# Microprice Drift and the Theta Wall: Why Order-Book Depth Cannot Time Option Execution

Charlie Yan*

August 2026

**Abstract**

Top-of-book depth imbalance is among the strongest known short-horizon predictors in equity microstructure. We test whether it works in options—and document, with three pre-registered results, why a signal can be genuinely informative and still useless. First, depth imbalance in SPY options is *real information*: cross-sectionally within each snapshot, imbalance predicts the next hour's delta-hedged mid-price move with a controlled rank IC of 0.104 ($t = 66.9$ over 9,958 snapshots), surviving residualization on moneyness, price level, relative spread, volume, and open interest, and a within-snapshot shuffle null. Posted depth is long established as a *liquidity* measure in options, alongside spread, volume and open interest; what we have not found in the literature is its *imbalance* tested as a predictor of option-specific returns. To our knowledge this is the first such measurement—and it is distinct from executed order flow, whose information content in options is established. Second, it cannot be traded directly: the toll structure of option markets is flat across horizons (median round trip $2.36 per contract at one hour and $2.37 at nine days) while the option-specific move grows faster—as $t^{0.72}$, because a decay term linear in $t$ superposes on a diffusive term in $\sqrt{t}$—so the move-to-toll ratio rises from 1.78× (hourly) through 6.8× (daily) to 34× (nine days): friction mechanically rewards patience, and the achievable value of an hourly signal with IC $\approx 0.10$ sits below any deployable threshold. Third—the sharpest result—depth cannot even time executions a trader was going to do anyway, because of what we call the *theta wall*: for a premium seller, delaying a sale by one hour forfeits approximately $1.2 per contract of option decay on the studied universe, while the depth signal's conditional fill improvement is $0.24–$0.46. Waiting for a favorable book loses money *even conditional on the signal being right* (−$1.19 per contract, $t = -4.31$); the apparent gain from delaying purchases is the same theta with the sign flipped, available without any signal. We reconcile this with Muravyev and Pearson's finding that execution timing sharply reduces effective option spreads: their effect operates at second-to-minute horizons where decay is negligible; ours at the hourly horizon where decay dominates. Together the two results bracket the usable timescale of option execution timing, and we state the boundary as a general bound: a liquidity taker benefits from signal-timed delay only where the signal-conditional improvement of the tradeable side outruns the instrument's deterministic decay, $g(h) > \theta_{\$}\, h$—a condition no improvement in signal accuracy at a given horizon can rescue.

*Draft v0.3 (journal-length), August 2026. All results SCREEN-basis: information coefficients and comparisons on quoted prices; no fill simulation, no queue modeling, no performance claims. Every test was pre-registered with fixed horizons, controls, and decision bars before any statistic was computed; the registration documents, including declared deviations, are summarized in Appendix A. Data: SPY option chain, hourly, 2020-09-01 to 2026-05-12, 1,429 trading days; top-of-book depth (bid and ask size) newly extracted from raw vendor files (20.8 million contract-snapshots, 100.00% join to the quote panel). Code, pre-registration documents and a data-availability statement: https://github.com/charlieyanhx/option-depth-theta-wall.

# 1 Introduction

In equity limit-order books, the imbalance between posted bid and ask depth is one of the most robust short-horizon predictors known: it forecasts the direction of the next price move, anchors the microprice, and underlies a generation of execution and market-making models (Gould et al., 2013; Cont et al., 2014; Stoikov, 2018). Options would seem a natural export market for the idea. Option quotes are wide, so the value of predicting the next quote move is high; option books are professionally made, so posted size plausibly encodes dealer inventory and hedging pressure; and the options literature has already established that *executed* order flow carries information (Muravyev, 2016), including delta-weighted order imbalance in index options (Sensoy and Omole, 2022).

