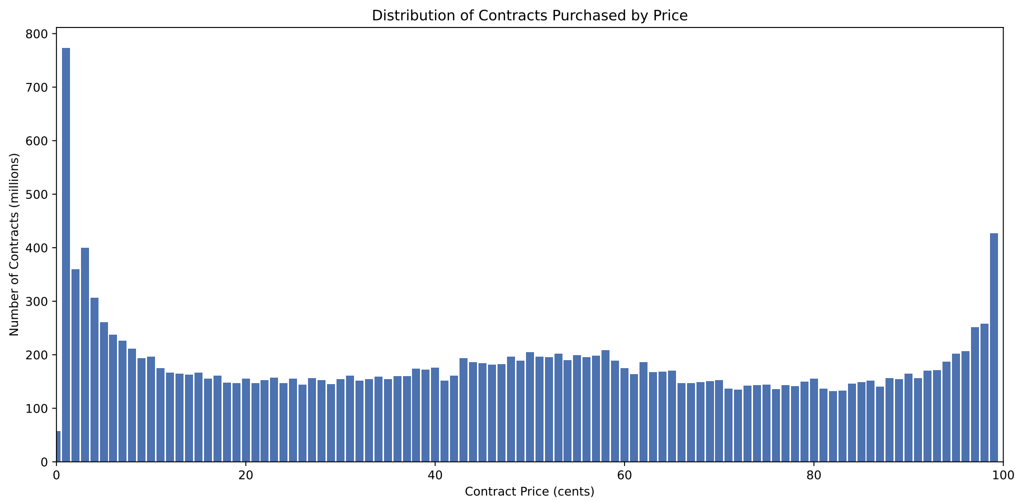

Figure 2: Total contracts traded by price. Contract volume peaks at 1 cent, where over 770 million contracts were traded, reflecting strong demand for cheap longshot positions despite their low expected value.

2.5 Trade Size Distribution

Figure 3 displays the average trade value across price levels. Trade sizes increase monotonically with price: the average trade at 99 cents is $406, compared to just $5.50 at 1 cent. This pattern reflects both mechanical effects (higher-priced contracts cost more per unit) and behavioral factors (traders may size positions based on conviction).

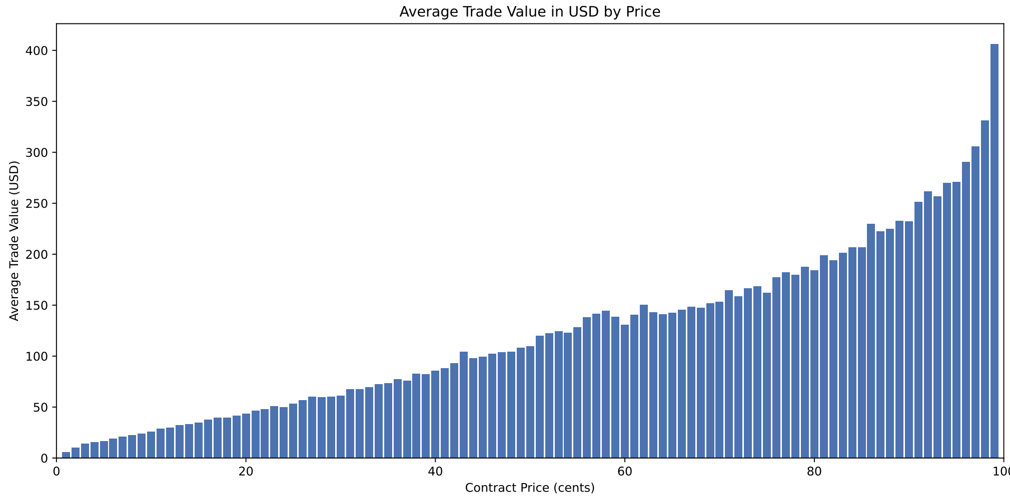

Figure 3: Average trade value (USD) by contract price. Trade size increases monotonically with price, reflecting both the higher per-contract cost and potentially greater conviction among traders of high-probability outcomes.

2.6 Statistical Framework

Our primary specification tests whether win rates deviate systematically from price-implied probabilities. For inference, we compute standard errors accounting for the binary nature of outcomes:

$$\mathrm{SE}(\mathrm{WinRate}) = \sqrt{\frac{p(1-p)}{n}} \tag{3}$$

where n is the number of trades in the bin. We report z-statistics computed as $z = \mathrm{ExcessWinRate}/\mathrm{SE}$ and corresponding p-values from the standard normal distribution.

Given our large sample sizes, virtually all deviations from perfect calibration are statistically significant at conventional levels. We therefore focus on economic magnitudes—the size of the calibration errors in percentage points—rather than statistical significance alone.

2.7 Sample Restrictions

We restrict our analysis to markets that reached final resolution with a binary yes/no outcome, had at least one trade executed, and were not cancelled or voided by the exchange. These restrictions yield our final sample of 67.8 million trades. We exclude the small number of trades at prices of exactly 0 or 100 cents, as these represent near-certain outcomes where calibration analysis is uninformative.

3 Market Calibration and the Longshot Bias

3.1 Aggregate Calibration

Our primary finding is a systematic and economically significant longshot bias across nearly 68 million trades. Figure 4 plots the empirical win rate against contract price for each integer cent value from 1 to 99. Under perfect calibration, all points would lie on the 45-degree line. Instead, we observe a clear pattern: low-priced contracts underperform their price-implied probability, while high-priced contracts slightly outperform.

