Fair Value Pricing in Short Dated Bitcoin Binary Markets: A Single Window Evaluation of a Candidate Statistical Arbitrage Edge

A single window evaluation of a fair value pricing lens and its limits in fast settling crypto prediction markets

Jackson Semenas

JS Financials
jackson.semenas@gmail.com

July 2026

Abstract

Short dated Bitcoin binary contracts settle over horizons brief enough that expected drift is negligible, which reduces their fair value to a function of current spot and short horizon volatility alone. This paper frames the contract as a cash or nothing digital option, derives a driftless fair value, and defines an entry rule that acts when the implied price and fair value differ by at least a fixed buffer, sizing by a model edge score. On 182 resolved contracts traded across 22 to 23 June 2026, the rule realised a positive statistical edge, a win rate of 67.6 percent against the 63 percent break even rate implied by the mean entry price of 0.63, an edge of 4.6 percentage points. The point estimates are positive across the headline measures, but the sample is too small to establish significance: bootstrap intervals on the edge include zero, and the model fair value does not improve on the market price as a probability forecast. The realised gain is concentrated in a small number of trades and coincides with a directional position, and an edge threshold selection rule does not persist out of sample. The result is interpreted as a statistical edge documented within a single window, whose significance and source, a genuine pricing signal against the established favorite longshot bias, remain open questions pending a larger sample. The contribution is the pricing framework and its evaluation.

Keywords: prediction markets, Bitcoin, binary options, digital option pricing, statistical arbitrage, market microstructure, favorite longshot bias, calibration, market efficiency
JEL classification: G12, G13, G14, C58, D84

1. Introduction

Prediction markets have attracted sustained academic attention, and the central finding is that market generated forecasts are typically well calibrated and hard to beat (Wolfers and Zitzewitz, 2004; Snowberg, Wolfers and Zitzewitz, 2013). That literature concentrates almost entirely on political and event markets that resolve over days, weeks, or months, and even the interpretation of a contract price as a probability rests on assumptions about trader preferences that need not hold (Manski, 2006). Fast settling crypto binaries occupy a different regime. A contract that pays one unit if Bitcoin finishes above a strike over a short window is, mechanically, a cash or nothing digital option written on a liquid underlying. Classical derivatives pricing applies directly (Black and Scholes, 1973; Merton, 1973), yet the venue is populated by flow that does not price the contract as an option. That mismatch is the subject of this paper.

The paper specifies one framework and evaluates it. The framework holds that fair value for these contracts collapses, over the relevant horizon, to a function of spot and short horizon volatility, because expected drift over minutes is dominated by volatility, and that a rule trading the gap between fair value and the implied price earns a positive return where the venue is mispriced. A live sample is then subjected to a sequence of standard tests. The strategy realises a positive edge within the sample, but the sample does not permit strong inference: the realised edge is positive yet not statistically separable from zero, the model does not outperform the market price as a forecast, the gain concentrates in a small number of trades and coincides with a directional exposure, and an edge selection rule does not persist out of sample. The contribution is the pricing framework and its evaluation on live data.

The source of the inefficiency is structural. Retail dominated flow trades direction and sentiment rather than modelled probability, while latency and capital frictions, including settlement cost on the underlying network,

deter the arbitrageurs who would compress the gap, as the limits to arbitrage literature predicts when arbitrage requires capital and carries risk (Shleifer and Vishny, 1997). Cryptocurrency markets in particular sustain large and recurrent arbitrage deviations for this reason (Makarov and Schoar, 2020), a departure from the efficient markets benchmark (Fama, 1970).

2. Venue and Contract Microstructure

The instrument studied is a binary contract on the price of Bitcoin resolving over a short fixed window. At resolution the contract pays a fixed amount if the underlying finishes on the specified side of a strike and zero otherwise. The traded price of the contract, bounded between zero and one after normalisation, is the market implied probability of finishing in the money. This equivalence between price and probability is what makes the venue directly comparable to a modelled fair value.

