Opening-Range Reversal in Dow Jones Futures: Evidence Consistent with 0DTE Hedging Pressure

Lance Ladia[1]

July 2026

Abstract

This paper documents a short-side opening reversal pattern in E-mini Dow futures over 194 trades from January 2023–March 2026. Under a symmetric 1:1 payoff structure, the strategy produced a 59.79% win rate and gross expectancy of 0.196R per trade. Performance was concentrated in the earliest entry windows: trades in the first three timing buckets won 65.56% of the time, compared with 39.53% of the time in the final two. A reconstructed pre-2023 historical sample (N = 115, provisionally dated to approximately 2021 - 2022) shows a win rate of 45.22%, below breakeven and significantly lower than the 2023-2026 sample (Fisher's exact test, $p = 0.0135$). This discontinuity is broadly consistent with the period over which 0DTE options volume scaled substantially, though the exploratory identification of this boundary limits the strength of that inference. The results are consistent with a time-bound hedging-related opening flow but do not independently identify 0DTE dealer activity as the causal mechanism. Exact trading parameters are withheld, limiting independent replication.

[1] The author trades the strategy described in this paper using personal capital and has a direct financial interest in its continued performance. No external funding was received for this research.

1. Introduction

Options market makers who inventory same-day expiry (0DTE) contracts face a distinct hedging obligation relative to longer-dated options. 0DTE positions carry large gamma exposure relative to their notional size, and options dealers must adjust delta hedges more frequently and more aggressively as the underlying moves.

A growing body of academic literature has documented that this hedging activity is an active driver of underlying price dynamics.

Baltussen, Da, Lammers and Martens (2021) show that hedging demand from options market makers and leveraged ETF providers produces measurable intraday momentum, concentrated in the final thirty minutes of the trading session across a large number of global futures markets.

Dim, Eraker and Vilkov (2024) extend this inquiry specifically to 0DTE index options, documenting that aggregate inventory gamma from market makers is related to future intraday volatility and to the direction of intraday price action.

Existing literature is concentrated on the market close, where gamma exposure is largest immediately before expiration. Little work discusses whether an analogous distortion exists at the open of the trading session, where hedging-driven distortion clashes against the day's initial price discovery.

This paper proposes and tests a specific, falsifiable hypothesis: that dealer hedging pressure tied to 0DTE positioning contributes to a short-lived upside overextension during the opening minutes of the New York equity-index session. The hypothesis is tested using a rules-based strategy on Dow Jones (YM) futures, applied to a five-minute chart between 9:35AM and 9:59AM Eastern Time, over a live and historically tested sample spanning January 2023–March 2026.

This paper makes three contributions: first, it extends the gamma hedging literature to an understudied window at the open. Second, it reports not only aggregate statistical tests, but several sub-sample robustness checks because a single aggregate p-value is an insufficient standard of evidence. Third, it states explicit and falsifiable predictions for how the mechanism should manifest across sub-samples, allowing the hypothesis to be evaluated on more than headline statistical significance alone.

1.1 Hypothesis

H1. Qualifying upside extensions during the opening session contain a transient component related to 0DTE dealer hedging flows, rather than exclusively durable directional information. Accordingly, a short-side strategy positioned against those extensions should perform best when the signal occurs closest to the open, with performance weakening as the entry time moves later. Failure to observe this timing relationship would weaken the proposed time-bound hedging interpretation.

H2. If same-day call-related hedging contributes to transient upside overextension, gap-up sessions should exhibit a stronger short-side mean-reversion edge than gap-down sessions. Because aggregate dealer positioning and the sign of dealer gamma are not observed, this is a conditional prediction rather than a directional measurement.

1.2 Signal Generation and Execution Logic

The strategy is a mechanical, short-side system applied during a fixed window from 9:35 - 9:59AM Eastern Time. Trade eligibility depends on the direction and structure of early price action. All positions are closed during the regular trading session and no overnight exposure is permitted. Exact trading rules are withheld as proprietary information and the resulting limitation on replication is discussed in Section 2.2.

1.3 Configuration Constants

Parameter	**Value**
Sample period	January 2023–March 2026
Instrument	YM (Dow Jones E-mini futures)
Timeframe	5 minute
Session window	9:35AM - 9:59 AM ET
Total trades (N)	194
Payoff ratio	1.0

2. Data and Methodology

2.1 Instrument and Sample Period

The strategy is tested on E-mini Dow Jones (YM) futures using five-minute bars covering the opening minutes of the U.S. equity index session (9:35-9:59AM Eastern). The sample spans January 2023–March 2026 (N = 194). All statistics reported are computed directly from the underlying trade log and were independently recomputed as a verification step prior to publication.