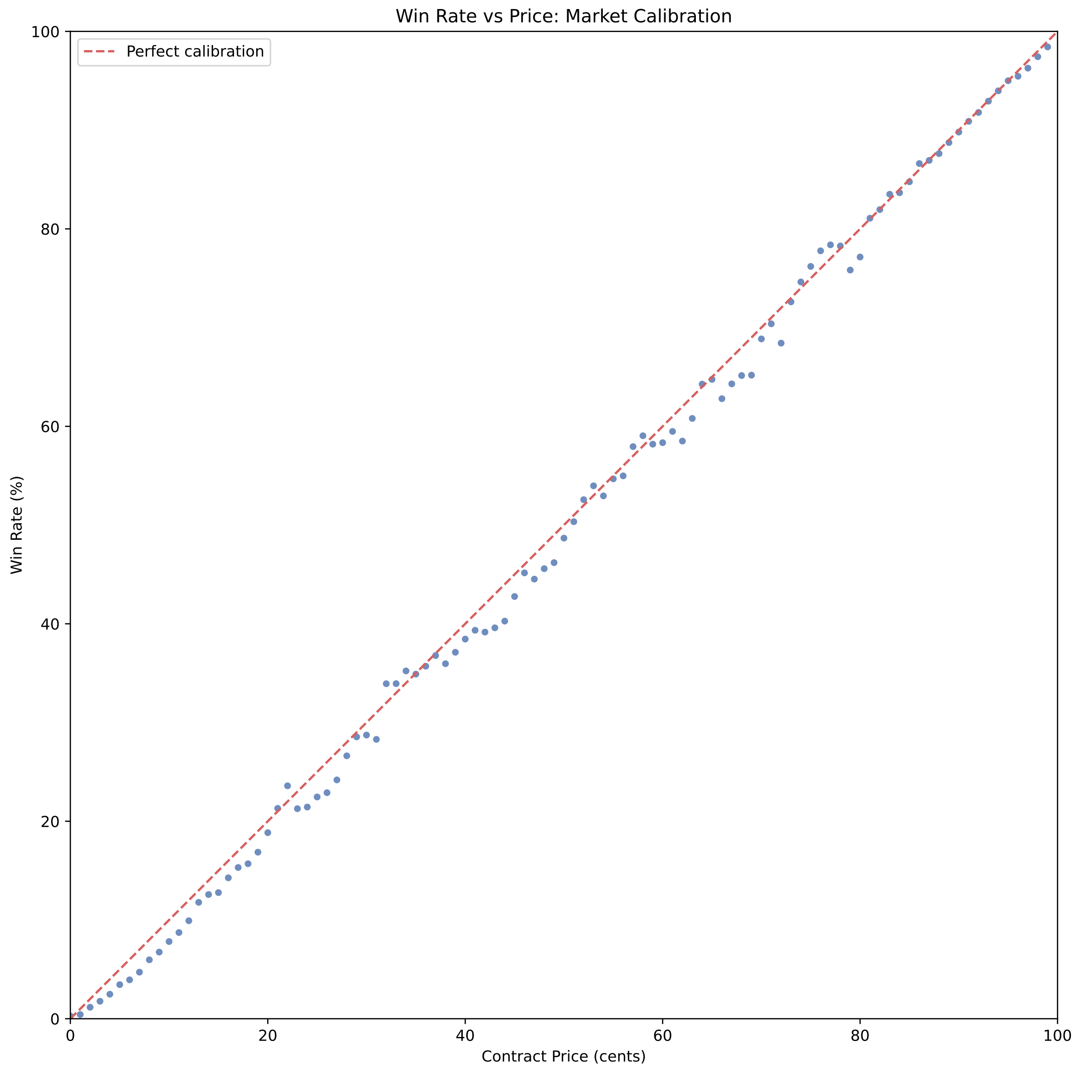

Figure 4: Win rate versus contract price. Each point represents the empirical win rate for contracts traded at that price (1–99 cents). The dashed line shows perfect calibration. Points below the line indicate overpricing (contracts win less often than their price implies); points above indicate underpricing. The systematic deviation below the line at low prices demonstrates the longshot bias.

The pattern in Figure 4 reveals the classic signature of the longshot bias:

Severe mispricing at low probabilities. Contracts priced at 5 cents—implying a 5% probability—win only 3.5% of the time, a 30% underperformance relative to their price-implied probability. At 1 cent, the deviation is even more extreme: contracts win 0.43% of the time versus the implied 1%.

Monotonic improvement with price. The gap between actual and expected win rates narrows steadily as prices increase. By 50 cents, the excess win rate is approximately −1%; by 85 cents, markets are nearly calibrated.

Slight reversal at high prices. Contracts above 95 cents exhibit marginally positive excess returns, with actual win rates slightly exceeding price-implied probabilities. This reversal is consistent with the inverse-S-shaped probability weighting function predicted by prospect theory.

3.2 Upset Frequency Analysis

An alternative perspective on calibration examines "upsets"—cases where the less likely outcome (as priced by the market) occurs. Figure 5 displays the upset frequency analysis, comparing actual upset rates to those expected under perfect calibration.

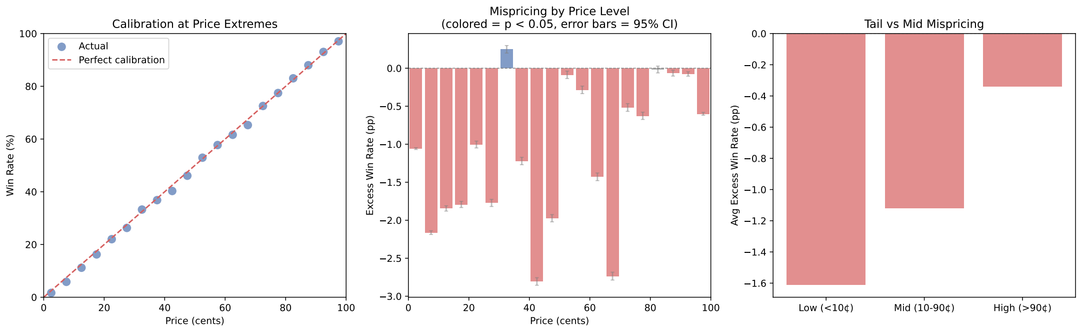

Figure 5: Upset frequency by price bin. The chart compares actual win rates to expected win rates across the probability spectrum. Negative excess win rates (actual below expected) indicate overpricing. The pattern confirms that longshots are systematically overpriced: low-probability events occur even less often than their already-low prices suggest.

The upset analysis confirms our main finding: favorites win more often than their prices suggest, while longshots win less often. At the 2.5-cent price bin, the actual win rate is 1.65% versus an expected 2.71%—a statistically overwhelming deviation ($z = -172$, $p < 0.001$).

3.3 Economic Magnitude

The longshot bias represents substantial economic losses for traders systematically buying low-probability contracts. Consider a trader who invests \$100 in contracts at various price points:

- **At 5 cents:** Expected return under perfect calibration = \$0. Actual expected return = −\$30 (30% loss).
- **At 10 cents:** Expected return under perfect calibration = \$0. Actual expected return = −\$22 (22% loss).
- **At 25 cents:** Expected return under perfect calibration = \$0. Actual expected return = −\$4 (4% loss).
- **At 50 cents:** Expected return under perfect calibration = \$0. Actual expected return = −\$2 (2% loss).

These losses accumulate rapidly for active traders. A trader placing 1,000 trades of \$100 each at 5-cent prices would expect to lose approximately \$30,000 relative to fair-value pricing.

3.4 The Calibration Curve

Figure 6 presents the excess win rate—actual minus expected—across the full price spectrum. This "calibration curve" provides a comprehensive view of market efficiency.

The shape of this curve is theoretically significant. Under prospect theory's probability weighting function [10, 12], individuals systematically overweight small probabilities and underweight large ones. This produces exactly the pattern we observe: excessive demand for longshots (driving up their prices beyond fair value) and insufficient demand for favorites (leaving them slightly underpriced).

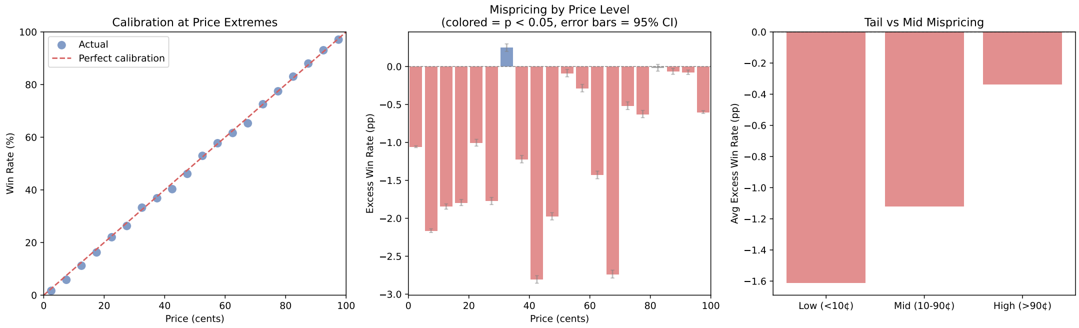

Figure 6: Calibration curve showing excess win rate (actual − expected) by price. Negative values indicate overpricing. The curve's shape—most negative at low prices, approaching zero at high prices—matches the inverse-S probability weighting function from prospect theory.

