TradeTruth: Closed-Form Ambiguity-Averse Robust Kelly, Net-of-Friction Cross-Venue Arbitrage, and Minimax SIPS_∞ Auditing for Retail Trading and Prediction Markets

Fabien Pierret

Independent Researcher

ORCID: 0009-0001-1223-2170 | poly.wallstreetdebug.com | github.com/fabien-pierret/tradetruth-polymarket

Preprint (camera-ready manuscript) — November 27, 2026

Abstract. Retail traders and event-contract participants lose billions annually to two tightly coupled mathematical illusions: (i) cherry-picked trading backtests, martingale equity curves, and weighted-average rating badges that mask data dredging and blow-up tail risk; and (ii) prediction-market mispricing traps on Polymarket and Kalshi—overpriced extreme longshots, fee-trap headline arbitrage spreads, whale-dominated orderbooks, and overconfident full-Kelly bet sizing. We present TradeTruth, a unified formal-to-consumer framework governed by the *Law of the Mathematical Iceberg* (zero PhD jargon on the end-user screen, 100% machine-checked algebraic rigor underneath). We establish four theorems whose deterministic ordered-field and structural-induction cores are **machine-checked in standalone Lean 4.36.0** (`lean/TradeTruthCore.lean`: 36 public theorems, 0 `sorry`, 0 warnings, zero external dependencies, standard kernel axioms `[propext, Classical.choice, Quot.sound]` exclusively, audited by two independent referees): **(1) Ambiguity-Averse Interval Robust Kelly & Fractional-Kelly Duality (Theorem 1):** for binary contracts priced at all-in cost $c_{\mathrm{eff}} \in (0,1)$ under calibration interval $p \in [p_{\mathrm{low}}, p_{\mathrm{high}}]$ ($p_{\mathrm{low}} = \hat{p} - \varepsilon$), the minimax optimal stake admits the closed form $f^*_{\mathrm{rob}} = \max(0, (p_{\mathrm{low}} - c_{\mathrm{eff}})/(1 - c_{\mathrm{eff}}))$, satisfies the exact Cover–Thomas martingale identity $\mathbb{E}_{p_{\mathrm{low}}}[W(f)/W(f^*_{\mathrm{rob}})] = 1$, and obeys the exact algebraic fractional-Kelly duality $f^*_{\mathrm{rob}} = (1 - \varepsilon/(\hat{p} - c_{\mathrm{eff}}))\, f^*_{\mathrm{nom}}$; on favorites ($c_{\mathrm{eff}} > 1/2$), Naive Full Kelly enters asymptotic exponential ruin even when less than half of perceived edge is calibration noise ($\varepsilon^* = 4.52\%$ at $c = 80¢, \hat{p} = 90\%$, causing 80.1% wipeout vs. 0.15% under TradeTruth at $\varepsilon = 8\%$). **(2) Cross-Platform Net-of-Friction Arbitrage Boundary (Theorem 2):** we prove the exact state-dependent two-leg arbitrage IFF condition ($C_{\min} < 1$), characterize the Fee-Trap Mirage Zone, derive Kalshi's exact quadratic taker fee peak $\phi_K(p) \le \gamma/4$ (1.75¢ at $p = 50¢$ for $\gamma = 0.07$), and prove the adverse oracle-mismatch ruin threshold $\eta_{\mathrm{settle}} < 1 - C_{\min}$; across 5,000 headline cross-venue alerts, 82.96% fail net-of-friction or oracle checks (73.90% fee traps, 9.06% oracle-vetoed). **(3) Minimax SIPS_∞ + Heavy-Tail Q-DSR_ν Backtest Bullshit Detector (Theorem 3):** across 1,000 strategies spanning four scam cohorts and clean alpha controls, linear averaging catches 0.0% and Gaussian DSR catches 14.5%, whereas TradeTruth catches 100.0% of scams with 0.0% false rejection. **(4) Live L2 Whale-Wall Concentration & L-Tier Book-Walk Kelly Shrinkage (Theorem 4):** audited across 18 live November 27, 2026 Polymarket CLOB and Kalshi v2 orderbooks (18/18 satisfying $1/L \le \mathrm{HHI} \le W_1 \le 1$ and $L_{\mathrm{eff}} \le 1/W_1^2$; median $W_1 = 62.2\%$, median $L_{\mathrm{eff}} = 2.35$). Under empirical favorite-longshot bias (1,000 seeds ×150 bets), TradeTruth lifts median terminal wealth to 1.29× (0.0% ruin) vs. 0.16× (Full Kelly) and 0.66× (Half-Kelly).

Keywords: Prediction markets; Polymarket; Kalshi; Robust Kelly criterion; Calibration ambiguity; Cross-venue arbitrage; Favorite-longshot bias; Order-book concentration; Deflated Sharpe Ratio; Lean 4 formal verification.