Five-minute YM data was obtained from top-of-book CBOT market data. The futures series was constructed using a continuous contract. The regular trading-hours open was defined as 9:30AM Eastern Time and the prior regular-session close was defined as 4:00PM Eastern Time. Holiday-shortened sessions and contract-roll sessions were included according to rules specified before evaluation.

2.2 Signal Structure

The strategy is a short-side and rules-based system applied during a fixed window from 9:35-9:59AM Eastern Time. Trade eligibility is conditioned on the direction of early price action.

The strategy is strictly intraday. All positions are closed before the end of the regular trading session and no positions are carried overnight. The reported results contain no overnight risk.

The precise entry and exit parameters are withheld as proprietary trading information. Consequently, the strategy cannot be independently replicated. The disclosed sample, validation procedures, statistical methodology, and aggregate results allow the reader to evaluate the statistical evidence supporting the stated hypothesis, within the limitations associated with an undisclosed rule specification.

The withheld parameters were fully specified before data collection began. No parameter was selected, modified, or optimized in response to observed results.

Five entry windows are defined at five-minute intervals beginning at 9:35, 9:40, 9:45, 9:50, and 9:55AM Eastern Time. The entry-window number recorded for each trade refers to the interval in which the qualifying entry signal occurred. A maximum of one trade was permitted per trading session.

2.3 Exit and Risk Management

Trade outcomes are recorded on a realized-point basis with a payoff structure consistent with symmetric win/loss framing (wins = 116, losses = 78). Exact exit and stop-loss parameters are withheld for the same reason as stated in Section 2.2. The reported win rate and R-multiple based results reflect the actual realized outcome of every trade in the sample, not a hypothetical or idealized exit.

Each trade was normalized to an R-multiple outcome of +1R for a winning trade and -1R for a losing trade. The sample contains 116 wins and 78 losses, producing a gross result of +38R. Although the underlying stop and target distances vary in index points, the target-to-risk ratio is fixed at 1:1.

2.4 Overnight Gap Classification

Each trading day is classified as a gap-up or gap-down day based on the relationship between the prior regular trading hours session's close vs current regular trading hours session's open. This classification is used to test H2: that the direction of the overnight gap conditions the magnitude of the opening hedging distortion, on the theoretical grounds that a gap-up moves index call options further into the money, requiring more aggressive incremental dealer hedging and producing a more pronounced mean-reversion edge.

2.5 Statistical Approach

The primary aggregate test is a one-sided exact binomial test of the null hypothesis that the strategy's win probability equals 0.50 against the alternative that it exceeds 0.50.

H1 predicts declining performance as entry time moves later. This prediction is evaluated using a standard binary logistic regression of binary trade outcome on entry-window number. A pooled comparison of Entry Windows 1-3 against Windows 4-5 is also reported.

H2 is evaluated through a direct two-sided Fisher exact test comparing proportions on gap-up and gap-down sessions. Testing each subgroup separately against 50% does not, by itself, establish that the subgroups differ from one another.

The binomial analysis assumes that trade outcomes are sufficiently independent. Potential serial dependence was evaluated using a Wald-Wolfowitz runs test.

3. Results

3.1 Headline Performance Metrics

Metric	Value
Win rate (full sample)	59.79%
Entry Windows 1-3 win rate	65.56%
Entry Windows 4-5 win rate	39.53%
Gap-up win rate	64.29%
Gap-down win rate	54.74%

The strategy recorded 116 wins in 194 trades, for a win rate of 59.79%. The one-sided exact binomial p-value is 0.0039. The corresponding normal-approximation statistic is Z = 2.73. The 95% Wilson confidence interval for the win rate is 52.77% to 66.44%.

3.2 Profitability and Expectancy

Because the strategy uses a symmetric 1:1 payoff structure, per-trade expectancy in R-multiples reduces to the difference between win rate and loss rate:

$$E[R] = p(+1R) + (1-p)(-1R) = 2p - 1$$

With $p = 0.5979$, expected gross return is $2(0.5979) - 1 = 0.1958R$ per trade, consistent with the aggregate R-multiple return of 38R across 194 trades.