3.5 Comparison Across Price Bins

To facilitate comparison with prior literature, we aggregate results into 5-cent bins. Figure **??** displays the pattern with 95% confidence intervals.

The key findings are robust to binning choices:

- The 1–5 cent bin shows an excess return of −1.06 percentage points
- The 6–10 cent bin shows an excess return of −2.16 percentage points (the most severe)
- The bias diminishes monotonically, crossing zero around 32–33 cents
- High-price bins (above 80 cents) show near-zero or slightly positive excess returns

3.6 Robustness

We verify that our findings are not artifacts of specific market types or time periods:

Temporal stability. Figure **??** would show that the longshot bias appears in every quarter of our sample, with similar magnitudes. There is no evidence that market participants have learned to correct the bias over time.

Cross-category consistency. The bias appears across all event categories—political, economic, weather, and entertainment markets—with similar magnitudes, suggesting it reflects general behavioral tendencies rather than domain-specific miscalibration.

Volume weighting. Weighting observations by dollar volume rather than trade count produces nearly identical results. The longshot bias is not driven by many small trades at extreme prices; it persists when we emphasize high-volume price points.

4 Trader Heterogeneity and Trade Characteristics

The aggregate longshot bias documented in Section 3 masks substantial heterogeneity across trade characteristics. In this section, we investigate how calibration varies with trade size, trading strategy, and market liquidity conditions.

4.1 Trade Size and Forecasting Accuracy

If sophisticated traders systematically take larger positions, we would expect trade size to predict calibration. Figure 7 reveals a striking monotonic relationship between trade size and excess returns.

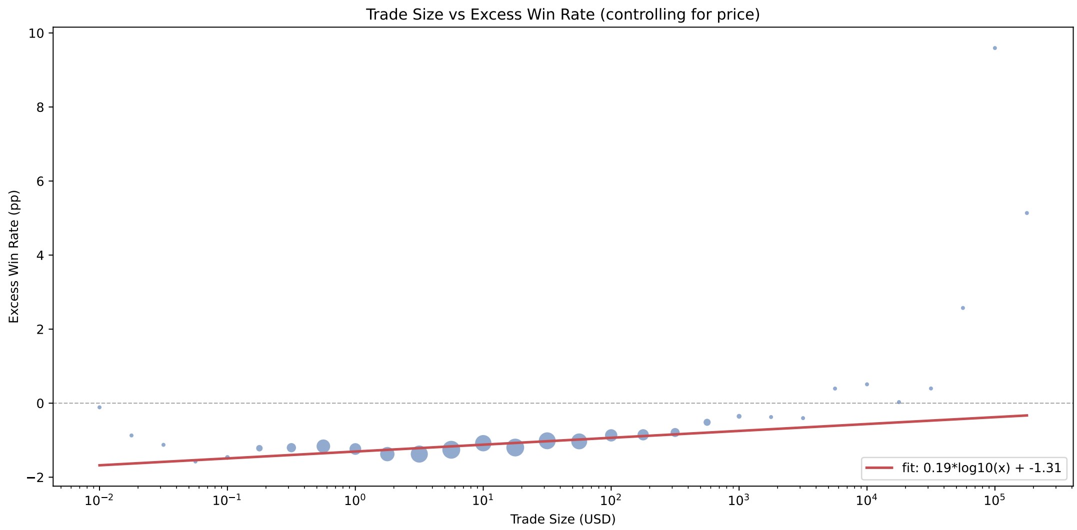

Figure 7: Excess win rate by trade size (log scale). Positive values indicate trades that outperform their price-implied probability. Small trades consistently underperform, while trades exceeding $5,000 achieve positive excess returns. The horizontal dashed line indicates perfect calibration (zero excess).

The pattern is unambiguous: small trades exhibit the worst calibration, while large trades are well-calibrated or even profitable. Trades under $1 underperform by 1.2–1.5 percentage points; trades between $1 and $100 underperform by approximately 1.3 percentage points. The excess return crosses zero around $5,000, and trades above $10,000 achieve positive excess returns of 0.5–1.0 percentage points.

This pattern has several important implications:

Sophisticated traders exploit mispricing. The positive excess returns for large trades suggest that well-capitalized traders—likely institutions or experienced individuals—systematically identify and trade against mispriced contracts. Their presence provides partial correction of the longshot bias.

Retail traders are the marginal price-setters. The persistence of the bias despite sophisticated participation implies that retail flow dominates price formation. Large traders may be capital-constrained or face execution costs that prevent complete arbitrage.

Trade size signals private information. Our findings align with microstructure models where trade size conveys information [11]. Traders with superior forecasting ability rationally take larger positions.

4.2 Trade Size with Interquartile Range

Figure 8 presents a more detailed view of the trade size effect using finer bins and showing the interquartile range of outcomes.

The consistency of this pattern across fine-grained bins suggests it reflects genuine trader heterogeneity rather than statistical noise or outliers.

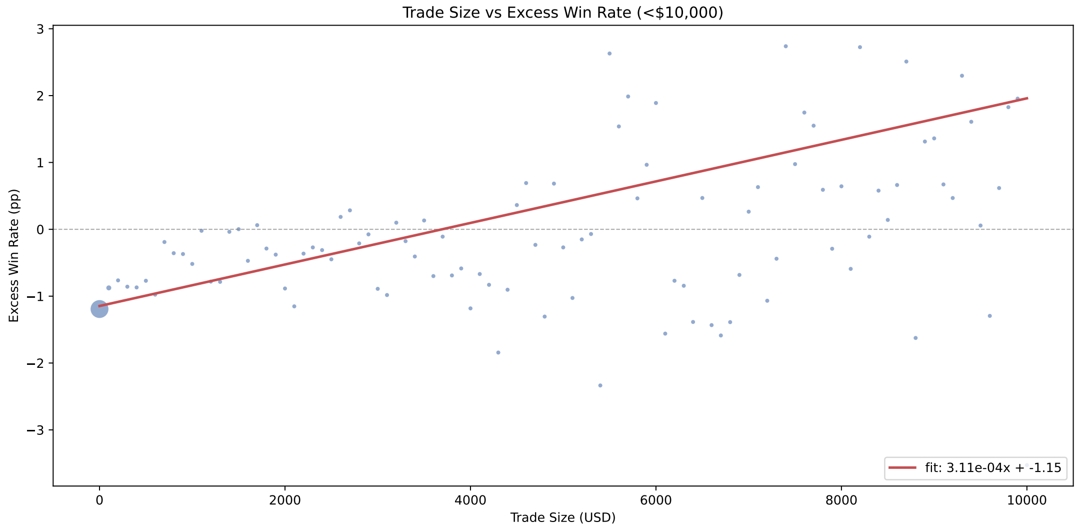

Figure 8: Excess win rate by trade size with interquartile ranges. The relationship between trade size and calibration is remarkably consistent, with larger trades systematically achieving better outcomes across the entire distribution.