This paper tests the export on 5.7 years of hourly SPY option chains, using a field that is routinely discarded: the posted bid and ask *size* attached to every quote, which we extract from raw vendor files (20.8 million contract-snapshots; the field survives in the raw data but had never entered our derived panels, and, we suspect, rarely enters anyone's). The answer comes in three pre-registered parts, and its interest lies in their combination:

1. **The signal is real** (§4). Within-snapshot, depth imbalance predicts the next hour's delta-hedged option mid move with controlled rank IC 0.104 ($t = 66.9$), surviving controls for moneyness, price level, relative spread, volume, and open interest—the last two mattering because the unsigned volume/open-interest family carries nothing here, so posted depth is incremental information, not repackaged flow.
2. **It cannot be a strategy** (§5). Option friction has a horizon structure that equity intuition does not prepare one for: the round-trip toll is *flat* in the holding period while the option-specific move grows as $t^{0.72}$. An hourly signal therefore fights friction $\sim 19\times$ harder per unit of movement than a nine-day signal, and an IC of 0.10 at the hourly clock is worth less, deployably, than nothing.
3. **It cannot even be an execution timer** (§6). The natural fallback —use the signal to time fills for trades one was doing anyway, where it bears no toll—dies on what we call the theta wall: an option seller who waits an hour forfeits $\approx$ \$1.2 per contract of decay to collect a $\approx$ \$0.3–0.5 depth tilt. Waiting loses money *conditional on the signal being right.*

The third result appears, at first reading, to contradict a well-known finding: Muravyev and Pearson (2021) show that traders who time option executions pay effective spreads less than 40% of conventionally measured ones. Section 7 reconciles the two: their timing operates at second-to-minute horizons, where the decay cost of waiting is negligible—linearly extrapolating our measured hourly decay gives $\sim$ \$0.02 per minute, and we emphasise this is an extrapolation, since our data cannot resolve sub-hourly decay directly—and predictable quote adjustment dominates; ours operates at the hourly horizon we do measure. The reconciliation is itself the paper's most portable statement: execution-timing signals in decaying instruments are governed by the inequality $g(h) > \theta_{\$}h$, and the two studies sit on opposite sides of its boundary.

# 2 Related literature

**Depth and imbalance in equities.** Gould et al. (2013) survey the limit-order-book literature; Cont et al. (2014) establish the price impact of order-book events; Stoikov (2018) formalizes the microprice as the depth-weighted expectation of the next mid. Cao et al. (2009) document that the open limit order book contributes to price discovery beyond the best quotes. Chordia et al. (2002) connect aggregate order imbalance to returns and liquidity at the daily horizon. Our contribution relative to this line is the transplant: the same object, measured in an option market, with the option-specific complications (a two-dimensional instrument grid, price levels spanning two orders of magnitude, and decay) handled explicitly. **Information in option markets.** Sensoy and Omole (2022) find that delta-weighted *trade* imbalance in index options predicts spot returns through a hedging channel; Muravyev (2016) shows executed option order flow predicts option returns and attributes the effect to inventory risk; the index-options order-imbalance literature (Sensoy and Omole, 2022) documents related effects in KOSPI options. Posted depth is a different object from executed flow—the standing schedule of intent rather than the record of transactions—and while it is a standard input to option liquidity measurement, we have not found its imbalance evaluated as a return predictor. Muravyev and Pearson (2021) is the closest relative on the execution side and is discussed throughout. **Costs and the premium.** Figlewski (1989) anticipated that option-market imperfections dominate fine-grained edges. The decay that defines our wall is the same object that generates delta-hedged option return premia (Bakshi and Kapadia, 2003); a premium seller's theta is not an inconvenience but the business, which is exactly why delaying its collection is expensive.

## 3 Data and the extracted field

Vendor chain files (hourly snapshots, seven per trading day, 2020-09-01 to 2026-05-12) carry top-of-book `bid_size` and `ask_size` for every contract. These fields are routinely dropped in research extractions; extracting them from all 1,430 raw daily files yields 20,823,814 contract-snapshots that join the quote panel at 100.00% on exact contract identity. In the studied belt (out-of-the-money, $|\Delta| \in [0.05, 0.40]$, 25–45 days to expiry) the touch is deep: median 195 contracts bid, 182 offered—roughly two orders of magnitude above a retail-scale clip, so depth is not a capacity constraint here; the question is whether its *imbalance*, imb = (bid size − ask size)/(bid size + ask size), is informative. It varies widely (5th–95th percentile −0.81 to +0.82; only 6% of quotes exactly balanced).