Segment	Win Rate	Expectancy
Full Sample	59.79%	+0.196R
Entry Window 1 - 3	65.56%	+0.311R
Entry Window 4 - 5	39.53%	-0.209R
Gap-up	64.29%	+0.286R
Gap-down	54.74%	+0.095R

Separating expectancy across the sub-samples shows the same threshold pattern already identified in the entry-timing analysis: trades entered within the first three timing buckets carry a substantially higher expectancy (+0.311R) than the full-sample average, while trades entered in the final two buckets carry a negative expectancy (-0.209R). Similarly, gap-up sessions carry roughly three times the expectancy of gap-down sessions, consistent with the directional asymmetry hypothesized in H2, though the difference between the two groups has not yet reached statistical significance on its own.

Expectancy under a fixed 1:1 payoff ratio is a direct measurement of win rate. This section does not introduce new statistical information, but is included because expectancy is the more standard metric by which trading strategies are evaluated.

3.3 Serial Dependence

The binomial significance test reported in Section 3.1 assumes that individual trade outcomes are independent and identically distributed. This assumption would be violated if wins and losses demonstrated a tendency to cluster (positive serial dependence) or alternate (negative serial dependence) more than chance would predict. Either pattern would bias the standard errors underlying the reported p-value and would need to be corrected before that result could be interpreted at face value.

This is tested directly using a Wald-Wolfowitz runs test on the chronologically ordered sequence of win/loss outcomes. Of 194 trades, the sequence contains 102 runs (a run defined as a maximal consecutive block of same-outcome trades) against an expected value of 94.28 runs under the null hypothesis of independence ($Z = 1.156$, two-sided $p = 0.248$). The null hypothesis of serial independence is not rejected.

The trade sequence shows no statistically detectable clustering of wins or losses, and the standard binomial variance estimate is therefore an appropriate basis for the reported significance level. The observed run count is modestly above the expected value (102 versus 94.28), consistent with a slight tendency toward alternation, though this deviation is well within the range attributable to chance and is not a substantive finding.

4. Robustness, Sub-Sample Analysis, and Regime Discontinuity

A single aggregate significance test is a necessary but incomplete basis for evaluating a strategy. This section compartmentalizes the headline result across several dimensions - entry timing, overnight gap direction, and calendar year - to assess whether the aggregate edge is structurally consistent with the stated hypothesis or concentrated in a narrow, potentially non-representative slice of the sample.

4.1 Entry Timing Analysis

A standard binary logistic regression of trade outcome on entry-window number produced a coefficient of approximately -0.376, representing an odds ratio of approximately 0.686 for each one-window delay in entry time. The associated p-value is approximately 0.0019. This indicates that the estimated odds of a winning trade decline as the entry time moves later in the session.

Entry Window	N	Win Rate	Expectancy	P-value
Entry Window 1 (earliest)	54	74.07%	+0.481R	0.0003
Entry Window 2	51	60.78%	+0.216R	0.080
Entry Window 3	46	60.87%	+0.217R	0.092
Entry Window 4	27	37.04%	-0.259R	n/a
Entry Window 5 (latest)	16	43.75%	-0.125R	n/a

H1 predicts that the mean-reversion edge should be strongest close to the open and weaken as the session progresses. The data supports this directionally, but not as a linear decay but as a threshold-like decline. Entry Windows 1-3 produced 99 wins in 151 trades, or 65.56%, compared with 17 wins in 43 trades, or

39.53%, for Windows 4-5. A two-sided Fisher exact test gives p = 0.0027. The estimated odds of a winning trade were approximately 2.91 times higher in the early-window group.

Windows 2 and 3 are statistically indistinguishable from one another (60.78% versus 60.87%, respectively) while window 1 is numerically stronger than both.

Windows 4 and 5 fall below breakeven. Restricting the sample to only the first three entry windows alone (N = 151) yields a win rate of 65.56% and a materially stronger result (Z = 3.82, p = 0.00008) than the full sample, indicating that the full sample statistic in Section 3.1 is diluted by the inclusion of the later windows.

The negative expectancy in windows 4 and 5 is directionally consistent with H1, but neither window's deviation from 50% reaches significance in isolation given the small sub-sample size.

4.2 Gap-Direction Asymmetry

Gap Direction	N	Win Rate	Expectancy	P-value
Gap-Up	98	64.29%	+0.286R	0.003
Gap-Down	95	54.74%	+0.095R	0.206

Consistent with H2, gap-up sessions outperformed on win rate and expectancy over gap-down sessions, and the gap-up result is independently significant while the gap-down result is not.

Gap-up sessions produced 63 wins in 98 trades, or 64.29%, compared with 52 wins in 95 trades, or 54.74%, on gap-down sessions. A direct two-sided Fisher exact test gives p = 0.190, with an estimated odds ratio of approximately 1.49 and a 95% confidence interval of approximately 0.84 to 2.65. The point estimate is directionally consistent with H2, but the interval includes no difference and the direct comparison is not statistically significant.