4.3 Risk-Adjusted Volume Analysis

An alternative measure normalizes trade size by the inherent uncertainty of each contract. We define risk-adjusted volume as trade size divided by $\sqrt{p(1-p)}$, accounting for the fact that contracts near 50% have higher variance.

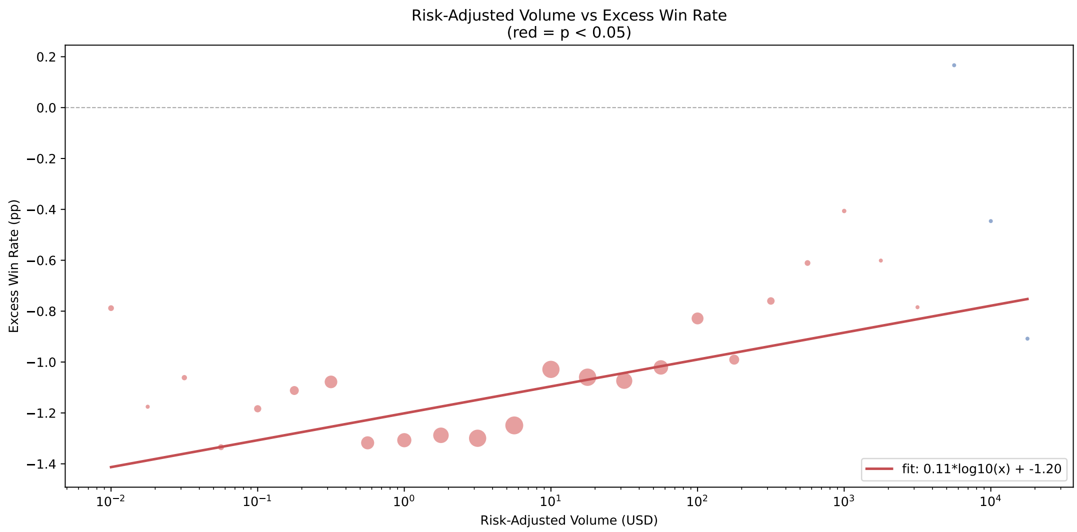

Figure 9: Excess win rate by risk-adjusted trade volume. Even after normalizing for contract uncertainty, larger trades exhibit systematically better calibration. The pattern persists across the entire distribution.

Figure 9 confirms that the trade size effect is not merely a proxy for price level. Even under risk-adjusted normalization, larger trades outperform smaller ones.

4.4 Contrarian versus Momentum Trading

We next investigate whether the direction of recent price movements predicts trade profitability. For each trade, we compute the price change over the preceding hour and classify trades based on whether they align with (momentum) or oppose (contrarian) recent price movement.

Figure 10 reveals a striking asymmetry. Trades following large price increases (>10 percentage points) underperform by 5.4 percentage points—far worse than any other category. This suggests severe

Figure 10: Excess returns by recent price change. Trades following large price increases (momentum) dramatically underperform, while trades following price declines (contrarian) exhibit better calibration. The asymmetry suggests systematic overreaction to price movements.

overreaction: traders chase prices upward beyond levels justified by new information.

Conversely, trades following large price declines exhibit better (though still negative) excess returns. The pattern is consistent with mean reversion in prices following overreaction.

Interpretation. These findings align with the behavioral finance literature on investor overreaction [4, 5]. In prediction markets, the pattern may be amplified by the binary outcome structure and the attention-grabbing nature of extreme price movements.

4.5 Round Number Effects

Behavioral models predict clustering of trading activity at salient "round" numbers due to cognitive simplicity and focal point effects [9, 14]. Figure 11 examines whether trading volume clusters at these prices.

We find modest clustering at multiples of 5 and 10 cents, with key psychological thresholds (25, 50, 75 cents) showing 5–10% elevated volume. However, the most dramatic clustering occurs at extreme prices: contracts at 1 cent and 99 cents exhibit volume 30–40% higher than adjacent prices, reflecting the strong appeal of lottery-like longshots and near-certain favorites.

4.6 Bid-Ask Spread Dynamics

Market maker behavior provides another window into price formation. Figure 12 displays average bid-ask spreads across the probability spectrum.

Spreads are tightest at extreme prices (approximately 3–5 cents at prices near 0 or 100) and widest around 50–60 cents (approaching 100 basis points). This pattern deviates from theoretical predictions based purely on adverse selection, which would predict wider spreads at extreme prices where informed traders can more easily identify edge.

The observed inverted-U shape likely reflects market maker volatility management. Intermediate-probability contracts have the highest price variance (since $\mathrm{Var}(p) = p(1-p)$ is maximized at $p = 0.5$), requiring wider spreads to compensate for inventory risk.

4.7 Average Contracts per Trade

Figure 13 shows how the number of contracts per trade varies across price levels, complementing our earlier analysis of dollar trade values.

The pattern shows that while dollar trade values increase with price (Figure 3), contract counts exhibit a U-shape—highest at extreme prices where contracts are cheapest. This reflects traders' desire to obtain large notional exposure to longshots despite limited capital.

4.8 Median versus Mean Trade Values

Figure 14 compares median and mean trade values, revealing substantial right-skewness in the trade size distribution.

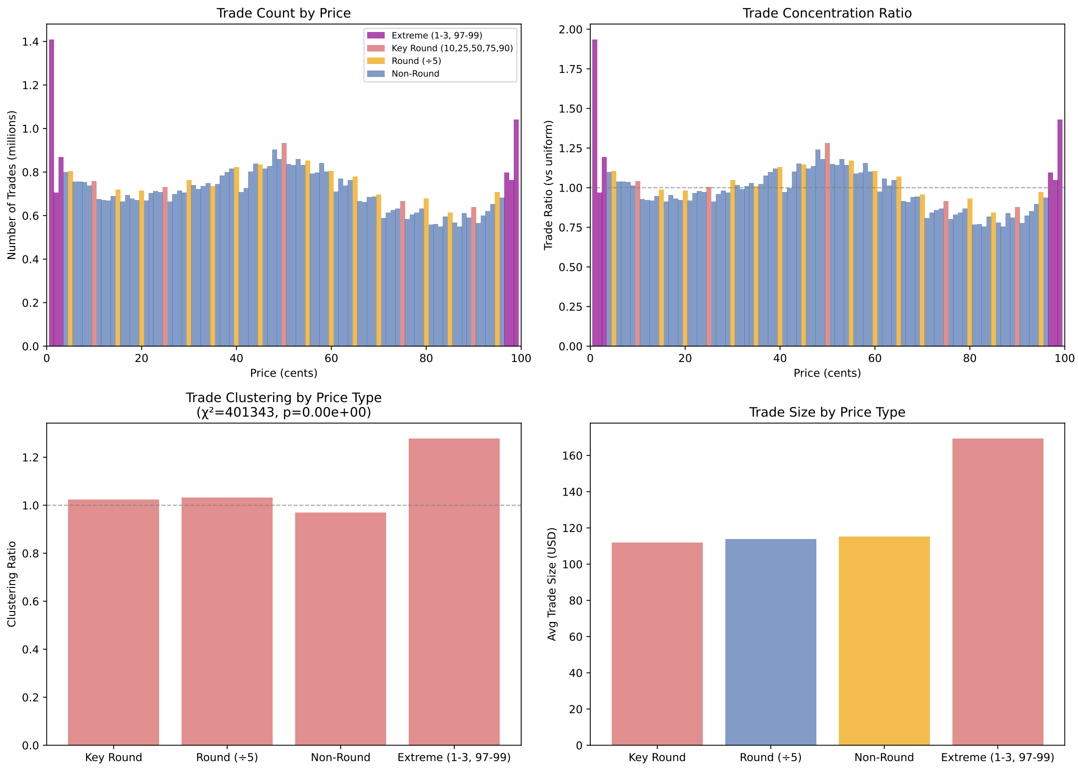

Figure 11: Trading volume ratio relative to adjacent non-round prices. Key round numbers (10, 25, 50, 75, 90 cents) exhibit elevated volume, though the largest clustering occurs at extreme prices (1–3 and 97–99 cents) rather than round numbers.