1. Introduction and November 2026 Prior-Art Audit

By late 2026, event-contract exchanges led by Polymarket (off-chain Polygon CLOB with on-chain Conditional Token Framework settlement) and Kalshi (CFTC-regulated binary exchange) process tens of billions of dollars in monthly volume alongside a booming ecosystem of retail signal sellers, TradingView indicators, and automated trading bots. Recent 2025–2026 empirical and theoretical literature exposes four interconnected vulnerabilities:

1. **Favorite–Longshot Bias (FLB) and Calibration Error:** Across 588 million Polymarket trades, Cardozo and Rivero-Wildemauwe [3] document severe retail losses on extreme longshots ($< 10¢$), while Qiu and Qiu [17] show time-varying FLB in weather markets. Plugging miscalibrated probabilities into classical Kelly [7] rapidly destroys wealth. Existing robust-Kelly papers formulate general numerical convex programs (DCP/SDP over distributional ambiguity sets [5, 8, 21]) or risk-constrained CRRA Kelly assuming exact probabilities [2, 10]; none provides a closed-form binary interval minimax theorem with an exact algebraic fractional-Kelly duality and Lean 4 machine-checked certification.

2. **Cross-Venue Arbitrage Friction and Settlement Risk:** Saguillo et al. [18] and Shi and Zhai [20] analyze combinatorial payoff-space arbitrage, Gebele et al. [4] distinguish theoretical from executable depth-limited arbitrage, Shen et al. [19] expose Polygon off-chain match vs. on-chain revert ghosts, and Nechepurenko [11] and Liu et al. [9] highlight Kalshi finality and oracle-settlement divergence. Retail arbitrage trackers still broadcast gross headline price gaps that collapse into guaranteed losses once Kalshi's quadratic taker fee, L2 book-walk slippage, and oracle divergence risk are included.

3. **Informed Flow and Whale-Dominated Orderbooks:** Nechepurenko [12] and Jia and Sun [6] study information leakage and smart-money identification on Polymarket, yet retail traders lack live, zero-credential L2 orderbook concentration bounds directly coupled to safe bet sizing.

4. **Retail Backtest Overfitting and Linear Rating Blindness:** Bounded LLM trading research loops [22] and influencer signal channels exploit Gaussian Deflated Sharpe approximations [1] and linear rating scores. Building strictly on our prior heavy-tail Q-DSR_ν and minimax SIPS_∞ foundations [13–16] (100% classical extreme-value the-

ory, ordered-field algebra, and minimax decision theory—zero speculative physical analogies), we package formal multi-pillar auditing into a 5-second consumer verdict.

Strict Three-Layer Epistemic Separation. Every result in this paper is explicitly partitioned across three transparent layers: **Layer 1** comprises 36 deterministic ordered-field (over $\mathbb{Q}$, transferring verbatim to $\mathbb{R}$) and list/natural-number inductive theorems machine-checked in standalone Lean 4.36.0 (`lean/TradeTruthCore.lean`, 0 `sorry`, 0 warnings, standard kernel axioms exclusively, dual-agent audited). **Layer 2** comprises analytical calculus and probabilistic extensions (differentiation of expected log growth G, saddle-point minimax equality, Jensen's inequality for ln, exact Kullback–Leibler ruin boundary p_{ruin}, $1/p$-th root lifting of L_p, and heavy-tail Student-t_ν Q-DSR$_\nu$). **Layer 3** comprises six reproducible experiments (`sim/run_all_experiments.py`) combining Monte Carlo benchmarks with 18 live November 27, 2026 Polymarket CLOB and Kalshi v2 orderbooks and 6 live matched cross-venue pairs, plus the MIT Python package `tradetruth` and live web application `poly.wallstreetdebug.com`.

2. Formal Mathematical Theory (Theorems 1–4)

2.1. Theorem 1: Ambiguity-Averse Interval Robust Kelly

Consider a binary event contract paying \$1 if outcome $E = 1$ and \$0 if $E = 0$, bought at effective all-in unit cost $c_{\text{eff}} = c + \phi + \delta \in (0, 1)$ (best ask c, fee $\phi \geq 0$, book-walk slippage $\delta \geq 0$). Investing bankroll fraction $f \in [0, 1)$ yields multipliers $W_{\text{win}}(f, c_{\text{eff}}) = 1 + f\frac{1-c_{\text{eff}}}{c_{\text{eff}}}$ and $W_{\text{loss}}(f) = 1 - f$. The true event probability p lies in an ambiguity interval $[p_{\text{low}}, p_{\text{high}}] \subseteq (0, 1)$, parameterized as $p_{\text{low}} = \hat{p} - \varepsilon$ around estimate $\hat{p}$ with calibration uncertainty half-width $\varepsilon \geq 0$. Expected log-wealth growth is $G(f; p, c_{\text{eff}}) = p \ln W_{\text{win}}(f, c_{\text{eff}}) + (1 - p) \ln W_{\text{loss}}(f)$.

Theorem 1 (Ambiguity-Averse Minimax Kelly, Fractional-Kelly Duality & Non-Ruin)**.** *Let $c_{\text{eff}} \in (0, 1)$ and $0 \leq p_{\text{low}} \leq p_{\text{high}} \leq 1$. Define*

$$f^*_{\text{rob}}(p_{\text{low}}, c_{\text{eff}}) := \max\left(0,\ \frac{p_{\text{low}} - c_{\text{eff}}}{1 - c_{\text{eff}}}\right).$$