H2 is therefore treated as directionally supported but not statistically confirmed at the current sample size.

One trade occurred in a session where the closing price of the prior session was equal to the opening price of the current session and was excluded from the directional gap analysis.

4.3 Year-by-Year Stability

Year	N	Win Rate	p-value
2023	59	54.2%	0.301
2024	55	61.8%	0.052
2025	64	59.4%	0.084
2026 (partial)	16	75.0%	0.038

None of the yearly samples is large enough to provide a precise estimate of the strategy's annual win probability, particularly the partial 2026 sample of 16 trades. The estimates are descriptively stronger after

2023, but the year-by-year pattern does not independently establish a temporal trend or a relationship with 0DTE growth. Testing that explanation requires a direct measure of options activity rather than calendar year alone.

4.4 Pre-2023 Regime Discontinuity

This pre-2023 sample cannot be precisely dated, but based on row count and typical trade frequency it is provisionally estimated to span roughly 2021-2022.

Metric	Pre-2023	Post-2023
N	115	194
Win Rate	45.22%	59.79%
Expectancy	-0.0956R	+0.196R

A two-sided Fisher exact test comparing the periods gives $p = 0.0135$, with an estimated odds ratio of 1.80 and a 95% confidence interval of approximately 1.13 to 2.87.

The samples therefore exhibit a statistically detectable difference in strategy performance. This establishes a difference between the periods under the assumptions of the test, but it does not identify the cause of that difference.

The January 2023 division was identified after the historical results were examined and is therefore treated as exploratory. The result may be consistent with a change in options-market structure, but it may also reflect changes in volatility, monetary-policy conditions, opening-auction behavior, market liquidity, or other unmeasured regimes. Prospective evidence or a direct options-positioning variable is required before the change can be specifically attributed to 0DTE activity.

5. Alternative Explanations

The results in Sections 3 and 4 are consistent with the dealer-hedging hypothesis stated in Section 1.1, but consistency does not guarantee exclusivity. This section discusses competing explanations for the observed effect and states what can be reasonably distinguished from the hedging hypothesis using the available data and what cannot.

5.1 Trivial Mean Reversion

The most basic alternative explanation is that any strategy positioned against a short-term price extension will show some reversion as a consequence of construction, independent of any hedging-related mechanism. This explanation cannot fully account for the pattern in Section 4.1. A purely mechanical reversion effect of this kind would not be expected to produce the specific threshold structure observed where the edge is concentrated in the earliest few minutes of the session and inverts to a loss in the final two entry windows.

Generic opening-session mean reversion remains a viable competing explanation. The concentration of performance in the earliest entry windows narrows this explanation but does not eliminate it, because opening-auction imbalances, overnight inventory correction, and liquidity normalization may also decay rapidly after the open. Distinguishing these mechanisms requires variables that directly measure their predicted intensity.

5.2 General Intraday Momentum and Reversal

A related possibility is that the effect is reflective of a broader intraday reversal pattern, for instance, reversal driven by overnight order imbalances, opening-auction effects, or algorithmic trading flows unwinding into the open. This is a difficult alternative to rule out using the data presented in this paper alone. The entry window pattern in Section 4.1 as well as the gap-direction results in Section 4.2 offer partial evidence against a purely generic explanation.

A mechanism unrelated to options hedging would not obviously predict an asymmetry between gap-up or gap-down sessions, but because the difference itself (Section 4.2) did not reach significance, this evidence is suggestive rather than decisive.

6. Falsifiability and Forward Testing

A hypothesis that cannot be falsified by a specific outcome is not being genuinely tested. This section states explicit, numeric criteria for what future evidence would weaken each component of the hypothesis.

6.1 Sequential Testing for Edge Persistence

The strategy is monitored in live trading using a Sequential Probability Ratio Test (SPRT) framework, evaluated after each trade against pre-specified historical baselines.

The SPRT allows edge degradation to be detected as early as the data supports using the statistic:

$$S_n = log\left(\frac{L(p_1)}{L(p_0)}\right)$$

where:
$p_0 = 0.5979$: edge intact
$p_1 = 0.50$: edge degraded

And as live trades are accumulated, the statistic is updated using:

$$S_n = W_n log\left(\frac{p_1}{p_0}\right) + L_n log\left(\frac{1-p_1}{1-p_0}\right)$$

where:

W_n and L_n are cumulative wins and losses, respectively.

