

Kalshi MLB Game-Winner Trading Strategy

Research Report: Live Trading Results and Historical Backtests

Prepared 2026-07-20 | Data window: 2026-05-12 through 2026-07-19

Executive Summary

This report consolidates the statistical evidence gathered from an automated Kalshi MLB game-winner trading bot: live settlement history, joined trade-context instrumentation, and two in-sample backtests over 1-minute price candles and play-by-play data (907 games through 2026-07-11; 868 games through 2026-07-19). The bot buys the favored team's contract late in games when its price sits in a defined band and its run lead clears a minimum margin, holding to settlement.

- Kalshi's in-game MLB prices are close to efficient. A price band alone (0.90-0.97) is slightly negative after fees; essentially all edge in this strategy comes from gating entries on run-margin, not from price selection.
- The run-margin ≥ 3 gate is the one robust, monotone finding across every cut of the data. It replicated in a 907-game backtest and a refreshed 868-game backtest.
- The strategy's net edge is small and appears to be eroding: +0.79 cents/contract in-sample through 7/11 fell to -0.11 cents/contract on the same rule set through 7/19, as a profitable early-May stretch rolled out of the ~68-day data window and three losses on 7/18-19 rolled in.
- Live results confirm this: 126 joined trades through 7/19 show a 92.1% win rate but are net -\$6.17 - running almost exactly at the breakeven win rate implied by the average entry price.
- All-time realized P&L across the full 231-settlement history is -\$35.50, verified directly against the Kalshi settlements API and reconciled against the account balance (Section 2). 84% of that loss (-\$29.72 of -\$35.50) came from the pre-fix era (before 2026-06-30), driven by a duplicate-buy bug that let the bot bet on both sides of the same game at once - a real trading defect, not a data artifact. The post-fix era alone is close to breakeven at -\$5.79.
- First-touch entry (buying the instant a game qualifies) outperforms waiting for confirmation; requiring persistence monotonically destroys value. No stop-loss level tested beats holding to settlement. Dwell time near the entry band carries no exploitable signal.
- One open, unresolved conflict: the season backtest says the 0.96-0.97 sub-band loses money and 0.90-0.92 is where the profit lives; the smaller live-fills sample suggests the opposite. Neither cut is statistically decisive; the question is left to forward validation.

1. Data Sources and Methodology

1.1 Live trading records

- Kalshi settlements API (GET /portfolio/settlements): ground-truth realized P&L per market (revenue in cents, yes/no total cost in dollars, fee cost in dollars).
- trade_context.csv: bot-side instrumentation logged at entry (inning, half, outs, scores, run margin, contracts) for every live buy since 2026-07-06.
- Joined by market ticker AND entry timestamp (within a 3-minute window) - a ticker-only join was found to contaminate results with unrelated fills on the same game from before a position-sizing bug fix, and was corrected during this analysis.

1.2 Backtests

- Kalshi settled markets + 1-minute yes_bid/yes_ask candlesticks, pulled via the public market-data endpoints (approximately 68-day rolling retention).
- MLB StatsAPI play-by-play, joined by game and timestamp, to reconstruct the exact run differential at every point in the game (free, no key required).
- Entry simulation rule: enter on the favored side the first minute its ask price falls in [0.90, 0.97] while its run margin is at or above the required minimum; one entry per game; live-game minutes only; taker fee modeled as $\text{ceil}(7 \times p \times (1-p)) / 100$ per contract, matching the bot's real fee schedule.
- Two runs: an initial pull covering 907 games (2026-05-05 to 2026-07-11), and a refreshed pull covering 868 games (2026-05-12 to 2026-07-19) after the data window rolled forward.

