Do Prediction Markets Match Option Prices? Bitcoin Threshold Evidence from Binance and Polymarket

Abstract

The digitization of financial markets has produced two classes of platforms that price, in principle, the same state-contingent payoffs: centralized crypto-option exchanges and blockchain-based prediction markets. This paper asks whether these digitally native venues agree on value. It provides the first option-implied benchmark test of prediction-market pricing for cryptocurrency threshold contracts. For each hour in a matched sample, I compare the Polymarket Yes price, $P_{poly,t}$, with the discounted risk-neutral binary value implied by a listed Binance call option on the same underlying, strike, and maturity, $P_{fair,t}$, and study the gap $D_t = P_{poly,t} - P_{fair,t}$. In the main September 2023 Bitcoin contract, the mean pricing gap equals 5.6 percentage points across 214 hourly observations ($t = 6.46$, $p < 10^{-9}$). Pooling three Binance-compatible Bitcoin threshold markets yields a mean gap of 6.3 percentage points across 287 observations, robust to HAC and block-bootstrap inference. The gap is persistent — with an AR(1) half-life of roughly four hours — yet mean-reverting, consistent with slow information transmission between segmented venues rather than mechanical noise. Cross-sectional regressions reveal that the wedge is largest at low option-implied probabilities and long maturities, a pattern consistent with speculative demand for prediction-market contracts rather than measurement error. A Deribit extension on the same three Bitcoin contracts produces a larger pooled gap of 11 percentage points, while a smaller Ethereum exercise yields mixed evidence. The results demonstrate that digital fragmentation of financial markets — across algorithmically matched option exchanges and decentralized prediction platforms — generates systematic, persistent pricing wedges even for economically identical payoffs.

Keywords: [JEL] G13, G14, G19

1. Introduction

The digitization of financial markets has produced two classes of platforms that price, in principle, the same state-contingent payoffs. On one side sit centralized crypto-option exchanges — algorithmically matched, continuously quoted, and integrated with global derivatives infrastructure. On the other side sit blockchain-based prediction markets — decentralized, oracle-settled, and populated by a participant base whose access to financial markets has been transformed by digital platforms. Whether these two digital ecosystems agree on the value of an identical binary payoff is an open empirical question with implications for how digitization shapes price discovery and cross-market information transmission.

Prediction markets are increasingly treated as probability aggregators, but this interpretation requires a benchmark. For a contract that pays one dollar if Bitcoin closes above a fixed threshold on a fixed date, the natural benchmark is not a survey forecast or a historical frequency — it is the market price of a financial claim with an identical payoff. This paper exploits that observation to conduct what is, to our knowledge, the first option-implied benchmark test of prediction-market pricing for cryptocurrency threshold contracts.

A vanilla call option traded on a centralized exchange can be inverted to recover the market-implied probability of a threshold event under standard no-arbitrage pricing. Comparing that probability with the prediction-market price for the same event yields a sharp cross-market test: if both venues price the same state-contingent payoff efficiently, their average difference should be indistinguishable from zero after accounting for frictions. If one market is systematically above or below the other, the difference should show up clearly in the data.

I implement this test using exact Binance historical option data, exact Binance spot bars, and Polymarket trade-level data for three Bitcoin threshold contracts from mid-2023. The empirical design is deliberately narrow. Only contracts with a precise option counterpart — identical underlying, strike, and maturity — enter the sample. This restriction sacrifices

breadth for comparability and is the right trade-off for a first study of this kind: it rules out contract-mapping ambiguity as an alternative explanation and ensures that any measured wedge is a statement about pricing, not about payoff structure.

The main finding is a robust positive wedge. Polymarket Yes prices were, on average, 5.6 percentage points above Binance-implied risk-neutral values in the September 2023 contract and 6.3 percentage points above in the pooled three-market panel. This gap is statistically precise under classical, HAC, and block-bootstrap inference. It is economically meaningful relative to estimated frictions. And it is persistent over intraday horizons — with an autoregressive half-life of approximately four hours — rather than transitory noise. These features rule out simple explanations based on bid–ask spreads or quote-staleness.

The structure of the wedge provides additional discipline. Cross-sectional regressions on the pooled panel show that the gap is largest when the option-implied probability is low and when time to expiry is long. This pattern is consistent with speculative or sentiment-driven demand for out-of-the-money prediction-market contracts. It is difficult to reconcile with the alternative hypothesis that the wedge is driven by liquidity premia or settlement-risk adjustments, which would predict a different sign and magnitude pattern.

The paper contributes to two literatures. The first is the literature on prediction market efficiency and information aggregation. Early work by **?**] and **?**] establishes that prediction markets produce well-calibrated forecasts for political and economic events, while more recent papers study limit-order dynamics [**?**] and arbitrage across platforms [**?**]. Our contribution is to provide an option-theoretic benchmark that is grounded in no-arbitrage pricing rather than ex-post calibration, allowing a sharp cross-market test independent of realized outcomes.