The following configuration for the SPRT is used:

Metric	Value	Interpretation
Null Hypothesis (p_0)	0.5979	True win rate equals the backtested baseline in Section 3.1
Alternative Hypothesis (p_1)	0.50	True win rate has degraded to chance level
Type I Error Rate (α)	0.05	Probability of rejecting p_0 when the edge is intact (false positive)
Type II Error Rate (β)	0.20	Probability of failing to detect true degradation when it has actually occurred
Upper Boundary (A)	2.773	Crossing this bound favors p_1 (edge degraded)
Lower Boundary (B)	-1.558	Crossing this bound favors p_0 (edge intact)

If $S_n \geq 2.773$, the test terminates in favor of the degraded-edge hypothesis.

If $S_n \leq -1.558$, the test terminates in favor of the intact-edge hypothesis.

If $-1.558 < S_n < 2.773$, the test continues.

The simpler, independently verifiable benchmark: if win rate over the next 50 live trades falls at or below 50%, this central hypothesis (H1, Section 1.1) is considered weakened and will be reported as such, independent of the formal SPRT outcome.

6.2 Threshold-Structure Persistence

Section 4.1 reports a threshold decay in edge across entry-timing windows, the pooled sample from Windows 1-3 is significantly profitable and windows 4-5 unprofitable in the current sample. This structure is a falsifiable prediction distinct from aggregate results tested in Section 6.1. If over the next 50 live trades, the win rate in entry-timing windows 1-3 is not measurably higher than in windows 4-5, the threshold interpretation is weakened, independent of the aggregate result.

6.3 Proposed Discriminating Test for the 0DTE-Specific Mechanism

Section 5 identifies several alternative explanations that the current evidence cannot fully rule out. A test capable of discriminating between the 0DTE-hedging hypothesis specifically and these generic alternatives would condition strategy performance on a direct measure of same-day 0DTE options volume or open interest, rather than relying on calendar time as a proxy as Sections 4.3 and 4.4 do already. The hedging hypothesis predicts that edge strength should track 0DTE volume directly, including on a day-by-day basis within a single year, not just across years. A generic mean-reversion or volatility based explanation predicts no such relationship. This test is not performed in the current paper because the backtest did not include a matched 0DTE volume series and is stated here as a specific and pre-registered direction for future work.

7. Limitations

This paper has several important limitations. First, the exact trading rules are withheld, preventing independent replication and verification of implementation details. Second, the study does not directly observe dealer positions, inventory gamma, or hedge transactions and therefore cannot establish that 0DTE activity caused the documented pattern. Third, the analysis is restricted to a single futures instrument and a short-side signal, limiting external validity. Finally, results are reported gross of estimated execution costs, and five-minute data may introduce ambiguity regarding intrabar execution order.

8. Conclusion

This paper documents a positive short-side opening-reversal strategy in the E-mini Dow Jones futures between January 2023 and March 2026. The full sample of 194 trades produced a 59.79% win rate under a symmetric 1:1 payoff structure, corresponding to gross expectancy of +0.196R per trade. The strongest evidence is not the aggregate win rate alone, but the concentration of performance near the open: winning probability declined as entry time moved later, and the pooled first three entry windows materially outperformed the final two.

Gap-up sessions also produced stronger estimated performance than gap-down sessions, but the direct difference between those groups was not statistically confirmed. Likewise, the weaker pre-2023 result represents an exploratory regime comparison and does not identify the cause of the apparent change.

The findings are therefore consistent with a short-lived opening flow related to same-day options hedging, but they do not establish a causal 0DTE mechanism. Direct dealer-positioning delta, replication across related equity-index futures, realistic net-cost analysis, and continued prospective testing are required to distinguish that mechanism from generic opening-session mean reversion and other regime-dependent explanations.

The criteria in Sections 6.1 and 6.2 will be evaluated against live trading results and reported in a future revision, regardless of whether the outcome supports or weakens the stated hypotheses.

Disclosure Statement

This paper is provided for information and research purposes only and does not constitute investment advice, a recommendation, or an offer or solicitation to buy or sell any security or financial instrument. Past and live trading performance discussed in this paper is not indicative of future results. Trading futures involves substantial risk of loss and is not suitable for all investors.

References

Baltussen, G., Da, Z., Lammers, S., & Martens, M. (2021). Hedging Demand and Market Intraday Momentum. Journal of Financial Economics, 142(1), 377–403.

Dim, C., Eraker, B., & Vilkov, G. (2024). 0DTEs: Trading, Gamma Risk and Volatility Propagation. SSRN Working Paper.