2. Live Trading Results (through 2026-07-19)

Metric	Value
Account balance	\$36.75
All-time realized P&L (after fees)	-\$35.50 (196W / 35L, 84.8%)
All-time gross P&L (before fees)	-\$32.65
All-time fees paid	\$2.85
Post-fix era (settlements since 2026-06-30)	117W / 10L = 92.1%, net -\$5.79
Joined trade-context sample (126 trades, Sec. 2.1 methodology)	116W / 10L = 92.1%, net -\$6.17
Last 7 days (2026-07-13 to 07-19)	2W / 3L (40.0%), net -\$10.43 - driven by the 07-19 collapse

All 231 settlement records were re-pulled directly from the Kalshi settlements API on 2026-07-20 and independently verified: zero duplicate (ticker, settled_time) pairs, a continuous date range of 2026-06-15 through 2026-07-19, and the resulting all-time net P&L (-\$35.50) reconciles exactly against the account's current balance and implied net deposits (balance \$36.75 = net deposits \$72.25 + realized P&L -\$35.50). This is a corrected, fuller figure than an earlier version

of this report, which under-collected settlements via an incomplete pagination pass and understated the loss as -\$31.69.

2.0 Pre-fix vs. post-fix breakdown

Era	Settlements	Record	Net P&L
Pre-fix (2026-06-15 to 06-29)	104	76W / 28L (73.1%)	-\$29.72
Post-fix (2026-06-30 onward)	127	117W / 10L (92.1%)	-\$5.79
All-time	231	193W / 38L (83.5%)	-\$35.50

The pre-fix era accounts for 84% of the all-time loss despite being less than half the trade count. The mechanism is a real trading defect, not a data error: before the 2026-06-30 fix, an in-memory duplicate-buy guard could fail and let the bot buy BOTH the YES and NO side of the same game. Kalshi nets offsetting YES/NO holdings on the same market at settlement, so the account only ever collects a net payout capped at the size of the larger leg - but it paid full price for every contract on both legs. Example verified against the raw settlement and fill records: on 2026-06-28 the bot bought 2 YES contracts (cost \$1.06) and 3 NO contracts (cost \$2.50) on the same NYY-vs-BOS game - \$3.56 paid in total for a position that could only ever pay out \$3.00 at best, and NO won for a net 1-contract payout of \$1.00, realizing a \$2.62 loss on a single game purely from betting against itself. This pattern, combined with the partial-exit stop-loss bug and 33 pre-fix active stop-loss sells noted in Section 7, is what drove the pre-fix era to -\$29.72. Since the fix, the duplicate-buy guard checks live Kalshi position state before every order and this failure mode has not recurred.

2.0.1 Post-fix era in isolation

Since the 2026-06-30 fix, the bot has traded close to breakeven (net -\$5.79 over 127 settlements). The 92.1% win rate looks strong in isolation, but at the average entry price traded (approximately 0.92), breakeven requires almost exactly that win rate once the taker fee is included - the post-fix era is therefore consistent with a strategy running at zero edge, not a validated winning edge. The most recent week (2026-07-13 to 07-19) is a clear negative outlier, net -\$10.43 on just 5 settlements, entirely from three margin-3+ losses on 07-18 and 07-19 (Section 2.2) - too small a sample to say the edge has turned negative, but a reminder that variance at 4-contract sizing is large relative to the per-trade edge.

2.1 Price-band split (live fills)

Entry price band	Record	Net P&L
0.90 - 0.91	54W / 8L	-\$8.64
0.92+	62W / 2L	+\$2.47

Caution: 8 of 10 total losses clustered in the 0.90-0.91 band, but this is not statistically significant (8 losses vs. approximately 6 expected under the null, p is approximately 0.25) and it directly contradicts the larger season backtest (Section 3.3), which finds 0.90-0.92 to be the profitable sub-band and 0.96-0.97 to be the losing one. Inning of entry carried no additional signal once margin and price were controlled for.

2.2 The run-margin gate: origin and durability

MIN_RUN_DIFF was introduced after joined ledger/context data through 2026-07-06 showed 1-run leads winning only about 80% against a roughly 92% breakeven requirement at typical entry prices - every live settlement loss up to that point had entered on a margin of 1 or 2 runs. The rule was set to 2, then raised to 3 in .env once the 907-game season backtest confirmed the gate; the "margin 3+ never loses" pattern (44W / 0L) held from early July until it was broken by three straight diff-3/diff-4 losses on 2026-07-18 and 07-19 - a reminder that a bigger margin reduces the loss rate but does not eliminate it.