The second is the literature on cross-market price discovery and limits to arbitrage. **?**] and **?**] show that capital constraints prevent full price equalization across markets even when the same security trades in multiple venues. **?**] documents slow-moving capital as a source of persistent mispricings in derivatives markets. Our finding of a persistent, mean-reverting wedge with a four-hour half-life is consistent with these mechanisms: information and capital

move across the Polymarket and Binance ecosystems, but not instantly and not without cost.

The closest antecedent is the literature on prediction market prices and implied probabilities from financial derivatives more broadly [**?** **?**]. Those papers compare prediction-market prices against alternatives constructed from equity or options markets for political and macroeconomic events. Our setting differs in three respects: the underlying event is a financial threshold (Bitcoin price), the option counterpart is exact rather than approximate, and the data are aligned at hourly frequency, allowing a time-series characterization of the wedge rather than a single-point comparison.

The paper also connects to the broader literature on digitization and financial inclusion. The expansion of digital prediction platforms has extended financial market access to participants who would not typically engage with listed derivatives, generating a new class of digitally native price signals. Understanding whether those signals are aligned with option-market benchmarks is a prerequisite for evaluating whether digital platforms genuinely aggregate information or instead reflect segmented, potentially mispriced demand. Our finding that the wedge is largest at low probabilities and long maturities is consistent with the hypothesis that digital platform access draws in participants who systematically over-weight low-probability outcomes, a pattern with direct implications for how digital financial inclusion shapes aggregate price discovery.

The rest of the paper is organized as follows. Section **??** describes the data and contract mapping. Section **??** develops the Black–Scholes benchmark, the delta-method uncertainty propagation, and the hypothesis testing framework. Section **??** presents the main results, pooled evidence, cross-sectional patterns, and persistence analysis. Section **??** describes the arbitrage-style proxy backtest. Section **??** reports the Deribit extension and the Ethereum exercise. Section **??** interprets the findings and discusses scope conditions. Section **??** concludes. Analytical derivations are collected in the Appendix.

2. Data

2.1. Contract mapping

The key empirical step is matching a prediction-market contract to an options-based payoff with the same economic meaning. A Polymarket Yes share pays one unit at maturity T if the terminal Bitcoin level S_T exceeds a threshold K. Abstracting from minor venue-specific settlement details, that payoff is the binary indicator $\mathbf{1}\{S_T > K\}$. Under a frictionless no-arbitrage benchmark with constant short rate r, its time-t price is

$$e^{-r\tau_t}\mathbb{E}^{\mathbb{Q}}[\mathbf{1}\{S_T > K\} \mid \mathcal{F}_t], \qquad \tau_t = \frac{T-t}{365}. \tag{2.1}$$

Under Black–Scholes dynamics, this expectation equals the value of a cash-or-nothing call option, providing a direct bridge from a listed vanilla option price to a prediction-market benchmark.

2.2. Sources and sample

The analysis combines three public data sources.

1. *Binance option data.* Option observations are taken from the public `EOHSummary` archive, which provides historical bid and ask information for all listed Bitcoin option contracts at end-of-hour frequency [**?**].
2. *Binance spot data.* Spot prices are taken from the public hourly bar archive at the same source [**?**].
3. *Polymarket data.* Prediction-market prices are constructed from public transaction-level trade data and, where available, from historical hourly prices [**?**].

The main specification studies the Polymarket contract "BTC above $27,000 at end of September?" (market id 252196), matched to the Binance call option `BTC-230929-27000-C`. For pooled robustness, I also examine two August 2023 Bitcoin threshold contracts matched

to `BTC-230901-28000-C` and `BTC-230901-26000-C`. The pooled exact-data panel contains 287 aligned hourly observations across these three contracts.

2.3. Variables

Table 1: Variable definitions

Variable	Definition
$C_{mkt,t}$	Binance option mid-price, average of best bid and ask
S_t	Binance spot reference level
K	Strike shared by the two contracts
τ_t	Time to Polymarket expiration (years, 365-day convention)
$\tau_t^{(c)}$	Time to Binance option expiration
$P_{poly,t}$	Polymarket Yes price on the unit interval
$P_{fair,t}$	Option-implied risk-neutral binary benchmark
$s_{B,t}$	Relative Binance bid–ask spread
s_P	Polymarket spread parameter
f_B, f_P	Binance and Polymarket fee parameters
D_t	Pricing gap $P_{poly,t} - P_{fair,t}$

All Polymarket observations are placed on the Yes-probability scale. A Yes trade with price p enters as p; a No trade at price p enters as $1 - p$. This ensures that every observation can be interpreted against the same binary event probability.

2.4. Time alignment

All series are aligned at hourly frequency. Binance spot and option observations are recorded on the hour. Polymarket observations are merged backward to the latest available quote or trade at or before the Binance timestamp. In the exact-data specification, the staleness tolerance is set to zero, so the panel uses only contemporaneous hourly matches. Observations with extremely short maturity are excluded to avoid numerically unstable implied-volatility estimates.