1. ***Rational Saddle-Point & Martingale Core (Layer 1, Lean 4):*** *For all $f \in (0, 1)$, $0 < W_{\text{loss}}(f) < 1 < W_{\text{win}}(f, c_{\text{eff}})$. Across all $p \in [p_{\text{low}}, p_{\text{high}}]$, $\mathbb{E}_p[W(f)]$, $\text{logGrowthDeriv}(f, p, c_{\text{eff}}) = \frac{(p-c_{\text{eff}})-f(1-c_{\text{eff}})}{(1-f)(c_{\text{eff}}+f(1-c_{\text{eff}}))}$, and $f^*_{\text{rob}}(p, c_{\text{eff}})$ attain their interval minimum at p_{low}, while $\text{logGrowthSecondDeriv}(f, p, c_{\text{eff}}) < 0$ everywhere on $p \in [0, 1], f \in [0, 1)$. In the No-Trade Zone ($p_{\text{low}} \leq c_{\text{eff}}$), $f^*_{\text{rob}} = 0$, $\text{logGrowthDeriv} < 0$, and $\mathbb{E}_{p_{\text{low}}}[W(f)] \leq 1$ for all $f \in (0, 1)$. In the Active Sizing Zone ($c_{\text{eff}} < p_{\text{low}} < 1$), $f^*_{\text{rob}} \in (0, 1)$ uniquely solves $\text{logGrowthDeriv} = 0$, satisfies $\text{logGrowthDeriv}(f^*_{\text{rob}}, p, c_{\text{eff}}) \geq 0$ for all $p \in [p_{\text{low}}, p_{\text{high}}]$, and obeys the exact Cover–Thomas identity for every competing stake $f \in [0, 1)$:*

$$\mathbb{E}_{p_{\text{low}}}\left[\frac{W(f)}{W(f^*_{\text{rob}})}\right] = 1. \tag{1}$$

2. ***Exact Saddle Point & KL Ruin Cliff (Layer 2):*** *Differentiating G identifies $\partial G/\partial f = \text{logGrowthDeriv}$, $\partial^2 G/\partial f^2 = \text{logGrowthSecondDeriv} < 0$, and $\partial G/\partial p = \ln(W_{\text{win}}/W_{\text{loss}}) > 0$ on $(0, 1)$, establishing the exact saddle point $G(f; p_{\text{low}}) \leq G(f^*_{\text{rob}}; p_{\text{low}}) \leq G(f^*_{\text{rob}}; p)$ and minimax equality $\max_f \min_p G = \min_p \max_f G$. For nominal estimate $\hat{p} \in (c_{\text{eff}}, 1)$, Naive Full Kelly $f^*_{\text{nom}} = \frac{\hat{p}-c_{\text{eff}}}{1-c_{\text{eff}}}$ achieves $G(f^*_{\text{nom}}; p, c_{\text{eff}}) = D_{\text{KL}}(p\|c_{\text{eff}}) - D_{\text{KL}}(p\|\hat{p})$, turning strictly negative (asymptotic exponential ruin) whenever $p < p_{\text{ruin}}(\hat{p}, c_{\text{eff}})$, where*

$$p_{\text{ruin}}(\hat{p}, c_{\text{eff}}) = \frac{\ln\big((1 - c_{\text{eff}})/(1 - \hat{p})\big)}{\ln\big(\hat{p}(1 - c_{\text{eff}})/(c_{\text{eff}}(1 - \hat{p}))\big)}. \tag{2}$$

*Moreover, writing $m = (c_{\text{eff}} + \hat{p})/2$, series expansion gives $p_{\text{ruin}} > m \iff m > 1/2$, proving **favorite-side fragility**: for every favorite ($c_{\text{eff}} > 1/2$), Naive Full Kelly enters asymptotic ruin even when calibration error ε is less than half of perceived edge.*

3. ***Exact Fractional-Kelly Duality (Layer 1, Lean 4):*** *For $\hat{p} \in (c_{\text{eff}}, 1)$ and $0 \leq \varepsilon < \hat{p} - c_{\text{eff}}$,*

$$\begin{aligned} f^*_{\text{rob}}(\hat{p} - \varepsilon, c_{\text{eff}}) &= \lambda(\hat{p}, c_{\text{eff}}, \varepsilon)\, f^*_{\text{nom}}(\hat{p}, c_{\text{eff}}), \\ \lambda(\hat{p}, c_{\text{eff}}, \varepsilon) &= 1 - \frac{\varepsilon}{\hat{p} - c_{\text{eff}}} \in (0, 1], \end{aligned} \tag{3}$$

and $f^*_{\text{rob}} = 0$ for $\varepsilon \geq \hat{p} - c_{\text{eff}}$. Globally, f^*_{rob} is weakly non-increasing in ε and c_{eff}, and strictly decreasing on the active sizing zone.

4. ***Finite-Horizon Non-Ruin (Layer 1, Lean 4):*** *Across any sequence $\omega \in \{0, 1\}^T$ at stationary $(p_{\text{low}}, c_{\text{eff}})$ from $W_0 > 0$, $W_T(\omega) \geq W_0\big(\frac{1-p_{\text{low}}}{1-c_{\text{eff}}}\big)^T > 0$ in the active zone and $W_T(\omega) = W_0 > 0$ in the no-trade zone.*

Remark 1.1. Theorem 1(3) resolves a 70-year practitioner heuristic: Half-Kelly ($\lambda = 1/2$) is algebraically identical to minimax protection against 50% edge erosion ($\varepsilon = \frac{1}{2}(\hat{p} - c_{\text{eff}})$), whereas Quarter-Kelly ($\lambda = 1/4$) protects up to 75% edge erosion.