3. Backtest 1 - Through 2026-07-11 (907 games, n=601 qualifying entries)

3.1 Price band alone has no edge

Simulating entries on price alone (band 0.90-0.97, no margin filter) nets -0.90 cents/contract after fees. Kalshi's in-game MLB prices are well-calibrated on their own; the price band by itself is not a source of edge.

3.2 Run-margin sweep (first-touch entry, hold to settlement)

Minimum run margin	Net / contract	Sample size
>= 0	-0.90c	-
>= 1	-0.70c	-
>= 2	-0.07c	-
>= 3	+0.79c	n = 601
>= 4	+0.48c	-

The sweep is monotone from margin 0 through margin 3, and margin 3 is the peak - this is the strongest single finding across the whole project and is what the live MIN_RUN_DIFF=3 setting is built on.

3.3 Price sub-band split at margin >= 3

- 0.90-0.92 entries: +1.82 cents/contract (profitable)
- 0.96-0.97 entries: -2.2 cents/contract (loses money in every cut tested, including at margin >= 2 where it lost -7.9 cents/contract)
- Combined rule (margin >= 3 AND price 0.90-0.95): +1.29 cents/contract over 515 games - the basis for the BUY_CEILING=0.95 setting, though single-parameter cuts of this size are individually within about 1 standard deviation of noise; the monotone trend across the full sweep is the stronger evidence.

3.4 Cross-venue check

A Kalshi-vs-Polymarket arbitrage check found no exploitable edge: pregame price deltas were 0-1 cent, and in-game deltas were stale-quote artifacts on thin books (~10 contracts), smaller than the round-trip fee. Kalshi carries 2-8x Polymarket's MLB dollar volume, confirming it as the more liquid, more efficient venue for this market.

4. Backtest 2 - Refreshed Through 2026-07-19 (868 games, n=577 qualifying entries)

The same rule set (margin ≥ 3 , band 0.90-0.97, first-touch entry) was re-run after the ~68-day data retention window rolled forward, dropping a profitable early-May stretch and adding the losing 07-18/07-19 weekend. This refresh also tested three new questions that the forward price_path.csv instrumentation was built to eventually answer out-of-sample: entry persistence, dwell time, and post-entry drawdown.

4.1 Headline: the edge compressed toward zero

Backtest window	Sample	Win rate	Net / contract
Through 2026-07-11 (original)	n = 601	-	+0.79c
Through 2026-07-19 (refreshed)	n = 577	93.24%	-0.11c

Nothing about the rule changed between these two runs - only the rolling data window moved forward by eight days. With a per-trade standard error on the order of 1 cent, both +0.79c and -0.11c are statistically indistinguishable from zero. Read together with the live joined-sample result in Section 2 (92.1% win rate, running at breakeven), the consistent picture is a strategy operating very close to zero edge after fees, not a durable, exploitable inefficiency.

4.2 Entry persistence - first-touch vs. waiting for confirmation

Consecutive qualifying minutes required (k)	Net / contract
k = 1 (first touch)	-0.11c
k = 2	-0.49c
k = 3	-0.72c
k = 5	-1.12c

Waiting for the price/margin condition to hold for multiple consecutive minutes monotonically destroys value - it only raises the average entry price paid, without filtering out any losing trades. This replicates the same monotone pattern found in the original backtest. First-touch entry remains the correct rule; a hypothesis that brief price spikes were unreliable "fakeouts" worth waiting out is not supported by the data.

4.3 Dwell time before cheap (0.90-0.921) entries

Time spent near the entry floor in the prior 15 minutes	Net / contract
Passed through (0-3 min)	+0.55c
Lingered (4-9 min)	+1.33c
Camped (10+ min)	+0.77c

No exploitable pattern: entries that lingered near the band floor before qualifying performed about the same as ones that passed straight through. Dwell time is not a useful additional filter.