3. Model and Inference

3.1. Black–Scholes benchmark

Let S_t denote the underlying Bitcoin level, K the strike, r the continuously compounded short rate, and σ the instantaneous volatility parameter. Under the Black–Scholes model [**?**], the European call price is

$$C_{BS}(S_t, K, r, \tau, \sigma) = S_t\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \tag{3.1}$$

where

$$d_1 = \frac{\log(S_t/K) + (r + \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \qquad d_2 = d_1 - \sigma\sqrt{\tau}. \tag{3.2}$$

The corresponding discounted cash-or-nothing call value is

$$P_{fair,t}(\sigma) = e^{-r\tau_t}\Phi(d_{2,t}(\sigma))\,. \tag{3.3}$$

This quantity serves as the option-implied risk-neutral benchmark for the prediction-market event. Its derivation is stated formally in Proposition **??**.

Black–Scholes is the natural starting point for two reasons. First, it is the standard benchmark for constructing an implied probability from a vanilla option, and any deviation from it is both quantifiable and replicable. Second, corrections for stochastic volatility would in general widen the measured gap: Bitcoin smiles are left-skewed for near-money strikes, implying that the lognormal model understates the risk-neutral probability of large upward moves and therefore overstates $P_{fair,t}$ at the margin. The Black–Scholes benchmark is thus conservative with respect to the sign of the main finding.

3.2. Implied-volatility inversion

At each time t, volatility is defined implicitly by

$$C_{mkt,t} = C_{BS}\Big(S_t, K, r, \tau_t^{(c)}, \sigma_t\Big)\,. \tag{3.4}$$

I estimate $\hat{\sigma}_t$ numerically as the root of $F_t(\sigma) = C_{BS}(S_t, K, r, \tau_t^{(c)}, \sigma) - C_{mkt,t}$. Once obtained, $\hat{\sigma}_t$ is substituted into (**??**) to construct $P_{fair,t}$.

3.3. Delta-method uncertainty propagation

Because $P_{fair,t}$ is a nonlinear function of an estimated implied volatility, it carries estimation uncertainty. I treat the observed option mid as a latent efficient price plus additive measurement error,

$$C_{mkt,t} = C_t^{\star} + \varepsilon_t, \quad \mathbb{E}[\varepsilon_t \mid \mathcal{F}_t] = 0, \quad \mathrm{Var}(\varepsilon_t \mid \mathcal{F}_t) = \varsigma_{C,t}^2. \tag{3.5}$$

A first-order implicit-function expansion with call vega $\mathcal{V}_t = \partial C_{BS}/\partial\sigma$ evaluated at $\hat{\sigma}_t$ gives

$$\hat{\sigma}_t - \sigma_t^{\star} = -\frac{\varepsilon_t}{\mathcal{V}_t} + o_p(\varsigma_{C,t}). \tag{3.6}$$

Writing $P_{fair,t} = g(\hat{\sigma}_t)$ with $g(\sigma) = e^{-r\tau_t}\Phi(d_{2,t}(\sigma))$, the delta method yields

$$\mathrm{Var}(P_{fair,t} \mid \mathcal{F}_t) \approx \left(g'(\hat{\sigma}_t)\right)^2 \frac{\varsigma_{C,t}^2}{\mathcal{V}_t^2}. \tag{3.7}$$

The sensitivity $g'(\sigma) = -e^{-r\tau_t}\phi(d_2)d_1/\sigma$ is derived in closed form in Proposition **??**. The plug-in standard error is

$$SE(P_{fair,t}) := |g'(\hat{\sigma}_t)| \, \frac{\hat{\varsigma}_{C,t}}{\mathcal{V}_t}, \tag{3.8}$$

where $\hat{\varsigma}_{C,t} = \max\{(C_{high,t} - C_{low,t})/2,\ 0.01\,|C_{mkt,t}|,\ 10^{-10}\}$ uses the intra-hour high–low range as a conservative scale for within-hour quote variation.

3.4. Friction-adjusted comparison band

I distinguish statistical uncertainty in the benchmark from explicit transaction costs. The total friction term is

$$TF_t = (f_B + f_P) + \frac{s_{B,t} + s_P}{2}, \tag{3.9}$$

where f_B and f_P are fee proxies for the option venue and Polymarket, and $s_{B,t}$ is the relative option bid–ask spread (replaced by a high–low proxy when quoted spreads are unavailable). The friction-adjusted band is

$$CI_{adj,t} = [P_{fair,t} - 1.96\, SE(P_{fair,t}) - TF_t,\ P_{fair,t} + 1.96\, SE(P_{fair,t}) + TF_t]\,. \tag{3.10}$$

Observations with $P_{poly,t} \notin CI_{adj,t}$ exhibit disagreement that is large relative to both measurement uncertainty and conservative friction estimates.

3.5. Discrepancy process and hypothesis testing

The central empirical object is

$$D_t = P_{poly,t} - P_{fair,t}. \tag{3.11}$$

The null hypothesis of average pricing equivalence is

$$H_0 : \mathbb{E}[D_t] = 0, \qquad H_1 : \mathbb{E}[D_t] \neq 0. \tag{3.12}$$

The baseline test statistic is the one-sample Student statistic $t_n = \bar{D}_n/(s_D/\sqrt{n})$. Because the discrepancy is observed at hourly frequency and is likely serially dependent, I supplement this with HAC confidence intervals based on the Newey–West covariance estimator [**?**] and with circular block-bootstrap confidence intervals using blocks of length approximately $n^{1/3}$.