The cost stack matters because it sets the threshold below which no divergence is tradable. It comprises the bid ask spread on the contract, the settlement and transaction cost on the underlying network (Nakamoto, 2008), and any slippage between the quoted and realised entry. Liquidity and its associated price impact are first order here, consistent with evidence that illiquidity is priced (Amihud, 2002) and that cryptocurrency venues exhibit distinctive liquidity and price discovery dynamics (Aleti and Mizrach, 2021; Barbon and Ranaldo, 2021). This file does not isolate the friction components directly, so rather than assert a cost figure we treat the combined cost as whatever the buffer must clear and let the empirical return in Section 5 speak to it directly. No claim in this paper depends on any detail of how the venue is accessed beyond this cost abstraction.

3. Fair Value Framework

Consider the contract as a cash or nothing digital call with strike K, current spot S, volatility sigma over the horizon, and time to resolution T expressed as a fraction of the volatility unit. Under a standard lognormal assumption for the underlying, the fair value of the contract, meaning the risk neutral probability of finishing in the money, is N(d2), where N is the standard normal cumulative distribution function. This is the digital option term of the Black and Scholes (1973) formula, which Merton (1973) placed on a rigorous footing, and it can equivalently be obtained as a limiting case of the binomial approach of Cox, Ross and Rubinstein (1979).

The simplification that makes this venue tractable is the horizon. Over minutes, the drift term in d2 is negligible relative to the volatility term, because expected return scales with T while the standard deviation of return scales with the square root of T. Setting the drift contribution to zero, d2 reduces to the log moneyness scaled by volatility and the square root of time. Fair value is then a function of spot and short horizon volatility alone. The quality of the volatility input is therefore the binding constraint, which situates the problem in the realized volatility literature built on high frequency data (Barndorff-Nielsen and Shephard, 2002; Andersen, Bollerslev, Diebold and Labys, 2003), though the specific estimator used here is out of scope. The reference implementation below computes the driftless fair value. It is generic: it takes spot, strike, volatility, and horizon as arguments and makes no commitment to how any of those inputs are estimated.

```
from math import log, sqrt, erf

def norm_cdf(x):
    # standard normal CDF via the error function
    return 0.5 * (1.0 + erf(x / sqrt(2.0)))
```

```
def binary_fair_value(spot, strike, sigma, horizon):
    # cash or nothing digital call, driftless short horizon reduction.
    # sigma is volatility over the horizon unit; horizon is time to
    # resolution in the same unit. returns P(finish in the money).
    if horizon <= 0 or sigma <= 0:
        return 1.0 if spot > strike else 0.0
    d2 = log(spot / strike) / (sigma * sqrt(horizon))
    return norm_cdf(d2)

def signed_edge(fair_value, implied):
    # positive when the model prices the contract richer than the market
    return fair_value - implied
```

This constitutes the complete pricing core; it is standard and reproducible. What separates a working system from the function above is the quality of the spot and volatility estimates supplied to it, and that estimation is out of scope here. The analysis rests on the pricing framework and the market structure argument, not on any proprietary estimation.

4. Entry Rule and Expected Value

Let f denote model fair value and p the market implied price. The signed divergence is f minus p: when f exceeds p the market underprices the contract and the capturing position is to buy it, and when p exceeds f the reverse holds. Gross expected value is proportional to the absolute divergence, since the contract pays at the model probability in expectation while entered at the implied price.

In deployment the divergence is not treated as a continuous dial. A trade is admitted only when the absolute gap clears a fixed buffer, set here at two points, which at minimum covers the cost stack and absorbs estimation error in f. That buffer is an on off gate rather than a sizing signal, so among admitted trades the divergence is constant by construction. What varies across admitted trades, and what the model uses to size and rank, is a scalar edge score that combines the cleared divergence with the model's confidence in f. The empirical question is therefore whether realised return scales with that edge score, which Section 5 tests directly. The edge score is reported here only as an anonymised quantity; its construction is out of scope.