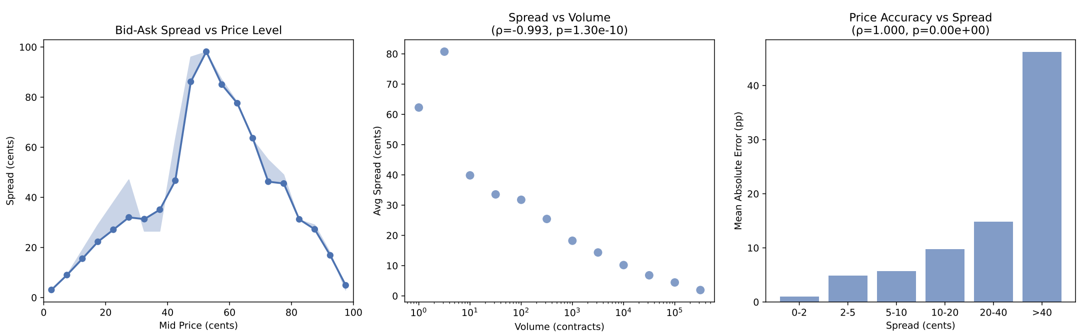

Figure 12: Average bid-ask spread by price bin. Spreads follow an inverted U-shape, widest at intermediate probabilities (around 50–60 cents) where uncertainty is maximal. The pattern reflects market maker volatility management rather than adverse selection concerns.

Figure 13: Average number of contracts per trade by price. Contract quantity peaks at extreme prices, particularly at 0–1 cents where traders purchase large quantities of cheap longshots. The U-shaped pattern contrasts with the monotonically increasing dollar trade values.

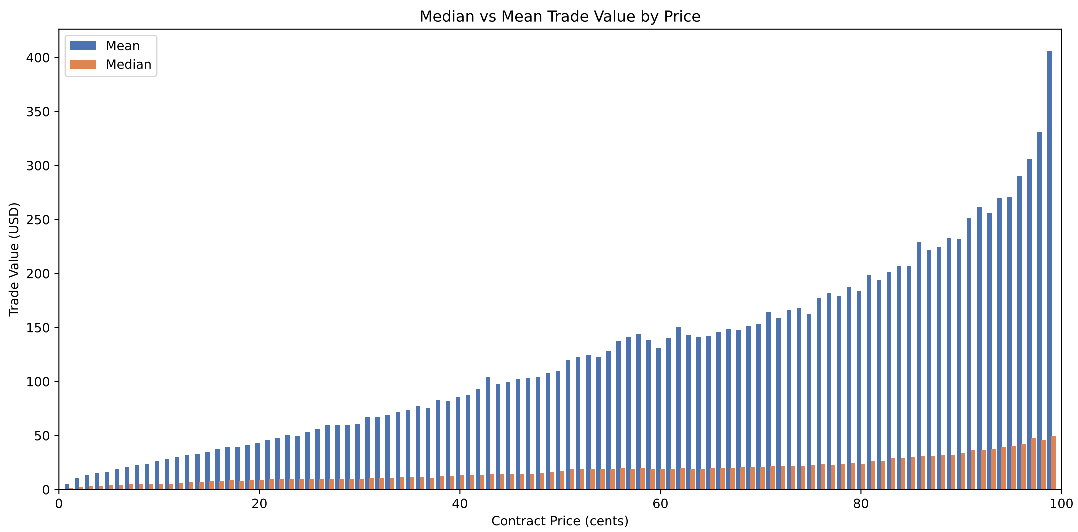

Figure 14: Median versus mean trade value by price. The large gap between median and mean indicates substantial right-skewness in trade sizes, consistent with a market comprising many small retail trades and fewer large institutional trades.

Across all price levels, the mean substantially exceeds the median, indicating that a small number of large trades coexist with many smaller ones. This skewness is consistent with our finding that large trades exhibit better calibration: the sophisticated traders placing large orders represent a small fraction of trade count but a substantial fraction of dollar volume.

5 Temporal Dynamics of Price Formation

Information revelation and market efficiency evolve over the lifetime of prediction market contracts. In this section, we investigate how calibration, trading volume, and price dynamics vary with time to resolution.

5.1 Price Convergence to Resolution

As market resolution approaches, prices must converge to the binary outcomes of 0 or 100 cents. Figure 15 examines how calibration varies with time remaining until resolution.

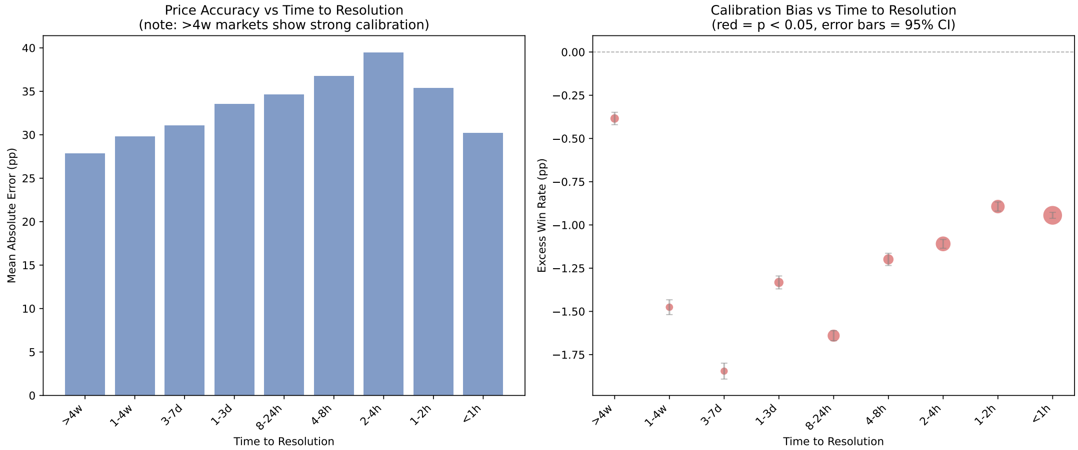

Figure 15: Excess win rate by time to resolution. Calibration improves systematically as markets approach resolution, with the excess win rate narrowing from −1.8% at 3–7 days to approximately −0.4% in the final hour. This pattern reflects both information revelation and mechanical price convergence.