Discrete pricing on Polymarket introduces mechanical rounding noise. Modeling the rounding error as $\varepsilon_P \sim \text{Unif}[-\Delta_P/2, \Delta_P/2]$ gives $\mathbb{E}[\varepsilon_P] = 0$ and $\text{Var}(\varepsilon_P) = \Delta_P^2/12$. This term is dominated by $\text{Var}(P_{fair,t})$ in the parameter range studied here, so it does not materially affect inference on $\mathbb{E}[D_t]$; the formal statement appears in Proposition **??**.

3.6. Dynamic specification

To characterize persistence, I model the discrepancy as the AR(1) process

$$D_t = a + \phi D_{t-1} + u_t. \tag{3.13}$$

When $|\phi| < 1$, this corresponds to the Euler discretization of an Ornstein–Uhlenbeck process with long-run mean $\theta = a/(1-\phi)$ and half-life $h_{1/2} = \log(0.5)/\log(\phi)$ hours. I report the AR(1) estimate alongside the Augmented Dickey–Fuller test for the null of a unit root [**?**].

4. Empirical Results

4.1. Main market

The main specification studies the Polymarket contract "BTC above $27,000 at end of September?" matched to Binance option `BTC-230929-27000-C`. The aligned sample contains 214 hourly observations. Sample averages of the key inputs are

$$\overline{C_{mkt}} = 253.45, \quad \overline{S_t} = 26{,}664.80, \quad \overline{P_{poly}} = 0.4043, \quad \overline{P_{fair}} = 0.3484.$$

The raw averages already indicate that the prediction-market price tends to lie above the option-implied binary benchmark. The mean discrepancy is

$$\bar{D} = 0.0558, \qquad s_D = 0.1264.$$

Economically, this means the Polymarket Yes price exceeds the Binance-implied risk-neutral binary price by roughly 5.6 percentage points on average.

Table **??** collects the principal inferential quantities.

Table 2: Main-market results: BTC above $27,000 at end of September

Statistic	Value
Aligned observations n	214
Mean discrepancy $\bar{D}$	0.0558***
Standard deviation s_D	0.1264
One-sample t statistic	6.46
One-sample p-value	6.86×10^{-10}
HAC 95% CI for $\mathbb{E}[D_t]$	$[0.0228,\ 0.0889]$
Block-bootstrap 95% CI for $\mathbb{E}[D_t]$	$[0.0212,\ 0.0911]$
Mean friction term $\overline{TF_t}$	0.0427
Mean standard error $\overline{SE(P_{fair,t})}$	0.0458
Share of obs. with $P_{poly,t} \notin CI_{adj,t}$	0.411
AR(1) coefficient $\hat{\phi}$	0.8485 ± 0.0511
Implied half-life (hours)	4.2
ADF statistic	-3.69 ($p = 0.004$)

Notes. Two-sided one-sample test of $H_0 : \mathbb{E}[D_t] = 0$. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. HAC standard errors use the Newey–West estimator. Bootstrap intervals use circular blocks of length $\lfloor n^{1/3} \rfloor$. The ADF test includes a constant and the default lag selection by AIC.

The inference is unambiguous. The classical test rejects H_0, the HAC interval remains strictly positive, and the block-bootstrap interval leads to the same conclusion. Moreover, 41.1% of hourly observations lie outside the friction-adjusted band, indicating that the discrepancy is not only statistically detectable but frequently large relative to the combined measurement uncertainty and explicit transaction costs.

The dynamic properties reinforce the interpretation of a genuine pricing wedge. The ADF test rejects a unit root ($p = 0.004$), confirming mean reversion. At the same time, the AR(1) estimate of $\hat{\phi} = 0.849$ implies an intraday half-life of 4.2 hours — long enough to make the gap exploitable in principle but short enough to rule out permanent divergence. Stationary but persistent discrepancies of this type are the signature of slow information transmission or capital segmentation across venues, not of independent measurement noise around a common latent benchmark.

Figure **??** visualizes the main time series. The upper panel plots $P_{poly,t}$, $P_{fair,t}$, and the

friction-adjusted band. The lower panel shows the gap process D_t directly. The positive wedge is not the product of a single outlier episode; it persists across a substantial fraction of the contract's life, with the sharpest deviations concentrated in the final ten days before expiry when the option-implied probability collapses but the prediction-market price adjusts more slowly.

Figure 1: Main-market time series. Upper panel: Polymarket Yes price (blue), Binance-implied benchmark (red), and friction-adjusted band (shaded). Lower panel: discrepancy process $D_t = P_{poly,t} - P_{fair,t}$.

4.2. Pooled evidence

Table **??** reports results for all three Binance-compatible contracts and for the pooled panel.