5. Empirical Results

This section reports evidence from 182 resolved contracts traded across 22 to 23 June 2026. The rule entered at a mean price of 0.63. Because a binary paying one unit costs its own win probability at fair value, that mean price sets the break even win rate at 63 percent: below it the position loses in expectation, above it the position gains. The realised win rate was 67.6 percent, so the strategy earned a positive edge of 4.6 percentage points over break even within the sample, equivalently a mean realised minus implied alpha of 0.046 per contract. The corresponding return on deployed capital over the window was 8.8 percent. These are in sample outcomes over a single two day window; they document the edge realised in the window rather than establish it as a stable property, a distinction carried through the interpretation below.

5.1 Divergence structure

Figure 1 plots the market implied price against model fair value for each contract. Because entry requires the implied price and fair value to differ by at least the fixed two point buffer, admitted trades do not scatter across the diagonal; they sit as a band offset from it by the buffer, which is the signature of the on off entry gate rather

than of a continuously varying divergence. Fair value across admitted trades spans 0.37 to 0.89 with a mean near 0.61, so the gate fires across a wide range of moneyness rather than clustering at one price.

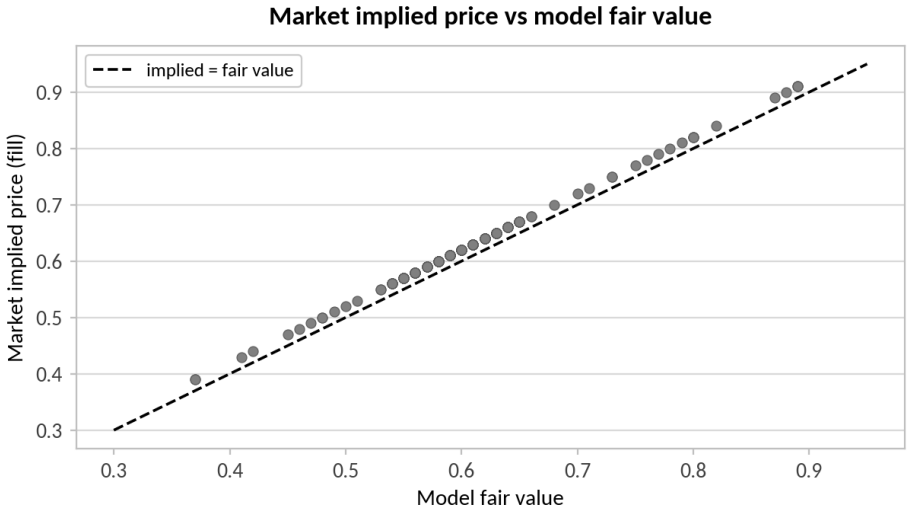

Figure 1. Market implied price against model fair value across the 182 admitted trades. The offset from the dashed diagonal is the fixed entry buffer.

5.2 Return and the edge threshold

Because divergence is fixed at the buffer for every admitted trade, the quantity that discriminates outcomes is the model edge score. On the full sample, raising the minimum edge requirement appears to lift return per dollar, roughly doubling from the full book toward the higher edge trades. That pattern is exactly the kind a threshold sweep manufactures by chance, so rather than report it at face value the sample is split in time into a training half and a held out test half, and the threshold is judged only on the half it was not chosen on.

Figure 2 reports the result. On the training half the return climbs with the edge requirement as expected, and the best training threshold near 0.065 returns about 32 percent per dollar in sample. On the held out half the same threshold returns only about 10 percent, below the roughly 15 percent that taking every admitted trade returns out of sample. The out of sample curve does not track the in sample one, and requiring more edge performs worse than requiring none. The edge threshold therefore does not generalise on this data: the apparent gain from selecting high edge trades is an in sample artifact, and the analysis does not rely on it. What remains is the weaker statement that admitted trades as a group returned positively in the window.

