

Kalshi 15-Minute Signals — Test Report

October 20, 2026 · Signal tests run on Jason's 60-day data (Aug 21 – Oct 20): BTC 5,696 windows, Gold 4,052, ETH 5,697 · Calibration tables refreshed the same day on the full since-launch history (Aug 13 – Oct 20): BTC 6,427 windows, Gold 4,616 · Direction calls are made at 6 minutes left; accuracy judged on chronological holdout halves so nothing is graded on data it was built from.

Bottom line

Eight different candidate signals were tested for **direction**. **None of them beat the simple lead** (spot vs target at 6:00). In every subgroup of every test, the side that was leading still won more than half the time — so no signal ever justifies flipping a call. Direction accuracy at 6:00 stands at BTC 79.4%, Gold 78.4%, ETH 82.0%, and that appears to be the ceiling of what the data supports. Re-checked on the full since-launch history it is unchanged (BTC 79.4%, Gold 78.4%) — the ceiling holds on every window that exists. Where real gains *were* found is **confidence calibration**: knowing which high-% calls are truly high and which are overstated. Three such fixes are now live (below), and one long-term data collection (order-book depth) is running.

Direction tests — all negative

Signal tested	Asset	Result (holdout half)
Price path / momentum (12→6 min shape)	BTC, Gold	No help — BTC 78.9% → 78.8%, Gold flat
Cross-asset: ETH move → BTC; BTC move → ETH	BTC, Gold	No help — accuracy identical to lead-only
Last-3-minute burst	BTC, Gold	No help — BTC worse; Gold gain reversed out-of-sample
Volatility-normalized lead	BTC, Gold	No help — 78.5–79.7% vs 78.9–79.9% baseline
Kalshi Yes drift / spread / volume pattern	BTC	No direction help — confidence sorting only
Multi-venue lead (Coinbase + Binance blend)	BTC	No help — venue gap median only \$5; in disagreements Coinbase was right 54%
Taker flow / CVD (aggressive buy vs sell \$)	BTC, Gold	No help — test-half 78–80% with flow vs 79–80% without; flow is mostly the lead
ETH as a BTC leader (12-min disagreements)	BTC	No — BTC followed itself 57.1%, ETH only 42.9%

One nuance: BTC taker flow did split mid-size leads (\$25–125) consistently — flow-agree held 84% vs flow-against 77%. But flow-against still wins ~77%, so it can never flip a call; it is the same confidence-downgrade family already live via the market check and the ETH downgrade.

What DID work — confidence calibration (now live)

1. Kalshi market-price check — BTC. Kalshi's Yes price adds no direction (in the 182 windows where it disagreed with our lead, it was right 50.5% — a coin flip). But it sorts confidence powerfully: filtered calls with the market 90¢+ on our side held 93.8–98.7%, vs 83.6–86.0% when softer. *Live rule*: \$75+ calls print 94/96/98% when strong, 86/84% + a "market soft" warning when soft; soft 3-min confirmations print 85% instead of 93/98%.

2. Kalshi market-price check — Gold (bigger effect). Gold's lead is measured in PAXG, but Kalshi settles on Pyth, so Gold's market carries real information: in 506 disagreement windows the market was right 63.2%. Confidence sorting is stronger than BTC's — calls ≥\$3 with a strong market held 97.6% vs 81.9% soft. The single biggest mislabel found anywhere: Gold 3-min confirmations printed 98% but soft-market ones held only 76.1%. *Live rule*: full strong/soft label tables on Gold 6-min calls; soft confirmations now print 76%.

3. ETH downgrade — BTC mid-leads. When BTC's lead is \$25–75 and ETH moved the other way during the window, those calls held only ~64% (holdout: 63.8–68.9% across both spot-ETH and Jason's Kalshi ETH sheet) while printing 75–82%. *Live rule*: those calls now print 64% with an "ETH against" note. Direction never changes.

ETH — added as a third asset

Jason's ETH sheet showed ETH's own 6-min lead is 82.0% accurate — the best of the three assets. ETH calls are now live with their own 60-day tables: ±\$7 or more → 98%, \$3–5 → 90–96%, \$5+ confirmations at 3:00 held 99.0%

(labeled 98%). One caution: ETH setups that just crossed flipped back 33.3% of the time, so ETH flip-risk warnings fire often.

Still collecting: order-book depth

Kalshi keeps no history of book depth (resting Yes vs No size), so it cannot be backtested. A silent logger now records the book at the 6:00 mark of every window for all three assets. After a few days of windows it can be graded honestly against settled results — it is the last untested direction candidate.

Update — full since-launch history (Oct 20)

Jason supplied the complete history of both markets since their Aug 13 launch (BTC 6,493 rows, Gold 4,745). The overlap with the 60-day files — 5,696 BTC and 4,052 Gold windows — matched **exactly**: zero result mismatches, zero checkpoint differences. Checkpoint agreement on the full history (12 / 10 / 8 / 6 / 4 / 2 min left): BTC 66.9 / 70.4 / 75.7 / 79.4 / 83.6 / 88.5%; Gold 65.2 / 69.5 / 73.9 / 78.4 / 82.1 / 86.4% — the same shape as the 60-day data at every horizon.

Label changes (live). Most buckets moved ≤ 1 pt. BTC small leads were under-labeled and are corrected: 6-min +\$0–10 now 59% (was 54%), +\$10–25 68% (67%), –\$25...–10 65% (63%), –\$50...–25 77% (76%), –\$150...–125 91% (92%); 3-min +\$0–10 58% (56%), +\$10–25 73% (71%), +\$25–50 78% (77%); the no-orderbook tier for \$125–200 leads is 93% (94%). Gold: 6-min +\$0–0.50 now 56% (57%), +\$1–2 73% (74%); the Gold 3-min table is unchanged.

Flip-risk corrections. The BTC crossed-setup warning quoted 18% — that figure covered only filtered big leads. On the full history, crossed setups flip back 38% for leads under \$75 and 20% for \$75+ leads; the warning now prints the rate for the lead size. BTC not-unanimous big leads flipped 10% (was quoted 8%). Gold crossed setups flipped back 34% (was quoted ~21%).

Confirmations re-graded. BTC confirm-grade windows held 96.8% overall; the \$75–100 tier held 92.7% (93% label stands) and \$100+ held 97.6% (98% label stands). Gold \$5+ confirmations held 99.4% (98% cap stands). The market-price strong/soft splits and the ETH downgrade come from candle backtests on the 60-day span and were not re-run; ETH has no since-launch file yet and keeps its 60-day tables.

Current live system

Asset	6-min call	3-min check	Extras
BTC	Always direction + %, full-history buckets; \$75+ confirmed 98% at \$5+ grade, FLIP line	Confirmed 98% at \$5+ grade, FLIP line	Flip-risk lines (33% flip-back)
Gold	Always direction + %, self-consistent PAXG buckets (full 98%), trading 75% on first \$5 bucket	Confirmed 98% at \$5+ grade, FLIP line	Flip-risk lines (33% flip-back)
ETH	Always direction + %, own 60-day buckets	Confirmed 98% at \$5+ grade, FLIP line	Flip-risk lines (33% flip-back)

All three assets also get a settled Result line after each window. Everything runs until Jason says stop. Confidence labels are capped at 98% — no test on the full history supports printing 99%.