Table 3: Market-level and pooled results

Market	n	Mean D	95% HAC CI	BB 95% CI	Share outside C
BTC > \$28,000 end of August	37	0.0957***	[0.0695, 0.1218]	—	0.405
BTC > \$26,000 end of August	36	0.0745**	[−0.027, 0.176]	—	0.028
BTC > \$27,000 end of September	214	0.0558***	[0.0228, 0.0889]	[0.021, 0.091]	0.411
Pooled panel	287	0.0633***	[0.0327, 0.0940]	[0.031, 0.095]	0.362

Notes. Two-sided one-sample test of $H_0 : \mathbb{E}[D_t] = 0$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. BB = circular block-boots
Individual-contract bootstrap intervals are omitted where the sample is too small for reliable block resampling.

The pooled panel mean of $\bar{D}_{pool} = 0.0633$ carries a one-sample statistic of $t = 8.24$ ($p = 6.26 \times 10^{-15}$). The HAC and bootstrap intervals are [0.033, 0.094] and [0.031, 0.095], respectively. The positive average wedge is not idiosyncratic to a single contract or resolution episode. The sign of the discrepancy is positive in all three markets; the August \$26,000 contract has a wider HAC interval, reflecting the fact that it spends much of its life near the upper boundary of the probability space where both venues converge to near-unity prices.

Figure **??** summarizes the pooled distribution and market-specific mean estimates.

(a) Pooled discrepancy distribution.

(b) Market-level mean discrepancy with 95% HAC CIs.

Figure 2: Pooled evidence across the three Bitcoin threshold markets.

4.3. Robustness and cross-sectional patterns

The pooled result is not driven by outliers. The median discrepancy is 0.0642, the 10% trimmed mean is 0.0576, and a sign test strongly rejects a 50–50 split: 68.6% of hourly observations have $D_t > 0$ ($p = 2.45 \times 10^{-10}$). The same patterns hold within the main September contract: median 0.0640, trimmed mean 0.0531, 68.2% positive observations ($p = 1.03 \times 10^{-7}$). Table **??** collects these outlier-robust summaries.

Table 4: Outlier-robust summary statistics

Sample	n	Median D	Trimmed mean	Share($D_t > 0$)
BTC > \$28,000 end of August	37	0.1000	0.0965	1.000
BTC > \$26,000 end of August	36	−0.036	0.0494	0.389
BTC > \$27,000 end of September	214	0.0640	0.0531	0.682
Pooled panel	287	0.0642	0.0576	0.686

Notes. Trimmed mean drops top and bottom 10% of observations within each sample.

The cross-sectional structure of the wedge is informative about its source. Table **??** reports a pooled OLS regression of D_t on the option-implied benchmark, its standard error,

and time to expiry, with market fixed effects and HC1-robust standard errors. Three findings stand out.

Table 5: Cross-sectional determinants of the pricing gap

Variable	Coefficient	Robust SE	p-value
$P_{fair,t}$	-0.398	0.063	< 0.001
$SE(P_{fair,t})$	0.039	0.019	0.040
Time to expiry (hours)	0.0008	0.0001	< 0.001

Notes. Dependent variable: D_t. Pooled panel, market fixed effects, 287 observations. $R^2 = 0.221$.

First, the gap is strongly decreasing in the option-implied benchmark probability (coefficient -0.398, $p < 0.001$): the wedge is largest when the option assigns a low probability to the threshold event. This is consistent with demand-side overpricing of low-probability prediction-market outcomes, a pattern documented in laboratory and parimutuel markets but not previously tested with an option-implied benchmark [**?** **?**]. It is not consistent with a settlement-risk or counterparty-premium story, which would predict a larger wedge at intermediate probabilities where settlement uncertainty is highest.

Second, the gap is increasing in benchmark uncertainty (coefficient $+0.039$): this suggests that venues diverge more when the Binance anchor is itself less precisely pinned, consistent with noise-driven divergence. Third, and most sharply, the wedge grows with time to expiry (coefficient $+0.0008$, $p < 0.001$): contracts with more time remaining display a wider gap, shrinking as expiry approaches and the option price becomes more informative about the binary event.

Figure **??** visualizes these patterns across terciles of the option-implied benchmark and time to expiry.

(a) Mean gap across terciles.

(b) Pooled regression coefficients.

Figure 3: Robustness and cross-sectional patterns. Panel (a): mean discrepancy across terciles of the option-implied probability (left) and time to expiry (right). Panel (b): pooled regression coefficients with 95% confidence intervals.

5. Arbitrage-Style Proxy Backtest

To ask whether the wedge is economically exploitable rather than merely statistically detectable, I run a delta-hedged proxy strategy. Because the digital payoff is not directly traded on Binance, I replicate the binary benchmark's local Greeks using the matched call, spot, and a cash account. At entry time t_0, the call quantity $q_0 = \mathcal{V}^D_{t_0}/\mathcal{V}^C_{t_0}$ matches the benchmark's vega, the spot position $x_{t_0} = \Delta^D_{t_0} - q_0\Delta^C_{t_0}$ matches its delta, and the cash account closes the value gap. The spot hedge is rebalanced hourly while q_0 is held fixed.