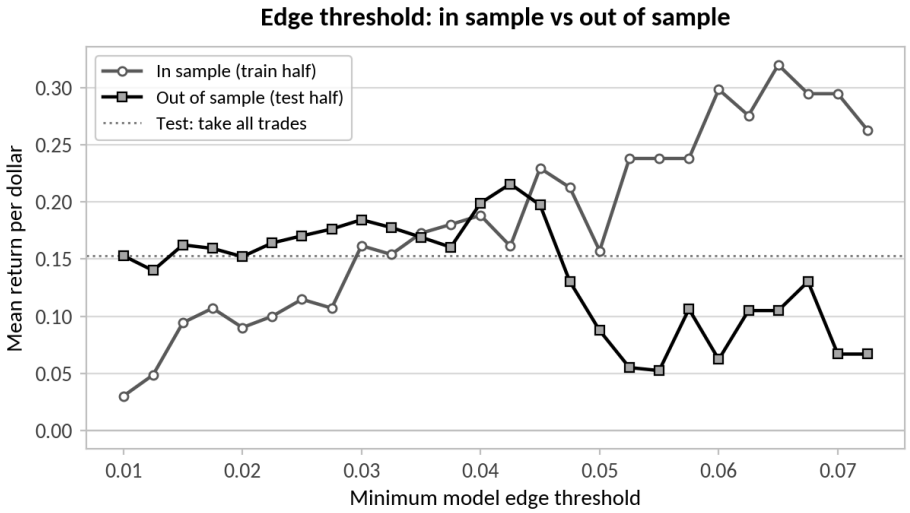

Figure 2. Mean return per dollar against the minimum edge threshold, computed on a training half and a held out test half. The test curve does not follow the training curve, and the dotted line marks the test return from taking all trades.

5.3 Fair value distribution

Figure 3 shows the distribution of model fair value across the admitted trades. The mass sits between roughly 0.55 and 0.65 with a mean near 0.61 and a tail reaching toward 0.9, indicating that most admitted contracts are moderately directional rather than near certain, and that the gate is not simply selecting lopsided contracts where a price is already close to one or zero.

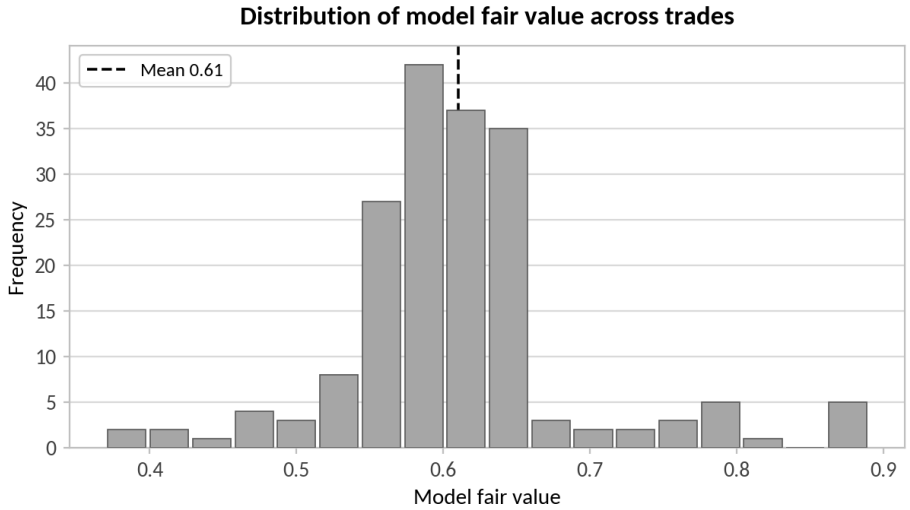

Figure 3. Distribution of model fair value across the 182 admitted trades.