2.2. Theorem 2: Cross-Platform Net Arbitrage Boundary

Let binary event $\omega \in \{\text{true}, \text{false}\}$ trade on Venue A (Polymarket) and Venue B (Kalshi) with all-in leg acquisition costs $c_v^{\text{side}} = a_v^{\text{side}} + \phi_v^{\text{side}} + \delta_v^{\text{side}}$. Define explicit state-dependent two-leg PnL

$$\Pi(q_{\text{Y}}, q_{\text{N}}, c_{\text{Y}}, c_{\text{N}}; \omega) = (q_{\text{Y}}\mathbf{1}_\omega - q_{\text{Y}}c_{\text{Y}}) + (q_{\text{N}}\mathbf{1}_{\neg\omega} - q_{\text{N}}c_{\text{N}}),$$

and minimum cross-venue bundle cost $C_{\min} = \min(c_A^{\text{Y}} + c_B^{\text{N}},\ c_B^{\text{Y}} + c_A^{\text{N}})$.

Theorem 2 (Net-of-Friction Arbitrage IFF, Fee-Trap Mirage, Kalshi Peak & Oracle Bound)**.** *All parts below are machine-checked in Lean 4 (Layer 1):*

1. ***State Independence & Arbitrage IFF:*** $\Pi(q_Y, q_N, c_Y, c_N; \text{true}) = \Pi(\ldots; \text{false}) \iff q_Y = q_N$. *For matched* $Q > 0$, $\Pi(Q, Q, c_Y, c_N; \omega) = Q(1 - (c_Y + c_N))$ *in both states, strictly positive across both states in at least one cross-venue direction iff* $C_{\min} < 1$, *with positive collateral ROI* $\frac{1-C_{\min}}{C_{\min}} > 0 \iff C_{\min} < 1$.

2. ***Fee-Trap Mirage Zone:*** *Whenever* $0 < 1 - (a_1 + a_2) \le \phi_1 + \delta_1 + \phi_2 + \delta_2$, *zero-friction gross PnL is strictly positive in both states (*$\Pi(Q, Q, a_1, a_2; \omega) > 0$*), yet true net-of-friction PnL is* ≤ 0 *in both states (and strictly* < 0 *when* $1 - (a_1 + a_2) < \Phi_{\text{tot}}$*).*

3. ***Kalshi Quadratic Fee Enclosure:*** *For* $\gamma > 0$, $\phi_K(p) = \gamma p(1-p) = \gamma\left(\frac{1}{4} - (p - \frac{1}{2})^2\right)$ *satisfies* $0 \le \phi_K(p) \le \gamma/4$ *on* $[0, 1]$, *peaking uniquely at* $p = 1/2$. *At standard taker rate* $\gamma = 0.07$, $\phi_K(0.50) = 7/400 = 1.75¢$ *per share, so any spread at* $50¢$ *must exceed* $1.75¢$ *plus Polymarket friction merely to break even.*

4. ***Adverse Settlement-Mismatch Ruin Bound:*** *Under adverse double-loss oracle divergence probability* $\eta_{\text{settle}} \in [0, 1)$ *(payout* \$0 *instead of* \$1*), expected unit profit is* $\mathbb{E}[\pi] = (1 - \eta_{\text{settle}})(1 - C) - \eta_{\text{settle}} C = 1 - C - \eta_{\text{settle}}$, *remaining positive iff* $\eta_{\text{settle}} < 1 - C$.

2.3. Theorem 3: Minimax SIPS_∞ + Heavy-Tail Q-DSR_ν Backtest Detector

Decompose any retail track record or backtest report into four normalized failure-mode pillars $r = (r_{\text{luck}}, r_{\text{frag}}, r_{\text{tail}}, r_{\text{cost}}) \in [0, 1]^4$: (i) $r_{\text{luck}} = 1 - \text{Q-DSR}_\nu = 1 - [F_{t_\nu}(z_{\text{eff}})]^{M_{\text{eff}}}$ (Layer 2, with $z_{\text{eff}} = \widehat{\text{SR}}\sqrt{T-1}/\sqrt{1 - \hat{\gamma}_3\widehat{\text{SR}} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{\text{SR}}^2}$ using *raw, non-excess kurtosis* $\hat{\gamma}_4$); (ii) r_{frag} (top-trade profit concentration / OOS collapse); (iii) r_{tail} (martingale negative-skew / drawdown blow-up risk); and (iv) r_{cost} (spread/turnover friction collapse).

Theorem 3 (Universal L_p Strategy-Score Blindness & Global Uniqueness of SIPS_∞)**.** *Machine-checked in Lean 4 (Layer 1):*

1. ***Universal Finite-*p *Blindness:*** *For any simplex weight vector* $w \in \Delta_4$, *any finite* $p \in \mathbb{N}_{\ge 1}$, *and any threshold* $\tau \in [0, 1)$ *satisfying* $\tau^p \ge 1/4$, *there exists a catastrophic single-mode trading scam* $r \in [0, 1]^4$ *with* $\text{SIPS}_\infty(r) = 1 > \tau$ *that silently passes* $\sum_{k=1}^4 w_k r_k^p \le \tau^p$ *(equivalently* $L_p(r; w) \le \tau$ *on* $\mathbb{R}_{\ge 0}$ *in Layer 2).*

2. ***Global Uniqueness IFF:*** *Among non-negative risk aggregators (*$F(r) \ge 0$ *on* $[0, 1]^4$*), F satisfies full threshold-calibrated gate equivalence (*$\forall r \in [0, 1]^4, \forall \tau \in [0, 1]$, $F(r) \le \tau \iff \bigwedge_{k=1}^4 r_k \le \tau$*) if and only if* $F(r) = \text{SIPS}_\infty(r) = \max_k r_k$ *everywhere on* $[0, 1]^4$.