Trades open when the absolute discrepancy exceeds one benchmark standard error plus the friction proxy, $h_t = SE(P_{fair,t}) + TF_t$, and close when the gap re-enters the neutral band $|D_t| \leq h_t$, or at expiry. Gross and net PnL per trade are

$$\Pi_i^{gross} = \eta_i \left(P_{poly,\tau_i} - P_{poly,t_i}\right) - \eta_i \left(V_{\tau_i} - P_{fair,t_i}\right), \tag{5.1}$$

$$\Pi_i^{net} = \Pi_i^{gross} - c_{i,\text{entry}}^P - c_{i,\text{exit}}^P - c_{i,\text{entry}}^C - c_{i,\text{exit}}^C - \sum_{t \in \mathcal{R}_i} c_t^S, \tag{5.2}$$

where $\eta_i \in \{1, -1\}$ is the sign of the Polymarket leg, $c_t^P = f_P + s_P/2$, $c_t^C = |q_0| C_t (f_B + s_{B,t}/2)$, and $c_t^S = |\Delta x_t| S_t \kappa_S$ with spot cost $\kappa_S = 5$ basis points one-way.

Table **??** summarizes the backtest results.

Table 6: Arbitrage-style proxy backtest

Market	Trades	Net PnL	Net α	p	95% HAC CI	Median hold (h)
BTC > \$28,000 Aug	4	0.274	0.221	0.103	[0.027, 0.416]	2.5
BTC > \$26,000 Aug	1	−0.345	—	—	—	1.0
BTC > \$27,000 Sep	11	1.183	0.018	0.092	[−0.003, 0.039]	6.0
Pooled panel	16	1.113	0.067	0.053	[−0.008, 0.143]	3.5

Notes. Alpha is defined as mean net return relative to deployed notional. HAC intervals use Newey–West. The August \$26,000 contract generates one trade; formal inference is not reported for it individually.

The pooled backtest generates 16 trades with gross PnL of 1.649, net PnL of 1.113 after all cost assumptions, net win rate of 69%, and median holding period of 3.5 hours. The pooled net alpha is 0.067 with $t = 2.10$, $p = 0.053$, and HAC interval $[-0.008, 0.143]$. The trading evidence is thus weaker than the pricing discrepancy evidence, for two reasons: the trade count is small, which limits power; and the August \$26,000 contract contributes a single losing trade that reduces pooled precision. Within the main September contract, net PnL is 1.183 and the alpha t-statistic is 1.86 ($p = 0.092$). Figure **??** plots cumulative gross and net PnL across trade exits.

Figure 4: Cumulative gross (blue) and net (red) PnL of the proxy backtest, plotted at trade-exit dates across the pooled sample.

The appropriate reading of these results is that the measured wedge does not vanish when one moves from statistical detection to a stylized trading exercise. At the same time, sixteen trades is not sufficient to make strong claims about economic significance in a frequentist sense. The backtest is best understood as an illustration of practical exploitability rather than evidence of a risk-adjusted anomaly.

6. Benchmark Extension: Deribit and Ethereum

6.1. Bitcoin against Deribit

To assess sensitivity to the options benchmark venue, I re-run the same three Bitcoin threshold markets against matched Deribit call options, holding the Polymarket contracts

fixed and changing only the options source. The Deribit extension uses the same delta-method pipeline; option marks are converted from BTC-denominated mid prices to USD using the concurrent perpetual price.

The Deribit results confirm the direction of the main finding but reveal stronger cross-market heterogeneity. In the September $27,000 contract, the mean discrepancy is 0.1248 over 645 aligned hourly observations (HAC CI: $[0.075,\ 0.174]$, $p = 2.5 \times 10^{-32}$). The pooled three-market Deribit panel shows a mean discrepancy of 0.1105 over 2,585 observations (HAC CI: $[0.074,\ 0.147]$, $p = 5.8 \times 10^{-64}$). At the market level, however, the August $28,000 contract is near zero under Deribit while the $26,000 and September $27,000 contracts remain strongly positive. Table **??** compares the Binance and Deribit specifications.

Table 7: Binance main specification vs. Deribit extension

Sample	Benchmark	n	Mean D	95% HAC CI	p-value
September $27,000	Binance	214	0.0558	$[0.023,\ 0.089]$	6.9×10^{-10}
September $27,000	Deribit	645	0.1248	$[0.075,\ 0.174]$	2.5×10^{-32}
Pooled panel	Binance	287	0.0633	$[0.033,\ 0.094]$	6.3×10^{-15}
Pooled panel	Deribit	2,585	0.1105	$[0.074,\ 0.147]$	5.8×10^{-64}

Notes. Deribit rows use the same three Polymarket Bitcoin contracts with Deribit calls and the Deribit perpetual reference price.

The useful interpretation is not that one benchmark dominates the other, but that the sign and economic relevance of the wedge are robust across option venues while the magnitude is sensitive to the choice of benchmark. This is consistent with Deribit and Binance option markets incorporating slightly different information, both of which differ systematically from the prediction-market pricing.