5.4 Calibration and the model versus market comparison

The premise of the strategy is that model fair value is a sharper probability than the market price. That premise is testable directly and should not be assumed. Figure 4 is a reliability diagram: model fair value on the horizontal axis against the realised win rate in each fair value bin on the vertical, with 95 percent Wilson intervals and the diagonal marking perfect calibration. Three of the four bins sit within confidence of the

diagonal. The exception is the bin near 0.57, where the realised win rate of 0.72 lies above its interval, indicating the model was underconfident there. On this sample the model is neither cleanly calibrated nor grossly broken; it is close to the diagonal with one significant local departure and wide intervals reflecting the small per bin counts.

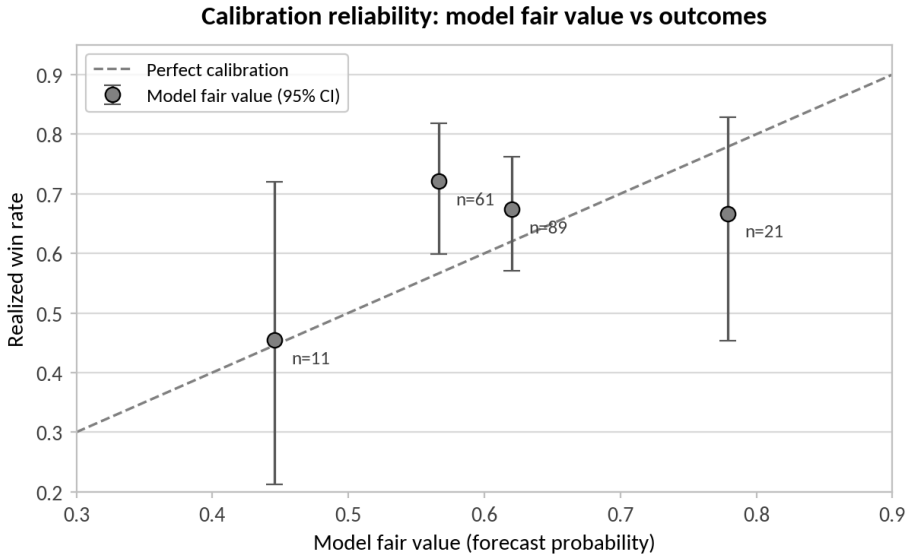

Figure 4. Calibration reliability of model fair value against realised outcomes. Marker area scales with bin count; bars are 95 percent Wilson intervals.

The stronger test compares the model and the market head to head as forecasters of the realised outcome. Scored on the traded book, the model fair value earns a Brier score of 0.227 against the market price at 0.225, and the log score favours the market by a similar slim margin. The paired bootstrap interval on the Brier difference includes zero, so the gap is not statistically meaningful, but the point estimate runs the wrong way for the premise: the market price is, if anything, the marginally better forecast on these trades. It follows that the realised edge in Sections 5.1 to 5.3 does not arise from the model being a globally superior probability estimate. Any edge present is confined to the selected book rather than explained by better calibration overall, which narrows the interpretation the sample supports.

5.5 Statistical assessment

Point estimates are not evidence on their own, so every headline quantity is reported below with a 95 percent bootstrap interval from 20,000 resamples. Because the trades cluster in time, the intervals were also recomputed with a two hour time block bootstrap that resamples whole blocks rather than individual trades. The two agree to within a fraction of a point, because the lag one autocorrelation of per trade return is near zero and the effective sample size is close to the nominal 182. The clustering, in other words, does not inflate the precision here, so the intervals below can be read at face value.

Metric	Estimate	95% interval	Reading
Win rate	67.6%	[61.0, 74.2]%	lower bound below 63% break even
Return on stake	8.8%	[-2.8, +20.3]%	interval spans zero
Per trade alpha	+0.046	[-0.023, +0.114]	interval spans zero
Return per dollar	9.1%	[-2.4, +20.7]%	interval spans zero
Edge return slope	2.42	t = 1.81, p = 0.07	marginal, not significant at 5%

Table 1. Headline quantities with 95 percent bootstrap intervals. Intervals are the iid resample; the two hour block bootstrap gives materially identical bounds.