Two structural facts shape all designs below. First, **depth and relative spread relate backwards** from equity intuition: the deepest quintile of the book carries the *widest* relative spreads (1.37% versus 0.83% for the thinnest)—a price-level artifact, since cheap far-from-the-money contracts carry large posted size. Any uncontrolled depth signal is therefore confounded with moneyness and price level by construction; the controls of §4 are mandatory. Second, depth has *fallen* across the sample (median total size 1,042 in 2020 to ∼390 in 2026), so nothing below is fitted to early-sample conditions.

## 4 Result 1: the information is real

**Design (pre-registered; family of 8—four horizons × {raw, controlled}, the shuffle nulls being diagnostics rather than tests—giving Bonferroni $|t| > 2.73$).** Within each snapshot, Spearman rank IC between imbalance and the forward delta-hedged mid move $[\Delta\text{mid} - \delta\,\Delta S] \times 100$ over $h \in \{1, 2, 7, 63\}$ snapshots, averaged across snapshots with Newey–West $t$ (Newey and West, 1987) at 10 lags—longer than the longest evaluation horizon within a trading day (seven snapshots), so the estimator spans any within-day dependence the overlapping targets induce. The delta-hedged move is the right target: the unhedged move is dominated by the underlying, which is the (closed) directional question, not the option-specific one. Controls: within-snapshot OLS residualization of imbalance on $\{|\delta|, \log(\text{mid}), \text{spread}\%, \log(1+\text{volume}), \log(1+\text{OI})\}$; a within-snapshot shuffle null must and does return zero (IC −0.0002 to +0.004, n.s.).

**Lag robustness.** Ten lags spans within-day dependence but not the 62-snapshot overlap of the nine-day target, so we re-estimated every $t$ across lag choices up to 126. The conclusions do not turn on it: at one hour $t$ falls from 66.9 (10 lags) to 32.5 (126 lags), and at nine days from 5.5 to 4.1—both still clear of the 2.73 bar. The overlap induces less dependence in the $IC$ series than in the return series it is built from, because the cross-sectional ranking turns over even while the forward windows overlap: the nine-day IC series' own autocorrelation is +0.087 at one lag and below +0.03 beyond ten.

| Horizon | Raw IC | **Controlled IC** | NW $t$ (controlled) | Snapshots |
|---|---|---|---|---|
| 1 hour | 0.110 | **0.104** | 66.9 | 9,958 |
| 2 hours | 0.083 | 0.076 | 57.1 | 9,957 |
| 1 day | 0.047 | 0.038 | 29.5 | 9,952 |
| 9 days | 0.017 | 0.011 | 5.5 | 5,251 |

Table 1: Within-snapshot rank IC of depth imbalance on the forward delta-hedged mid move. All eight declared tests significant at the Bonferroni bar; the shuffle null is zero.

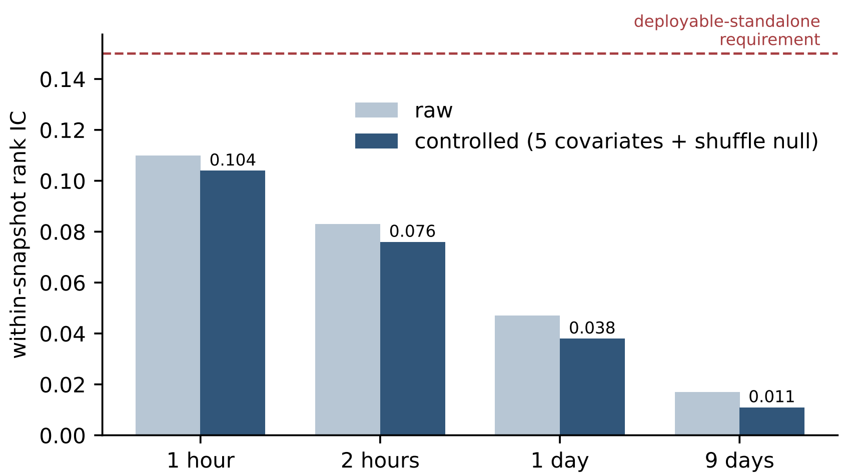


Figure 1: Raw and controlled IC by horizon. The signal is real, strongest at the shortest horizon, decays monotonically—and never approaches the level a standalone deployment requires.