6.2. Ethereum exercise

I also assemble a smaller Ethereum exercise using four Polymarket contracts of the form "ETH above $1,700" from February–March 2023, matched to hourly Deribit ETH calls. The pooled four-market ETH panel shows a mean discrepancy of 0.0129 over 2,737 observations (HAC CI: $[-0.000,\ 0.026]$), statistically fragile at the pooled level. The key

finding is instability across contracts: the three February markets have slightly negative mean discrepancies, while the March 24 market shows a large positive wedge of 0.1899 (HAC CI: [0.145, 0.234]). This heterogeneity suggests that the Bitcoin finding does not extend uniformly to other assets and periods. The Ethereum exercise is best read as supplementary evidence about the scope of the phenomenon rather than a second main result.

7. Discussion

7.1. Scope conditions

The benchmark is model-based. Black–Scholes is useful as a clean, replicable benchmark but is not a literal description of cryptocurrency return dynamics. The main result should therefore be read as a deviation from the standard option-implied risk-neutral benchmark, not as a direct statement about physical probabilities. As noted in Section **??**, stochastic-volatility corrections would in general widen rather than narrow the gap, so this scope condition is conservative with respect to the direction of the finding.

The contract mapping is close but not exact. Polymarket contracts resolve under platform-specific oracle procedures, while the options benchmark is defined by exchange convention. This approximation is economically natural for Bitcoin threshold events — the oracle simply reads the Coinbase or CMC closing price, which tracks Binance closely — but it is not identical to a listed exchange payoff. A more granular settlement-data analysis is left for future work.

The sample is intentionally narrow. Only three Binance-compatible Bitcoin contracts meet the exact-data criteria. This restriction is a feature, not a bug, for the main identification: it eliminates contract-mapping ambiguity entirely. But it does limit external validity. The Deribit Bitcoin extension and the Ethereum exercise provide qualitative breadth at the cost of less precise data alignment.

7.2. Interpretation of the wedge

Three non-mutually-exclusive interpretations are consistent with the evidence.

Demand-side overpricing. The cross-sectional regression shows the wedge is largest at low option-implied probabilities and long maturities. This pattern is the signature of retail or sentiment-driven demand for prediction-market contracts: participants overweight the probability of unlikely events relative to the option-market consensus [**?**]. The analogy to longshot bias in parimutuel markets [**?**] is instructive: prediction-market contracts for low-probability outcomes attract disproportionate speculative interest, bidding prices above risk-neutral benchmarks.

Segmented capital and slow arbitrage. The four-hour half-life is the signature of limits to arbitrage. Capital that could equalize prices across Binance and Polymarket is either absent (most participants in each venue do not participate in the other) or costly to deploy (the delta-hedged proxy strategy requires simultaneous positions in an options market, a spot market, and a prediction-market platform). The persistence is consistent with models of slow-moving capital [**?**] and segmented-market pricing [**?**].

Risk-neutral measure differences. If the two platforms embed different risk adjustments — for example, if Polymarket traders are risk-neutral over prediction-market payoffs but option-market participants price variance risk — then a systematic gap between the venues would persist even in frictionless equilibrium. This interpretation requires a structural model of cross-market preference heterogeneity and is left for future work.

7.3. Directions for future research

Three extensions are most promising. First, extending the contract universe to later Kalshi data and a credible Solana options benchmark would allow a richer cross-asset comparison; the current constraints are data availability, not empirical design. Second, quote-level spot and order-book data would sharpen the microstructure interpretation of the half-life. Third, testing whether the wedge narrows following large cross-venue traders or market makers' entry into prediction markets would provide a direct test of the segmented-capital interpretation.

8. Conclusion

This paper introduces an option-implied benchmark test of prediction-market pricing and applies it to Bitcoin threshold contracts on Polymarket and Binance. The empirical design is deliberately narrow: recover implied volatility from a matched vanilla call, map it into a risk-neutral binary benchmark, propagate estimation uncertainty through the delta method, adjust for transaction costs, and test whether the prediction-market price differs systematically from the option-implied value.

The results are clear and consistent. In the main September 2023 Bitcoin threshold contract, the mean discrepancy is 5.6 percentage points across 214 hourly observations ($p < 10^{-9}$). In the pooled three-market panel, the mean discrepancy is 6.3 percentage points across 287 observations ($p < 10^{-14}$). Outlier-robust summaries based on medians, trimmed means, and sign tests point in the same direction. The gap is largest at low option-implied probabilities and long maturities — a cross-sectional signature consistent with speculative demand for prediction-market contracts rather than mechanical measurement error.

The wedge is persistent but mean-reverting. An AR(1) half-life of 4.2 hours combined with rejection of a unit root is the hallmark of a genuine pricing gap that is exploitable in principle but not arbitraged away instantaneously. A delta-hedged proxy backtest generates positive net PnL after conservative transaction costs, though with marginal statistical precision given the small number of trades. The Deribit extension preserves the direction and economic magnitude of the finding across both contracts and both benchmark venues, while the Ethereum exercise delivers more mixed cross-contract evidence.

The main takeaway is institutional as much as statistical. Two markets that price, in principle, the same state-contingent payoff do not agree on its value, and the disagreement is structured: it is largest where speculative pressure on prediction markets is plausibly strongest. This finding has implications for how prediction-market prices should be interpreted as probability estimates and for how researchers and market participants understand the relationship between digital prediction platforms and centralized derivatives markets.