The pattern is consistent across every row. The point estimates are positive, and the realised win rate exceeds its break even in the sample, but no headline quantity is separable from zero at the 5 percent level. The edge return slope, the closest to significance, carries a p value near 0.07. This reflects the limited power of a two day sample: 182 binary outcomes cannot resolve an edge of a few points per trade at conventional confidence levels, independent of its true magnitude. The sample is thus consistent with a positive edge and with no edge, and a larger sample is required to distinguish the two.

5.6 Directional decomposition

A result concentrated in a single window admits an alternative explanation, which is examined here directly. Two thirds of the book, 121 of 182 trades, was on the Down side, and the return differs substantially by side. Figure 5 shows the Down trades returning about 14.6 percent on stake with a 70.2 percent win rate, and the Up trades returning a negative 2.5 percent with a 62.3 percent win rate. All realised profit in the window originates on the Down side; the Up side lost. The bootstrap intervals qualify this: the Down return only just excludes zero at its lower bound, the Up interval spans zero, and the difference in per trade alpha between the two sides, though nearly nine points in point estimate, has an interval that spans zero.

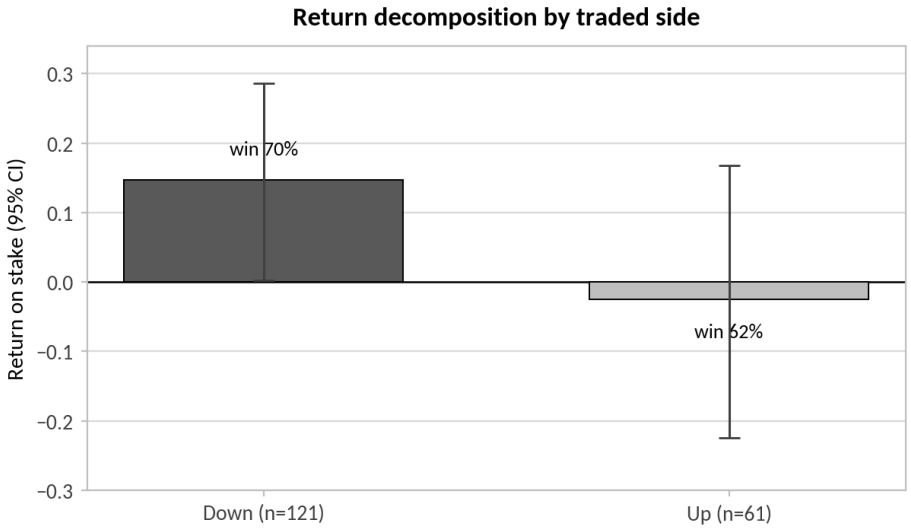

Figure 5. Return on stake by traded side with 95 percent bootstrap intervals. Win rate annotated per side. All realised profit originates on the Down side.

This constitutes a material confound. Over a two day window in which the book was net short, a Down bias and a falling underlying cannot be separated from a Down side pricing edge using this sample alone. The realised return is consistent with the model identifying genuine mispricing on the short side, and equally consistent with a directional short position that paid in a down move. Distinguishing the two requires either a fair value series that isolates the model signal from the underlying return, which this dataset does not contain, or a sample spanning both directions of the market, which a single window does not provide. On this sample the edge is therefore entangled with direction, and no stronger claim is made.

5.7 Benchmark nulls

A return means little without something to beat. Two nulls are natural on this venue: entering a random side at the quoted price, and mechanically buying the market favorite on every contract the strategy touched. Figure 6 places the strategy against both. Random entry returns essentially zero, as it must when contracts are priced near fair, so the strategy exceeds that benchmark. The favorite is the stronger null: buying the favorite on the same contracts returns 6.6 percent per dollar on its own, against the strategy's 9.1 percent. The strategy selects the favorite side on 95 percent of its trades, so it is, to first order, a favorite buying rule with a selective overlay.

