

Sample Test Report: Daily BTC Time-Series Momentum (TSMOM_D_V1)

Project Anchor · independent strategy test · **Verdict: FAIL**

Test date: 2026-10-24 · Report reissued as a sample: 2026-10-27 · Not investment advice.

This is a real test we ran on one of our own candidate strategies. It is published as an example of what a tester receives. It is a useful example because the strategy was profitable in-sample and out-of-sample, after fees, and it still failed.

1. Strategy as submitted

The source was external: the published time-series momentum idea, in a simple daily form. These are the rules as written before any run.

- **Instrument:** Kraken PF_XBTUSD (BTC perpetual). The signal uses daily bars, built in UTC from completed 15-minute candles.
- **Signal:** at each daily close, compare today's close with the close L days ago. If the return is positive, go long. If it is negative, go short.
- **Entry:** the next 15-minute open after the daily close. Only one position is open at a time.
- **Exit:** after exactly 7 days (a time exit), or earlier at a protective stop of $2 \times$ daily ATR(14). The stop is never tighter than 0.35% of entry. There is no take-profit.
- **Variants:** $L \in \{10, 30, 60\}$ days. The variant was chosen on in-sample data only, by highest in-sample net mean R among variants with at least 30 in-sample trades.

This family was one of 3 families tested in the same batch, with 7 configurations in total ($3 + 2 + 2$).

2. Pre-registered pass bar (written before running)

The rules file (PARAMS.md) was locked on **2026-10-24 at 12:46:04 (UTC+8)**, before any backtest ran. Its SHA-256 is:

4657ac56e6f5c8eacec015fed9ff07d9f85ba84d06527695ed8d897c9a9c7b39

The test script recomputes this hash and refuses to run if the file has changed. A strategy passes only if all of the following hold:

#	Criterion
Gate	In-sample mean gross move captured per trade $\geq 2 \times$ the cost wall. The cost wall is 8 bp round-trip commission plus 0.033%/day $\times$ the mean holding days. If the gate fails, the verdict is FAIL with no out-of-sample evaluation.
P1	Out-of-sample mean R, net of costs, > 0
P2	Out-of-sample trades ≥ 30
P3	Out-of-sample mean R $>$ the 95th percentile of a random-direction control. The control uses the same decision times and the same exits, over 200 random seeds.
P4	In-sample mean R > 0 , and the gate passed
P5	Anchor-specific: the simulated probability of passing our own \$10k prop evaluation, at \$100 risk and using the out-of-sample trades, must be $\geq$ our current baseline + 5 percentage points. In a test for you, this criterion is replaced by one you and we agree on before the run.

3. Data and costs

- **Data:** Kraken PF_XBTUSD public 15-minute and 1-hour candles. Daily bars are aggregated in UTC from completed bars only.
- **In-sample (IS):** 2022-10-03 $\rightarrow$ 2025-04-01 UTC. **Out-of-sample (OOS):** 2025-04-01 $\rightarrow$ 2026-10-23 UTC. The OOS period was run once.
- **Costs:** commission of 0.04% per side, plus a 0.033% per day margin fee, accrued every 4 hours while a position is open.
- **Fills:** entries fill at the next 15-minute open. Within a bar, the stop is checked first. If price gaps through the stop, the fill is at the worse open price. No separate slippage term was modelled.

- **R:** each trade is sized so that a stop-out, including round-trip commission, costs 1R. All R figures below are net of costs.

4. Results: in-sample vs out-of-sample (net of costs)

Variant	Period	Trades	Mean R	Total R	Win rate	Mean gross move	Gate (needs $\geq 2\times$ wall)
L30 (selected)	IS	140	+0.108	+15.2	50%	68.6 bp	68.6 vs 58.4 bp ✓
L30 (selected)	OOS	88	+0.046	+4.0	50%	42.3 bp	–
L10	IS	142	+0.041	+5.9	46%	29.3 bp	29.3 vs 57.8 bp ✗
L60	IS	139	+0.100	+13.9	50%	93.2 bp	93.2 vs 58.1 bp ✓

The mean holding time was about 6.4 days, because some trades were stopped out before day 7. OOS figures for the variants that were not selected were computed for transparency only and were not used for selection: L10 had 87 trades at +0.011R, and L60 had 88 trades at +0.025R. The in-sample edge roughly halved out-of-sample.

5. Random-entry baseline

The question is whether the *direction* signal adds anything beyond the exit template and the market's own drift and volatility. To test this, we replaced the long/short decision with a coin flip. Everything else stayed the same: the decision times, the 7-day hold, the $2\times$ ATR stop and the costs. The participation rate was matched to the strategy's trade count, so the probability of taking each decision bar was 0.989. We ran the OOS period 200 times with different seeds.

	OOS mean R per trade
Random direction, mean over 200 seeds	+0.002
Random direction, 95th percentile	+0.139
Strategy (L30)	+0.046

The strategy's +0.046R lies well inside the range that random-direction trading produced with the same exits.

6. Verdict and why

Gate	P1	P2	P3	P4	P5
✓	✓	✓	✗	✓	✓

FAIL on P3. The strategy made money after costs in both periods, but not by more than luck plausibly explains. Over 88 trades, a coin flip with these exits beat +0.046R often enough that we cannot attribute the result to the signal.

P5 passed (45.5% against a 25.6% baseline), but we do not read that as evidence of an edge. That number is a bootstrap of the OOS trades. It reflects low per-trade variance plus a slightly positive sample mean. We did not compute the same figure for the random-direction control, so it cannot show that the signal adds anything.

7. What would need to change

- **A materially larger OOS effect.** On this sample, the strategy would have needed to beat about +0.14R per trade to clear the random p95.
- **More trades or more markets,** so that a real but small effect can be separated from noise. That would mean a new pre-registration, not a re-run of this one.
- **No retuning on this OOS period.** Any revised version (a V2) must be locked and hashed first, then tested on data after 2026-10-24.

What actually happened next: we pre-registered a 10-coin weekly trend-following version (MULTICOIN_TREND_V1), with its rules hash posted publicly on X before the run. It also failed. Its OOS Sharpe was 0.31, below the random-baseline p95 of 0.86.

8. Limitations

- There is one instrument on one venue, and about 4 years of data. 88 OOS trades is a small sample.
- The margin fee is modelled as a flat 0.033% per day, not as the realised funding path. Slippage beyond gap fills is not modelled.
- The 95th-percentile estimate rests on 200 seeds, so the threshold itself carries some sampling noise.
- This is a backtest. It says nothing about live execution, capacity or future market regimes.

9. Disclosures

- **We were both the proposer and the tester.** This was Anchor's own research, not a third-party submission.
- **The hash was not published before this run.** It was recorded locally and enforced by the script. Outsiders therefore have to take the lock timestamp on trust for this particular test. Since the next test, we post the hash publicly before running.
- **There was a process slip in the same batch.** The rules said that a family failing the cost gate gets no OOS evaluation, but the script still computed OOS numbers for the two families that failed the gate (FUNDING_CONTRA_V1 and HOUR_SEASON_V1). Those numbers were labelled info-only, and the verdicts followed the gate as written. TSMOM passed the gate, so its OOS evaluation was allowed, and the slip does not affect this verdict.
- **Results across the whole batch:** all 3 families and all 7 configurations failed. Since nothing passed, there is no multiple-testing false positive to correct for.
- The rules file, results file and test script are available on request.

This report describes a historical simulation. It is not investment advice, it is not a recommendation to trade, and it makes no promise of returns.

Contact: reply on LinkedIn or on X at x.com/ProjectAnchorHQ.