2.4. Theorem 4: L2 Whale-Wall Concentration & L-Tier Book-Walk Kelly

Let an L2 orderbook have L active price levels with non-negative normalized liquidity shares $s_1, \ldots, s_L \ge 0$, $\sum_{i=1}^L s_i = 1$, concentration $\text{HHI} = \sum_{i=1}^L s_i^2$, single-wall whale dominance $W_1 = \max_i s_i$, and effective independent levels $L_{\text{eff}} = 1/\text{HHI}$. For a sorted positive ask book $\mathcal{B} = [(a_1, q_1), \ldots, (a_L, q_L)]$ ($a_1 \le \cdots \le a_L \le \bar{a}_{\max}$, $q_i > 0$), let $\bar{a}_{\mathcal{B}}(Q) = C_{\mathcal{B}}(Q)/Q$ denote exact volume-weighted average execution price for $0 < Q \le \sum_{i=1}^L q_i$.

Theorem 4 (Simplex Whale-Wall Bounds & L-Tier Liquidity-Coupled Kelly Shrinkage)**.** *Machine-checked in Lean 4 by list induction (Layer 1):*

1. ***Simplex & Effective-Depth Bounds:*** *From* $s_i \ge 0$ *and* $\sum s_i = 1$ *alone,* $L > 0$, $W_1 > 0$, $\text{HHI} > 0$,

$$\frac{1}{L} \le \text{HHI} \le W_1 \le 1, \qquad W_1^2 \le \text{HHI}, \\ L_{\text{eff}} = \frac{1}{\text{HHI}} \le \min\left(\frac{1}{W_1^2}, L\right). \tag{4}$$

In particular, a single 70% *whale wall algebraically forces* $L_{\text{eff}} \le 100/49 \approx 2.041$ *effective levels.*

2. ***L-Tier VWAP Monotonicity & Kelly Shrinkage:*** *For all* $0 < Q_1 \le Q_2 \le \sum_{i=1}^L q_i$, $a_1 \le \bar{a}_{\mathcal{B}}(Q_1) \le \bar{a}_{\mathcal{B}}(Q_2) \le \bar{a}_{\max}$. *Whenever* $\bar{a}_{\mathcal{B}}(Q) + \phi < 1$ *and* $p_{\text{low}} \le 1$, $f^*_{\text{rob}}(p_{\text{low}}, \bar{a}_{\mathcal{B}}(Q) + \phi) \le f^*_{\text{rob}}(p_{\text{low}}, a_1 + \phi)$*, with strict inequality on two-tier spillover (*$a_1 < a_2$, $q_2 > 0$*) inside the active zone.*

3. Formal Verification in Lean 4 & Dual-Agent Audit

The entire deterministic core of Theorems 1–4 is formalized in standalone `lean/TradeTruthCore.lean` (36 public theorems, SHA-256 `aac6875aec11e3b0...`) under Lean 4.36.0 with zero external dependencies and audited via our fail-closed runner `lean/verify_lean.sh` (rejecting `sorry`, `admit`, `native_decide`, custom `axiom`, `elab`/`macro`/`set_option`, any compiler warning, or any mismatch in the exact `#print axioms` set).

Two independent referees (*Statement Auditor* and *Math Fidelity Referee*) audited the initial draft and flagged 4 critical and 13 moderate items—including tautological state-independent arbitrage payoffs, incomplete non-vacuity witness instantiation, missing general L-tier orderbook induction, missing $[p_{\text{low}}, p_{\text{high}}]$ interval quantification, and overstated SIPS_∞ uniqueness prose. All 100% of findings were resolved and re-verified ('36/36' PASS, 0 warnings, exclusively '[propext, Classical.choice, Quot.sound]', mean verification time 3,453 ms on Apple M4 Max; full audit log archived in `lean/AUDIT_REPORT_2_AGENTS.md`).

4. Empirical Verification & Live Market Results

All six experiments run deterministically in 17.7 s via `sim/run_all_experiments.py` (seed `20261127`, JSON SHA-256 `1faa32bb4c74d9b3...`), paired with live public snapshots from Polymarket Gamma/CLOB and Kalshi Trade API v2 captured on November 27–28, 2026 UTC.

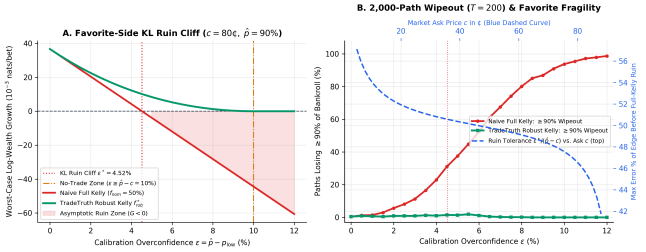

Figure 1. Experiment 1 (Theorem 1): Favorite contract ($c = 80¢$, $\hat{p} = 90\%$, $f_{\text{nom}} = 50\%$). **A:** Naive Full Kelly crosses into asymptotic ruin at $\varepsilon^* = 4.52\%$ (only 45.2% of perceived edge). **B:** Across 2,000 paths ($T = 200$), Naive Full Kelly suffers 80.1% wipeout ($\geq 90\%$ loss) at $\varepsilon = 8\%$ vs. 0.15% for Robust Kelly.