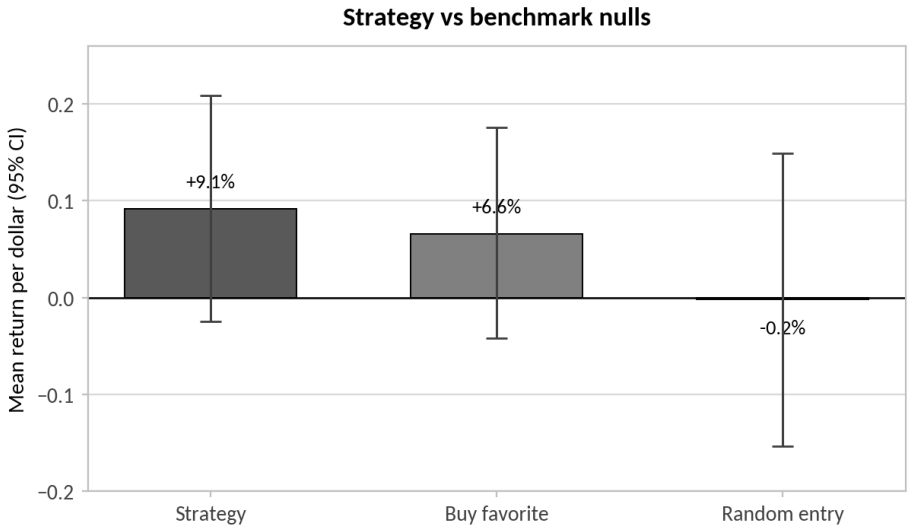

Figure 6. Mean return per dollar for the strategy against a random entry null and a buy the favorite null, with 95 percent bootstrap intervals.

The increment the strategy adds over favorite buying is 2.5 points per dollar, with a bootstrap interval from negative 4.7 to positive 10.0 points that spans zero. This locates the finding. The majority of the window's return is the payoff to holding favorites while the market trended, a directional component rather than a pricing effect, and the model's selective contribution above it is small and not statistically established in this sample. The strategy differs from random entry, but on this evidence it is not distinguishable from a favorite buying rule.

5.8 Concentration and robustness

A positive aggregate can rest on a small number of outcomes, so the final check is how the profit is distributed. It is heavily concentrated. The five best trades account for 42 percent of total profit and the ten best for 69 percent; beyond the top twenty the remaining trades are net negative as a group. Of 182 trades, 123 won and 59

lost. Removing the ten best trades reduces return on stake from 8.8 percent to 2.9 percent, and removing the seventeen best, under a tenth of the book, eliminates the profit entirely. The realised edge is located in a small subset of the sample rather than in a broad tendency across trades. This is the expected shape for a small sample of high variance binary payoffs, but it indicates that the window's return is sensitive to a few outcomes and cannot be treated as a stable per trade effect.

6. Conditions and Limitations

The edge documented here is conditional, and the conditions are stated explicitly. The mechanism is fragile along four axes. Competition compresses it as other participants price the contracts as options. Latency erodes it, since the edge assumes entry near the quoted price. The cost stack caps it, as any rise in spread, settlement cost, or slippage lifts the break even and removes the marginal trades first. And volatility regime governs it, because the driftless reduction is most accurate when short horizon volatility is stable and well estimated.

These are general statements about where the mechanism holds and where it does not, offered as forward looking caveats rather than as a report on any particular deployment. That the edge should compress as arbitrage capital arrives is the central prediction of Shleifer and Vishny (1997), and the persistence of crypto arbitrage documented by Makarov and Schoar (2020) suggests the frictions that sustain it can also decay slowly. Venue design matters too: the mechanism by which prices are formed and quotes are subsidised shapes how quickly divergence is competed away (Hanson, 2007). A reader should treat the empirical results as evidence that the divergence existed in the sample and that the pricing lens explains it, not as a claim that the edge is permanent or invariant to the conditions above.