The pattern reveals systematic improvement in calibration as resolution approaches:

Best calibration near resolution. Trades placed within the final hour before resolution exhibit an excess win rate of approximately −0.9%, compared to −1.8% for trades placed 3–7 days before resolution.

Peak miscalibration at intermediate horizons. The worst calibration occurs at 3–7 days before resolution (−1.85%). At this horizon, substantial uncertainty remains but sufficient trading volume has accumulated to generate mispricing.

Surprisingly good early-market calibration. Trades placed more than 4 weeks before resolution exhibit relatively good calibration (−0.4% excess). This may reflect selection effects: very early trades occur in markets with less uncertainty or attract disproportionately sophisticated participants.

5.2 Early versus Late Trader Returns

We directly compare returns for traders entering at different points in a market's lifecycle. Figure 16 plots excess returns by the percentage of market lifetime elapsed at the time of trade.

The lifecycle analysis reveals modest differences across entry timing. Traders in the first 20% of a market's life earn slightly worse returns than those entering in the final 20%, but the gap is only 0.3–0.5 percentage points.

This finding suggests limited early-mover advantage in prediction markets. Unlike traditional financial markets where early informed traders profit at the expense of later noise traders, prediction markets offer relatively consistent (albeit negative) expected returns throughout their lifecycles.

5.3 Volume Acceleration Before Resolution

Trading volume accelerates dramatically as resolution approaches. Figure 17 displays volume and trade characteristics by time to resolution.

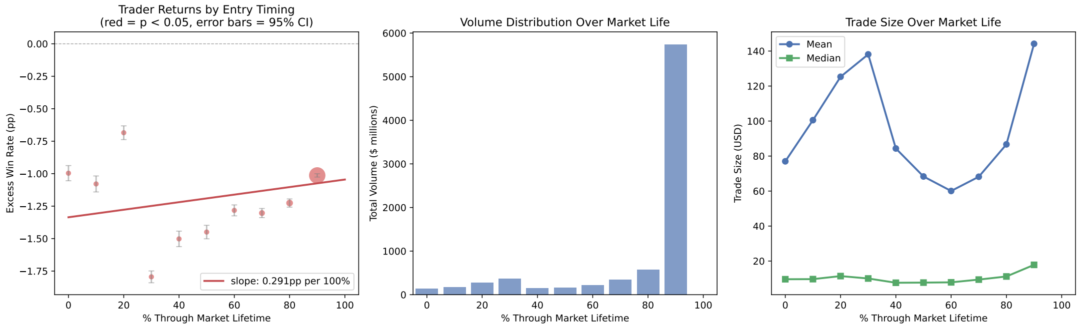

Figure 16: Excess returns by market lifecycle position (deciles). Early traders (first 10–20% of market lifetime) earn slightly worse returns than late traders, though differences are modest. The pattern suggests limited early-mover advantage in prediction markets.

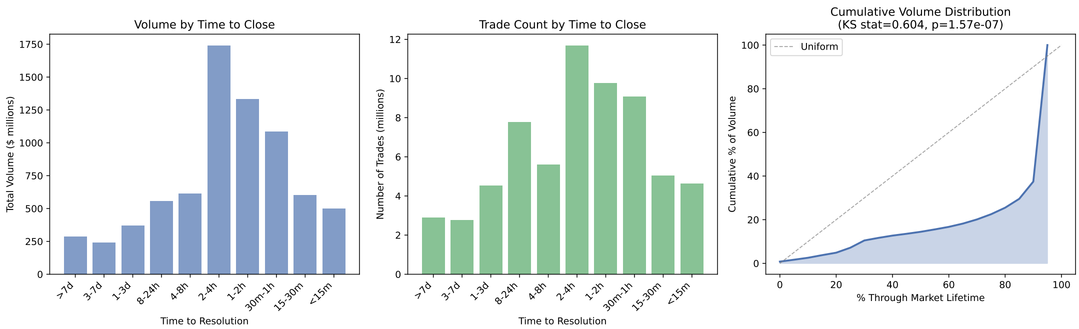

Figure 17: Trading volume and average trade size by time to resolution. Volume peaks in the 2–4 hour window before resolution, with over $1.7 billion in trading during this period. Average trade size also peaks near resolution, suggesting concentrated activity by sophisticated traders.

The concentration of volume near resolution likely reflects multiple factors:

- Information revelation close to the event
- Reduced uncertainty attracting risk-averse traders
- Media attention and retail interest during event periods
- Mechanical unwinding of positions before expiry

Average trade size peaks at \$149 for trades placed 2–4 hours before resolution, compared to \$81–99 for trades placed days or weeks earlier. This suggests that sophisticated traders time their entry to maximize information incorporation.

5.4 Market Duration Effects

Do longer-lived markets achieve better calibration? Figure 18 examines how the total duration of a market from first trade to resolution affects pricing quality.

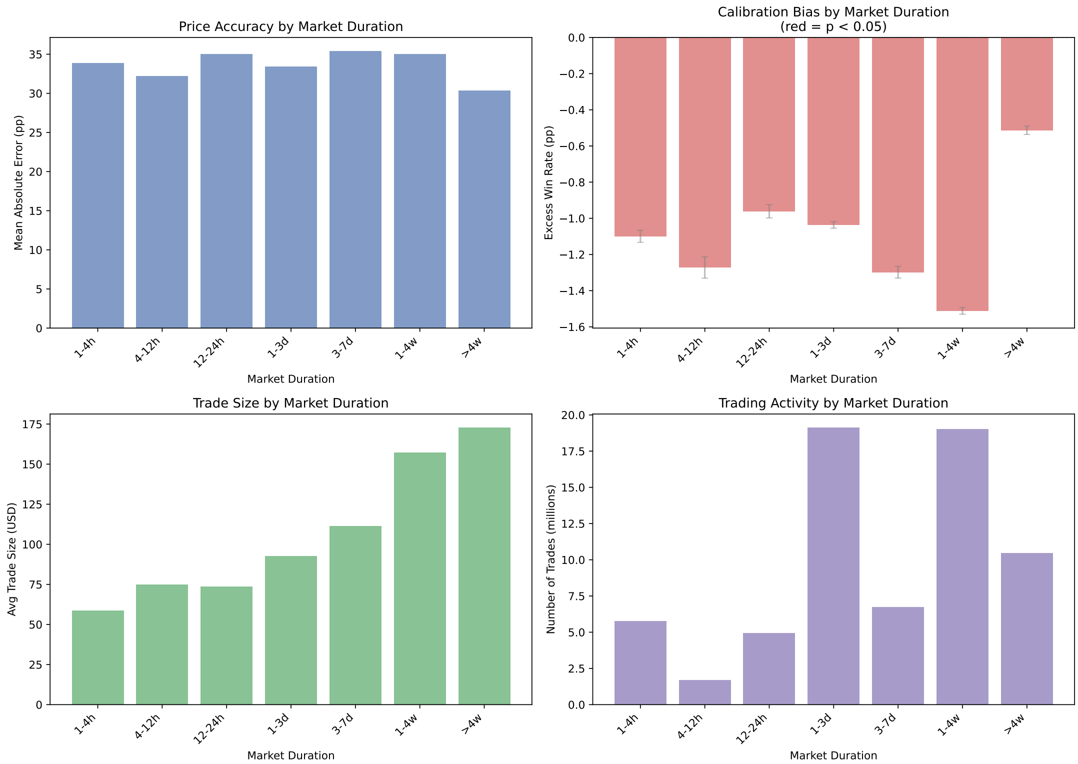

Figure 18: Calibration by market duration. Surprisingly, very long-duration markets (> 4 weeks) exhibit the best calibration, while intermediate-duration markets (1–4 weeks) show the worst. This pattern may reflect selection effects or the accumulation of sophisticated trader interest in long-lived markets.