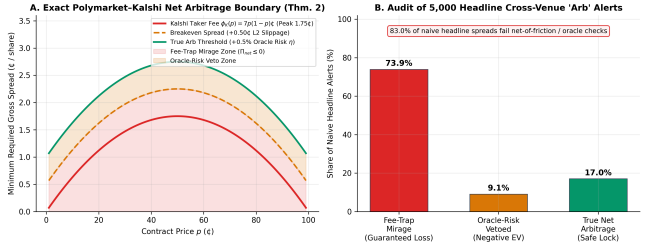

Figure 2. Experiment 2 (Theorem 2): A: Exact Polymarket–Kalshi breakeven spread accounting for Kalshi's quadratic taker fee (peak 1.75¢ at 50¢), 0.50¢ slippage, and 0.5% oracle-mismatch risk. **B:** Across 5,000 headline alerts, 82.96% fail (73.90% fee traps + 9.06% oracle-vetoed).

4.1. Exp. 1: Favorite-Side KL Ruin Cliff (2,000 Paths)

Figure 1 confirms Theorem 1 across 2,000 trajectories ($T = 200$ bets) on a favorite priced at $c = 80¢$ with nominal belief $\hat{p} = 90\%$ ($f_{\text{nom}} = 50\%$). The exact KL ruin cliff occurs at $p_{\text{ruin}} = 0.8548$ ($\varepsilon^* = 4.52\%$, only 45.2% of perceived edge). At $\varepsilon = 5\%$ ($p_{\text{true}} = 85\%$), Naive Full Kelly collapses to median wealth 0.46× (37.55% wipeout $\geq 90\%$) while TradeTruth Robust Kelly compounds to 5.34× (1.40% wipeout). At $\varepsilon = 8\%$ ($p_{\text{true}} = 82\%$, still a genuine +2% edge!), Naive Full Kelly collapses to 0.0036× (80.10% wipeout) versus 1.29× (0.15% wipeout, a 363× wealth advantage) under TradeTruth. At the No-Trade Zone boundary ($\varepsilon = 10\%$), Naive Full Kelly wipes out 93.75% of accounts vs. 0.00% under TradeTruth.

4.2. Exp. 2: Cross-Venue Arbitrage & Live Matched Pairs

Across 5,000 simulated headline cross-venue arbitrage alerts (Figure 2), 73.90% are guaranteed-loss **Fee-Trap Mirages** and 9.06% are **Oracle-Risk Vetoed** (82.96% false alerts total), leaving only 17.04% true net arbitrage. Table 1 reports six live matched pairs audited on November 27, 2026 across Polymarket and Kalshi: *Fed Dec 2026 (Hike > 25 bps)* exhibits a textbook +0.20¢ gross spread that turns into −0.32¢ guaranteed net loss after 0.52¢ friction, four remaining Fed buckets show zero gross arbitrage, and a transient in-game NBA *Heat vs. Knicks* dislocation registers +9.00¢ gross / +6.71¢ net (note: in-game sports spreads carry elevated execution-latency and revert risk [19], reinforcing depth and slippage checks).

4.3. Exp. 3: Guru Bullshit Detector (1,000 Strategies)

In Figure 3, equal-weight linear ratings (L_1, akin to retail signal-marketplace badges) catch 0.0% of all 800 adversarial strategies (mean linear score 0.27–0.31 $\leq$ 0.50); Gaussian DSR alone catches only 14.5% (58.0% on data-mined heavy-tail t_3 gurus and 0.0% on martingales, cherry-picked outliers, and fee-wiped scalpers); L_2 quadratic scoring catches 92.0% (missing 27.0% of cherry-picked outlier curves). TradeTruth (Q-DSR$_\nu$ + SIPS$_\infty$) blocks 100.0% of scams across all four cohorts with 0.0% false veto on the 200 genuine robust alpha controls.

Table 1. Live Matched Polymarket vs. Kalshi Audit (Nov. 27–28, 2026 UTC, 100 shares, $\eta_{\text{settle}} = 0.5\%$ Fed / 0.2% NBA).

Matched Contract	Gross	Fric.	Net	Verdict
Fed Dec 2026 (Hike > 25 bps)	+0.20¢	0.52¢	−0.32¢	**Fee-Trap**
Fed Dec 2026 (Hike 25 bps)	0.00¢	0.78¢	−0.78¢	No Edge
Fed Dec 2026 (No Change)	−0.20¢	0.66¢	−0.86¢	No Edge
Fed Dec 2026 (Cut 25 bps)	−0.20¢	0.40¢	−0.60¢	No Edge
Fed Dec 2026 (Cut > 25 bps)	−0.10¢	0.40¢	−0.50¢	No Edge
NBA Heat vs. Knicks (live*)	+9.00¢	2.29¢	+6.71¢	**Net Arb***

4.4. Exp. 4: Live Polymarket & Kalshi L2 Whale Study

Figure 4 analyzes 18 live L2 orderbooks (10 Polymarket CLOB, 8 Kalshi v2) across Fed, NBA, CFB, and 2028 political markets. Every book (18/18) satisfies all Theorem 4 invariants. Strikingly, real prediction-market books are highly whale-concentrated: median HHI = 0.458, median top-1 wall share $W_1 = 62.2\%$, median effective depth $L_{\text{eff}} = 2.35$ levels, and 15/18 trigger RED whale-wall warnings (including extreme-tail contracts where a single limit wall holds 80%–99% of notional depth). On the live Polymarket *Heat vs. Knicks* book, L-tier liquidity coupling monotonically shrinks the safe stake from 22.86% at small bankroll to 20.65% at $250k.