7. Related Work and Interpretation

Two literatures reframe what the window most likely captured. Makarov and Schoar (2020) document large, recurrent arbitrage deviations in cryptocurrency prices that transaction costs alone cannot explain and that widen during large bitcoin appreciations, sustained by capital controls and slow moving arbitrage capital. Their mispricing is a cross venue price level dislocation rather than a derivative gap, but the shared lesson is limits to arbitrage, and the detail that their deviations co move with directional bitcoin moves bears on the confound in Section 5.6: crypto mispricing measured over a directional window is difficult to separate from the direction itself.

The prediction market literature is more directly relevant. The favorite longshot bias, the regularity that favorites are underbet and therefore underpriced while longshots are overbet, is documented across betting markets by Snowberg and Wolfers (2010) and attributed to systematic misperception of probability, and recent venue level evidence reports the same pattern on Polymarket, with high probability outcomes underpriced and longshots overpriced. This is relevant because the strategy selected the market favorite on 95 percent of its trades, and buying favorites mechanically returned 6.6 percent over the window. The realised return is therefore consistent with the established favorite longshot bias as much as with a distinct fair value edge, and the calibration result, in which the model does not forecast outcomes better than the market price, does not distinguish the two. The source of the in sample edge remains an open question that a larger sample would resolve.

8. Conclusion

Short dated Bitcoin binaries are options in all but name, and over their brief horizon their fair value reduces to a spot and volatility problem, which makes them a suitable venue for a pricing study. Applied to a live sample, the strategy realised a positive statistical edge within the window, a 4.6 percentage point advantage over the break even win rate. The sample does not permit stronger inference: the edge is not statistically separable from zero, is concentrated in a small number of trades, coincides with a directional exposure, does not persist under an out of sample selection rule, and cannot be distinguished from the established favorite longshot bias, since the model does not forecast better than the market price. The contribution is the pricing framework and its evaluation on live data. Establishing whether the edge is durable, and whether its source is a distinct pricing signal, requires a larger sample spanning both market directions; nothing in the framework depends on any specific estimation method.

References

Aleti, S. and Mizrach, B. (2021). Bitcoin spot and futures market microstructure. Journal of Futures Markets, 41(2), 194 to 225.

Amihud, Y. (2002). Illiquidity and stock returns: cross section and time series effects. Journal of Financial Markets, 5(1), 31 to 56.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579 to 625.

Barbon, A. and Ranaldo, A. (2021). Cryptocurrency liquidity and market microstructure. Swiss Finance Institute Research Paper No. 21 to 45.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B, 64(2), 253 to 280.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81(3), 637 to 654.

Cox, J. C., Ross, S. A. and Rubinstein, M. (1979). Option pricing: a simplified approach. Journal of Financial Economics, 7(3), 229 to 263.

Fama, E. F. (1970). Efficient capital markets: a review of theory and empirical work. Journal of Finance, 25(2), 383 to 417.

Hanson, R. (2007). Logarithmic market scoring rules for modular combinatorial information aggregation. Journal of Prediction Markets, 1(1), 3 to 15.

Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics, 135(2), 293 to 319.

Manski, C. F. (2006). Interpreting the predictions of prediction markets. Economics Letters, 91(3), 425 to 429.

Merton, R. C. (1973). Theory of rational option pricing. Bell Journal of Economics and Management Science, 4(1), 141 to 183.

Nakamoto, S. (2008). Bitcoin: a peer to peer electronic cash system. White paper.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. Journal of Finance, 52(1), 35 to 55.

Snowberg, E. and Wolfers, J. (2010). Explaining the favorite longshot bias: is it risk love or misperceptions? Journal of Political Economy, 118(4), 723 to 746.

Snowberg, E., Wolfers, J. and Zitzewitz, E. (2013). Prediction markets for economic forecasting. In Handbook of Economic Forecasting, Vol. 2, 657 to 687. Elsevier.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. Journal of Economic Perspectives, 18(2), 107 to 126.