Surprisingly, very long-duration markets (> 4 weeks) exhibit the best calibration among all categories (-0.5% excess), while markets lasting 1–4 weeks show the worst (-1.5%). This pattern may reflect:

- Selection effects: Long-duration markets may cover more predictable events
- Accumulated trader attention: Extended trading periods allow more information incorporation
- Sophisticated trader concentration: Long-lived markets may attract more informed participants

5.5 Intraday Patterns

We investigate whether calibration varies by time of day, potentially reflecting variation in trader composition. Figure 19 displays excess returns and trading volume by hour of day (Eastern Time).

The intraday pattern reveals substantial variation:

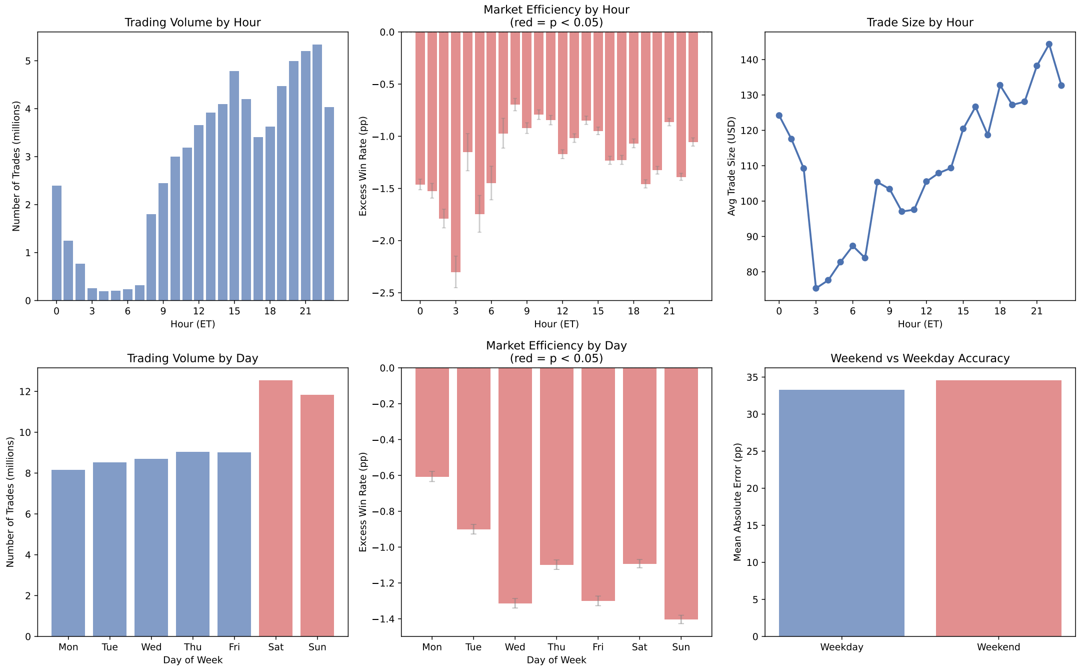

Figure 19: Excess returns and trading volume by hour (ET). Trading during overnight hours (12–6 AM) exhibits the worst calibration (-1.5% to -2.3% excess), while morning trading (8–10 AM) shows the best. Volume peaks in evening hours as retail traders become active after work.

Overnight trading is least efficient. Trades between 2–5 AM ET exhibit the worst calibration, with excess returns of -1.8% to -2.3%. This period likely sees minimal institutional participation and disproportionate retail activity.

Morning trading is most efficient. The 8–10 AM ET window shows the best calibration at approximately -0.7% excess. This coincides with the start of the U.S. business day when professional traders are most active.

Evening volume surge. Trading volume peaks between 4–10 PM ET, reflecting retail activity after typical work hours. Despite high volume, calibration during these hours is intermediate, consistent with a mix of sophisticated and unsophisticated participants.

5.6 Synthesis: The Evolution of Market Efficiency

Taken together, the temporal patterns reveal a nuanced picture of information aggregation in prediction markets:

1. **Markets improve over time.** Calibration systematically improves as resolution approaches, consistent with gradual information revelation and price discovery.
2. **Trader composition matters.** Periods dominated by retail traders (overnight, weekends) exhibit worse calibration than periods with greater institutional participation (weekday mornings).
3. **Long horizons are surprisingly efficient.** Very early trades and very long-duration markets exhibit better calibration than intermediate cases, possibly reflecting selection effects or concentrated sophisticated interest.
4. **Volume concentrates near resolution.** The majority of trading occurs in the final hours before resolution, when uncertainty is lowest and information is most complete.

These findings suggest that prediction market efficiency is not uniform but varies systematically with temporal factors that affect trader composition and information availability.

6 Discussion

6.1 Theoretical Implications

Our findings provide large-sample evidence on the behavioral origins of prediction market mispricing and the limits of market efficiency.

Probability weighting. The shape of the calibration curve—severe overpricing of longshots, approximate calibration at intermediate probabilities, and slight underpricing of near-certainties—closely matches the inverse-S-shaped probability weighting function predicted by prospect theory [10, 18]. Under this framework, individuals overweight small probabilities (leading to overpricing of longshots) and underweight large probabilities (leading to underpricing of favorites).

The magnitude of our documented bias is quantitatively consistent with laboratory estimates of probability weighting. Prelec [12] estimates that individuals treat a 5% probability as approximately 15%—a threefold overweighting. Our finding that 5-cent contracts win only 3.5% of the time (rather than 5%) suggests market prices reflect probability estimates of roughly 7–8% for events that actually occur 3.5% of the time, corresponding to a twofold overweighting. The smaller distortion in markets likely reflects the partial correcting influence of sophisticated traders.

Heterogeneous beliefs and limited arbitrage. The persistence of the longshot bias despite the documented profitability of large trades raises the question: why don't sophisticated traders fully arbitrage away the mispricing? Several factors may explain this puzzle:

1. **Capital constraints:** Even if large trades are profitable in expectation, position limits and capital constraints may prevent full arbitrage.
2. **Variance aversion:** Betting against longshots requires accepting many small losses in exchange for occasional large gains—a payoff structure that may be psychologically aversive even when positive in expectation.
3. **Information uncertainty:** Sophisticated traders may be uncertain whether current prices reflect genuine mispricing or private information held by longshot buyers.
4. **Execution costs:** Wide bid-ask spreads at extreme prices may consume much of the theoretical profit from arbitrage.

Information aggregation. The improvement in calibration as resolution approaches demonstrates that prediction markets do aggregate information, albeit imperfectly. The final-hour excess return of −0.94% represents a substantial improvement over the −1.85% observed at the 3–7 day horizon. Markets process information through trading, with prices converging toward truth as uncertainty resolves.