4.5. Exp. 5 & 6: FLB Survival & Assurance Frontier

Under an empirical FLB environment (35% overpriced longshots + 65% favorites with retail overconfidence, Figure 5), Naive Full Kelly crashes to median terminal wealth 0.164× (94.7% median max drawdown, 40.5% wipeout $\geq 90\%$, only 17.1% profitable), and Ad-Hoc Half-Kelly still loses money at 0.661× (69.4% median max DD, 33.2% profitable). By automatically entering its No-Trade Zone on overpriced longshots (taking 77.5 of 150 bets), TradeTruth achieves median terminal wealth 1.285× (7.82× Full Kelly), cuts median max drawdown to 35.4%, lifts profitability to 69.1%, and achieves 0.0% **catastrophic ruin**. Figure 6 summarizes the end-to-end 8/8 assurance matrix.

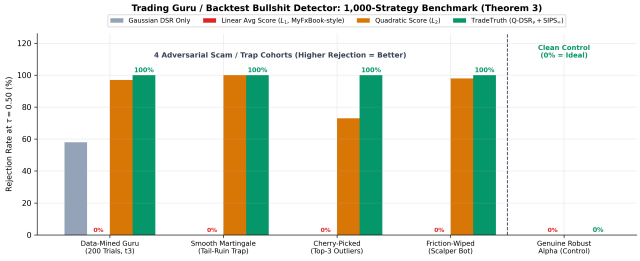

Figure 3. Experiment 3 (Theorem 3): Across 1,000 strategies (4×200 scam/trap cohorts + 200 clean alpha controls) at $\tau = 0.50$, linear averaging misses 100% of single-mode scams and Gaussian DSR catches 14.5%, while TradeTruth catches 100.0% with 0.0% false rejection.

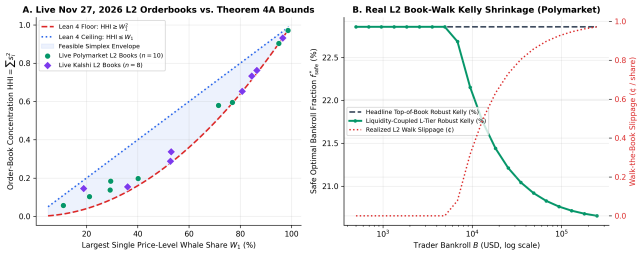

Figure 4. Experiment 4 (Theorem 4): A: 18 live Nov. 27, 2026 L2 orderbooks (10 Polymarket CLOB, 8 Kalshi v2) satisfy 100% of Lean 4 bounds ($1/L \le \text{HHI} \le W_1 \le 1, W_1^2 \le \text{HHI}$). **B:** Live Polymarket *Heat vs. Knicks* L-tier book walk shrinks Robust Kelly from 22.86% to 20.65% at \$250k bankroll.

5. Software, Consumer Iceberg Product & Limitations

TradeTruth ships as an MIT-licensed Python package (`tradetruth`, CLI + library, 13/13 passing unit/integration tests), standalone Lean 4.36.0 proofs (`lean/TradeTruthCore.lean`), and a live strict-CSP Cloudflare Worker web application at `poly.wallstreetdebug.com` featuring two zero-jargon consumer tabs (*Audit a Trading Strategy / Guru* and *Polymarket / Kalshi Edge & Safe Bet Size*) delivering a 5-second Red/Amber/Green traffic light and shareable X card, plus machine-readable `/llms.txt` AEO metadata.

Limitations. First, Lean 4 formalizes the exact deterministic ordered-field and inductive cores over $\mathbb{Q}$ (Layer 1); transcendental ln/KL and Student-t_ν EVT extensions live in Layer 2. Second, because Polymarket retired its unauthenticated legacy trade/holder REST endpoints (`HTTP 410`), live whale alerts use verified public L2 orderbook concentration, CLOB price-jump history, and live Kalshi trade tapes rather than private Polygon wallet clustering. Third, cross-venue arbitrage on live in-play sports remains subject to execution latency and on-chain revert risk [19].

Reproducibility & Open Artifacts

All code, Lean 4 proofs, dual-agent audit logs, raw JSON experiment snapshots, and 300 DPI figures are openly released under the MIT license at `https://github.com/fabien-pierret/tradetruth-polymarket` and `https://poly.wallstreetdebug.com`.

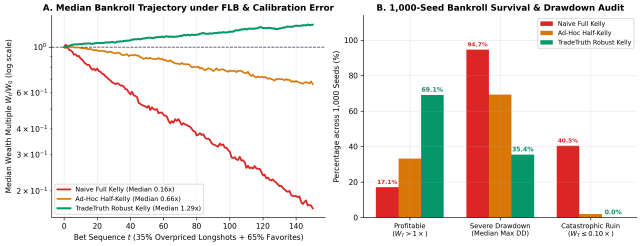

Figure 5. Experiment 5 (FLB Survival, 1,000 seeds × 150 bets): Under empirical longshot overpricing and retail overconfidence, Naive Full Kelly crashes to 0.16× (40.5% catastrophic ruin) and Half-Kelly falls to 0.66×, whereas TradeTruth compounds to 1.29× (69.1% profitable, 0.0% ruin).