Appendix A. Analytical Derivations

Proposition Appendix A.1 (Risk-neutral valuation of the binary payoff)**.** *Let S follow Black–Scholes dynamics under the risk-neutral measure, $dS_t/S_t = r\,dt + \sigma\,dW_t^{\mathbb{Q}}$, with constant r and σ. The time-t value of the unit payoff $\mathbf{1}\{S_T > K\}$ at maturity T is*

$$P_{fair,t} = e^{-r\tau}\Phi(d_2), \qquad \tau = T - t, \tag{A.1}$$

where $d_2 = \Big[\log(S_t/K) + (r - \sigma^2/2)\tau\Big]/(\sigma\sqrt{\tau})$.

Proof. Under $\mathbb{Q}$, $\log S_T = \log S_t + (r - \sigma^2/2)\tau + \sigma\sqrt{\tau}Z$ with $Z \sim N(0,1)$. Therefore $\mathbb{Q}(S_T > K \mid \mathcal{F}_t) = \Phi(d_2)$. Discounting gives $e^{-r\tau}\mathbb{E}^{\mathbb{Q}}[\mathbf{1}\{S_T > K\} \mid \mathcal{F}_t] = e^{-r\tau}\Phi(d_2)$. □

Proposition Appendix A.2 (Black–Scholes vega)**.** $\mathcal{V}(S_t, K, r, \tau, \sigma) = \partial C_{BS}/\partial\sigma = S_t\phi(d_1)\sqrt{\tau}$.

Proof. Differentiate C_{BS} with respect to σ: $\partial C_{BS}/\partial\sigma = S_t\phi(d_1)\partial d_1/\partial\sigma - Ke^{-r\tau}\phi(d_2)\partial d_2/\partial\sigma$. Using $S_t\phi(d_1) = Ke^{-r\tau}\phi(d_2)$ and $\partial d_1/\partial\sigma - \partial d_2/\partial\sigma = \sqrt{\tau}$ yields the result. □

Proposition Appendix A.3 (Volatility sensitivity of the binary price)**.** *Let $g(\sigma) = e^{-r\tau}\Phi(d_2(\sigma))$. Then $g'(\sigma) = -e^{-r\tau}\phi(d_2)d_1/\sigma$.*

Proof. By the chain rule, $g'(\sigma) = e^{-r\tau}\phi(d_2)\,\partial d_2/\partial\sigma$. Differentiating d_2 gives $\partial d_2/\partial\sigma = -[\log(S_t/K) + r\tau]/(\sigma^2\sqrt{\tau}) - \sqrt{\tau}/2$. Since $d_1/\sigma = [\log(S_t/K) + r\tau]/(\sigma^2\sqrt{\tau}) + \sqrt{\tau}/2$, one obtains $\partial d_2/\partial\sigma = -d_1/\sigma$, so $g'(\sigma) = -e^{-r\tau}\phi(d_2)d_1/\sigma$. □

Proposition Appendix A.4 (First-order volatility inversion and delta-method variance)**.** *Let $\hat{\sigma}_t$ solve $C_{BS}(S_t, K, r, \tau_t^{(c)}, \hat{\sigma}_t) = C_{mkt,t}$ with $\mathcal{V}_t \neq 0$. Then $d\hat{\sigma}_t \approx dC_{mkt,t}/\mathcal{V}_t$ and* $\mathrm{Var}(\hat{\sigma}_t) \approx \mathrm{Var}(C_{mkt,t})/\mathcal{V}_t^2$. *Consequently,* $\mathrm{Var}(P_{fair,t}) \approx (g'(\hat{\sigma}_t))^2\,\mathrm{Var}(\hat{\sigma}_t)$.

Proof. A first-order perturbation of $F_t(\hat{\sigma}_t) = 0$ gives $\mathcal{V}_t\,d\hat{\sigma}_t = dC_{mkt,t}$; rearranging and taking variances yields the inverse-vega formula. The delta-method step follows by the chain rule applied to $g(\hat{\sigma}_t)$. □

Proposition Appendix A.5 (Tick-noise contribution to discrepancy variance)**.** *If* $P^{obs}_{poly,t} = P^{\star}_{poly,t} + \varepsilon_t$ *with* $\varepsilon_t \sim \mathrm{Unif}[-\Delta_P/2, \Delta_P/2]$ *independent of the option-implied pricing error, then* $\mathrm{Var}(D_t) \approx \mathrm{Var}(P_{fair,t}) + \Delta_P^2/12$.

Proof. For $\varepsilon_t \sim \mathrm{Unif}[-a, a]$, $\mathbb{E}[\varepsilon_t] = 0$ and $\mathrm{Var}(\varepsilon_t) = a^2/3 = \Delta_P^2/12$. Independence gives $\mathrm{Var}(D_t) = \mathrm{Var}(P^{\star}_{poly,t} - P_{fair,t}) + \Delta_P^2/12$. Treating the latent prediction-market component as locally fixed relative to $P_{fair,t}$ yields the stated approximation. □