4.4 Post-entry drawdown and stop-loss evaluation

Win rate conditioned on how far the position's price dipped after entry, before settlement:

Minimum post-entry price reached	Sample size	Win rate
Never dipped below 0.85	n = 331	100.00%
Dipped to 0.70 - 0.85	n = 137	99.27%
Dipped to 0.50 - 0.70	n = 47	97.87%
Dipped below 0.50	n = 62	40.32%

This is the sharpest signal in the entire dataset: a price crash below 0.50 after entry is a genuine danger sign (win rate falls to a coin flip), while any shallower dip - even into the 0.50-0.70 range - is still won almost 98% of the time. Despite this clear split, actively cutting losses at any tested stop-loss level still underperforms simply holding to settlement, because closing a position early always pays a second taker fee and permanently forfeits the positions that would have recovered:

Rule	Net / contract vs. holding
Hold to settlement (baseline)	-0.11c
Stop-loss at 0.85	-2.28c
Stop-loss at 0.80	-1.48c
Stop-loss at 0.70	-1.22c
Stop-loss at 0.60	-0.72c
Stop-loss at 0.50	-0.20c

Every stop-loss threshold tested loses to holding, and the closer the threshold is to the entry price, the worse it performs. This confirms the bot's current STOP_LOSS_ENABLED=false setting is correct given the data so far.

5. Unresolved Questions and Risks

5.1 Price sub-band conflict

The season backtest (n=601+, Section 3.3) finds 0.96-0.97 entries losing money and 0.90-0.92 entries carrying the profit. The live-fills cut (n=126, Section 2.1) finds the reverse - most live losses clustered in 0.90-0.91. Neither result is statistically decisive on its own (the live cut is not significant at p is approximately 0.25; the backtest's sub-band split is a single-parameter cut within about 1 standard deviation of noise). This is the single most important open question for the strategy and should not be resolved by intuition - it needs more out-of-sample data.

5.2 Margin ≥ 3 is not a shield

The "44 wins, 0 losses" pattern for margin-3+ entries held for roughly six weeks before breaking with three consecutive losses on 07-18 and 07-19. This does not invalidate the gate (it remains the strongest and most monotone finding in the data) but confirms it is a favorable tilt, not a guarantee, and the bot should be expected to occasionally lose even on high-margin entries.

5.3 Sample size and window instability

The entire backtest is drawn from a single rolling ~68-day window of publicly available Kalshi candlestick history, and the headline net-edge figure moved by nearly a full cent per contract (+0.79c to -0.11c) simply from that window rolling forward eight days. Conclusions about the overall profitability of the price/margin rule should be treated as provisional until a longer, non-overlapping history is available.

6. Forward Validation Plan

A price-path logger (price_path.csv, added 2026-07-19) now records, on every discovery cycle, the timestamp, matched market, run margin, leading side, and fill price for every live game the bot observes - including games that were skipped by the margin gate and games where a position is already held. This creates a live, out-of-sample counterpart to the backtest questions in Section 4: first-touch vs. persistence, dwell time, and post-entry drawdown, measured against real order-book fills rather than 1-minute candle closes. This dataset is the intended arbiter for the unresolved price sub-band conflict in Section 5.1, and should be reviewed again once a meaningful number of forward trades has accumulated.

7. Current Strategy Configuration (as of 2026-07-19)

Parameter	Value	Basis
BUY_THRESHOLD	0.90	Season backtest price-band sweep
BUY_CEILING	0.95	Season backtest sub-band split (Sec. 3.3); unresolved per Sec. 5.1
MIN_RUN_DIFF	3	Monotone margin sweep, Sec. 3.2, confirmed on refresh
CONTRACTS_PER_TRADE	4	Scaled with account balance over July 2026
STOP_LOSS_ENABLED	false	Every tested stop-loss level underperforms holding (Sec. 4.4)
Entry timing	First-touch (no persistence wait)	Persistence sweep, Sec. 3 and 4.2

Conclusion: every configuration decision the bot currently makes is supported by at least one backtest cut, and the two independent data refreshes agree on the qualitative shape of the strategy (margin gate helps, price band alone does not, waiting hurts, stop-losses hurt). The open question is no longer "is the current configuration reasonable" but "how much edge, if any, remains after fees" - the honest reading of both the live and backtested evidence is that the

strategy is operating close to breakeven, and the forward price-path data should be the deciding evidence going forward.