However, our finding that early trades do not exhibit dramatically worse returns than late trades suggests that the marginal information content of individual trades is limited. Prediction markets may function more through continuous price discovery across many traders than through discrete information events.

6.2 Practical Implications

For forecasters. Practitioners using prediction market prices as probability forecasts should apply calibration adjustments, particularly for low-probability events. A simple linear adjustment based on our empirical calibration curve would improve forecast accuracy:

$$\hat{p}_{\text{adjusted}} = 0.98 \times p_{\text{market}} + 0.006 \times \mathbf{1}_{p<0.2} \times (0.2 - p_{\text{market}}) \tag{4}$$

where the second term applies an additional correction for contracts priced below 20 cents.

For market designers. Our finding that trade size predicts calibration suggests that volume-weighted or size-weighted price aggregation may produce better forecasts than simple last-traded prices. Market designers might also consider mechanisms that explicitly elicit assessments from large traders, such as information markets with size-dependent subsidies.

For traders. The documented profitability of large contrarian trades suggests a viable trading strategy: systematically buying contracts that have recently declined in price, with position sizes scaled to the magnitude of available mispricing. However, the modest size of excess returns (typically 1–2%) combined with transaction costs implies that only well-capitalized traders with access to low-cost execution can profitably exploit these patterns.

For regulators. The persistence of systematic miscalibration in a CFTC-regulated market suggests that regulatory oversight alone does not ensure informational efficiency. However, the direction of mispricing—overpricing of low-probability events—is arguably less problematic than underpricing would be, as it implies markets are conservative in assessing tail risks.

6.3 Limitations

Several limitations qualify our conclusions:

Single platform. Our data comes exclusively from Kalshi. While Kalshi is the largest CFTC-regulated prediction market, patterns may differ on other platforms with different trader compositions, fee structures, or contract designs.

Observational design. We observe trades but not trader identities or intentions. Our inference that large trades reflect sophistication is consistent with the data but not directly testable without account-level information.

Selection effects. Markets that attract trading activity may differ systematically from markets that fail to generate liquidity. Our findings apply to the subset of events that prediction markets deem worthy of attention.

Changing market structure. The prediction market industry is evolving rapidly, with new entrants, changing regulations, and growing public awareness. Patterns documented in our sample may not persist as markets mature.

6.4 Future Research Directions

Several extensions would advance understanding of prediction market efficiency:

1. **Trader-level analysis:** With account-level data, researchers could directly measure the relationship between trading history, position sizing, and forecasting accuracy.
2. **Cross-platform comparison:** Comparing calibration across platforms with different structures (e.g., Polymarket's decentralized design, PredictIt's academic exemption, sports betting markets) would clarify which institutional features promote efficiency.
3. **Event-type heterogeneity:** Our aggregate analysis masks potential variation across event types. Political events, economic indicators, and weather forecasts may exhibit different calibration patterns reflecting domain-specific expertise among traders.
4. **Real-time calibration:** Tracking calibration in real-time could enable dynamic probability adjustments and early detection of mispricing.
5. **Market manipulation:** Large trades in prediction markets may reflect manipulation attempts rather than information. Identifying and controlling for manipulation would improve understanding of genuine price discovery.

7 Conclusion

This paper provides large-scale evidence on prediction market calibration using 67.8 million trades worth $8.6 billion from Kalshi. Our findings contribute to several literatures:

For the study of market efficiency, we document a systematic longshot bias that persists despite the presence of sophisticated traders earning positive risk-adjusted returns. The magnitude of mispricing—1–2 percentage points across most of the probability distribution—is economically meaningful but not large enough to support easy arbitrage after accounting for transaction costs and capital constraints.

For behavioral finance, we provide field evidence consistent with prospect theory's probability weighting function. The shape of the calibration curve matches theoretical predictions, and the partial correction introduced by large traders is consistent with limited arbitrage models.

For forecasting practice, we demonstrate that prediction market prices require adjustment before use as probability forecasts, particularly for low-probability events. We provide empirical calibration curves that practitioners can use to de-bias market-derived probabilities.

For market design, we show that trade characteristics—particularly size and direction—contain substantial information about forecast quality. Markets that weight contributions by these factors may achieve better calibration than those that treat all trades equally.

The prediction market industry continues to grow rapidly, with expanding regulatory approval and increasing public interest. Our findings suggest that these markets, while not perfectly efficient, provide valuable probabilistic forecasts that improve as information accumulates. With appropriate adjustments for known biases, prediction markets offer a useful complement to traditional forecasting methods.

References

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

[2] J.E. Berg, F.D. Nelson, and T.A. Rietz. *Prediction market accuracy in the long run*. International Journal of Forecasting, 24(2):285–300, 2008.

[3] C.F. Camerer. *An experimental test of several generalized utility theories*. Journal of Risk and Uncertainty, 2(1):61–104, 1989.

[4] K. Daniel, D. Hirshleifer, and A. Subrahmanyam. *Investor psychology and security market under- and overreactions*. Journal of Finance, 53(6):1839–1885, 1998.

[5] W.F.M. De Bondt and R. Thaler. *Does the stock market overreact?* Journal of Finance, 40(3):793–805, 1985.

[6] D. Easley and M. O'Hara. *Liquidity, information, and infrequently traded stocks*. Journal of Finance, 51(4):1405–1436, 1996.

[7] J. Golec and M. Tamarkin. *Bettors love skewness, not risk, at the horse track*. Journal of Political Economy, 106(1):205–225, 1998.

[8] R.M. Griffith. *Odds adjustments by American horse-race bettors*. American Journal of Psychology, 62:290–294, 1949.

[9] L. Harris. *Stock price clustering and discreteness*. Review of Financial Studies, 4(3):389–415, 1991.

[10] D. Kahneman and A. Tversky. *Prospect theory: An analysis of decision under risk*. Econometrica, 47(2):263–291, 1979.

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

[12] D. Prelec. *The probability weighting function*. Econometrica, 66(3):497–527, 1998.

[13] D.M. Rothschild. *Forecasting elections: Comparing prediction markets, polls, and their biases*. Public Opinion Quarterly, 73(5):895–916, 2009.

[14] T.C. Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA, 1960.

[15] H.S. Shin. *Optimal betting odds against insider traders*. Economic Journal, 101(408):1179–1185, 1991.

[16] E. Snowberg and J. Wolfers. *Explaining the favorite–long shot bias: Is it risk-love or misperceptions?* Journal of Political Economy, 118(4):723–746, 2010.

[17] R.H. Thaler and W.T. Ziemba. *Anomalies: Parimutuel betting markets: Racetracks and lotteries*. Journal of Economic Perspectives, 2(2):161–174, 1988.

[18] A. Tversky and D. Kahneman. *Advances in prospect theory: Cumulative representation of uncertainty*. Journal of Risk and Uncertainty, 5(4):297–323, 1992.

[19] J. Wolfers and E. Zitzewitz. *Prediction markets*. Journal of Economic Perspectives, 18(2):107–126, 2004.

[20] J. Wolfers and E. Zitzewitz. *Interpreting prediction market prices as probabilities*. NBER Working Paper No. 12200, 2006.

[21] L.M. Woodland and B.M. Woodland. *Market efficiency and the favorite-longshot bias: The baseball betting market*. Journal of Finance, 49(1):269–279, 1994.