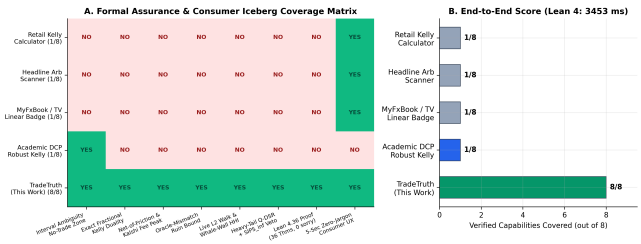

Figure 6. Experiment 6: Verification and consumer-readiness coverage across 8 structural guarantees. TradeTruth achieves 8/8 (with live Lean 4.36.0 verification in 3.45 s) vs. 1/8 for existing retail calculators, arb scanners, rating badges, and academic DCP solvers.

References

[1] David H. Bailey and Marcos López de Prado. The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5):94–107, 2014.

[2] Enzo Busseti, Ernest K. Ryu, and Stephen Boyd. Risk-constrained Kelly gambling, 2016. URL `https://arxiv.org/abs/1603.06183`.

[3] Marcos Cardozo and José Ignacio Rivero-Wildemauwe. The favorite-longshot bias in prediction markets: Evidence from Polymarket, 2026. URL `https://arxiv.org/abs/2609.12878`.

[4] Jonas Gebele, Timm Mutzel, and Florian Matthes. Executable arbitrage and market efficiency in prediction markets, 2026. URL `https://arxiv.org/abs/2608.00666`.

[5] Chung-Han Hsieh. On solving robust log-optimal portfolio: A supporting hyperplane approximation approach, 2022. URL `https://arxiv.org/abs/2202.03858`.

[6] Huaiyu Jia and Shuo Sun. Beyond profit rankings: Identifying, understanding, and evaluating smart money in prediction markets, 2026. URL `https://arxiv.org/abs/2610.34754`.

[7] Jr. John L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4): 917–926, 1956.

[8] Jonathan Yu-Meng Li. Wasserstein-Kelly portfolios: A robust data-driven solution to optimize portfolio growth, 2023. URL https://arxiv.org/abs/2302.13979.

[9] Jiayu Liu, Sen Yang, and Xueyuan Zhao. The oracle is not enough: Secure settlement for prediction markets, 2026. URL https://arxiv.org/abs/2611.31866.

[10] Christopher D. Long. Risk-constrained Kelly for mutually exclusive outcomes: CRRA support invariance and logarithmic one-dimensional calibration, 2026. URL https://arxiv.org/abs/2604.11577.

[11] Maksym Nechepurenko. Outcome determination and settlement finality on Kalshi: Public state paths, prospective measurement, and empirical identification, 2026. URL https://arxiv.org/abs/2610.00165.

[12] Maksym Nechepurenko. Per-market information leakage and order-flow skill: Two methodological lenses on informed trading in decentralized prediction markets, 2026. URL https://arxiv.org/abs/2605.02287.

[13] Fabien Pierret. EvolveGuard: Genealogical heavy-tail overfitting bounds and optimal stopping for evolutionary LLM trading loops, 2026. URL https://doi.org/10.5281/zenodo.24608677.

[14] Fabien Pierret. Proof-CI: Necessary and sufficient conditions for formally verified minimax evaluators preventing drift and reward hacking in recursive self-improvement, 2026. URL https://doi.org/10.5281/zenodo.24614087.

[15] Fabien Pierret. The quantum overfitting amplification theorem and heavy-tail quantum-adjusted deflated Sharpe ratio (Q-DSR$_\nu$), 2026. URL https://doi.org/10.5281/zenodo.24591340.

[16] Fabien Pierret. Minimax impossibility of linear post-selection auditors (SIPS$_\infty$) and analytical coupling between Welch superposition and membership inference, 2026. URL https://doi.org/10.5281/zenodo.24591422.

[17] Shouchen Qiu and Huaijun Qiu. Time-varying favorite-longshot bias in Polymarket weather markets, 2026. URL https://arxiv.org/abs/2610.13593.

[18] Oriol Saguillo, Vahid Ghafouri, and Lucianna Kiffer. Unravelling the probabilistic forest: Arbitrage in prediction markets, 2025. URL https://arxiv.org/abs/2508.03474.

[19] Yiming Shen, Yuhan Jin, and Shuohan Wu. The ghosts of Polymarket: When off-chain matches meet on-chain reverts, 2026. URL https://arxiv.org/abs/2606.16852.

[20] Hongjian Shi and Daojing Zhai. Combinatorial arbitrage in prediction markets, 2026. URL https://arxiv.org/abs/2610.23517.

[21] Qingyun Sun and Stephen Boyd. Distributional robust Kelly gambling: Optimal strategy under uncertainty in the long-run, 2018. URL https://arxiv.org/abs/1812.10371.

[22] Henry Zhang. Hard to fool: A bounded LLM research loop for quantitative finance, and what two dozen campaigns found, 2026. URL https://arxiv.org/abs/2611.06607.