The controlled IC of 0.104 at one hour survives everything we pre-registered against it, and its survival of the volume/open-interest controls is the substantive point: posted depth is not a repackaging of activity; it is standing information about the direction of the next quote adjustment—the options analogue of the microprice (Stoikov, 2018).

## 5 Result 2: the horizon gradient

Before testing the signal we bounded the driver (a Gate-0 oracle in our internal terminology): perfect foresight of the sign of the delta-hedged move, charged the full quoted spread at entry and exit plus hedge costs. Table 2 and Figure 2 report the inputs that matter.

| Horizon | Median \|hedged move\| | Median round trip | Move/toll | Oracle ceiling (Sharpe) |
|---|---|---|---|---|
| 1 hour | \$4.21 | \$2.36 | 1.78× | 10.1 |
| 2 hours | \$6.64 | \$2.36 | 2.81× | 12.6 |
| 1 day | \$16.18 | \$2.38 | 6.80× | 12.9 |
| 9 days | \$81.28 | \$2.37 | 34.3× | — |

Table 2: The toll is flat in horizon; the move grows as $t^{0.72}$ (fitted across the four rows), faster than diffusion alone. Ceilings are perfect-foresight bounds, not attainable performance.

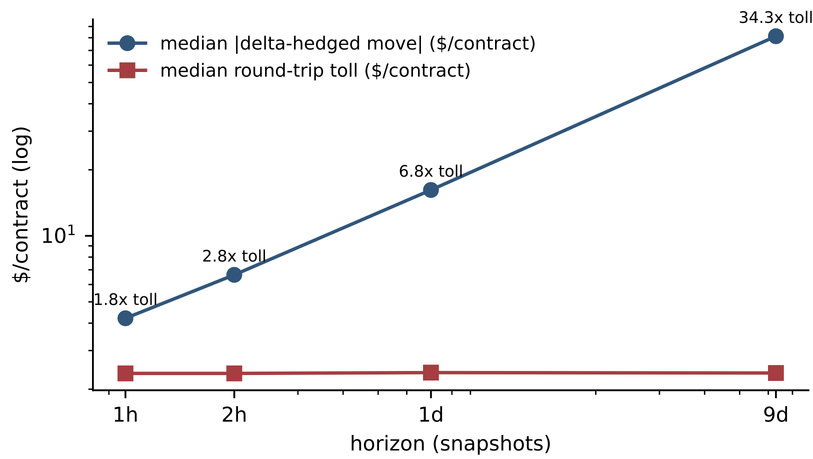


Figure 2: The horizon gradient. Friction taxes hourly trading $\sim 19\times$ harder than nine-day trading per unit of option-specific movement.

The flat toll is a structural fact of option markets worth stating on its own: the bid-ask spread is paid per transaction, not per unit time, while the option-specific move accumulates with horizon. It accumulates *faster* than diffusion: the fitted exponent across the four rows is $t^{0.72}$, not $t^{0.5}$, because the delta-hedged move superposes a deterministic decay linear in $t$ on a diffusive component in $\sqrt{t}$. One consequence deserves emphasis because it cuts against a naive reading of the last column. At the long horizons the decay term *is* most of the move. Measuring it directly on the same belt—the median *signed* hedged move, whose drift is the decay—gives −\$0.93 at one hour, −\$2.01 at two, −\$7.68 at one day and −\$65.48 at nine days, so decay accounts for 22% of the median absolute move at one hour and 81% at nine. (We measure rather than extrapolate deliberately: scaling the hourly figure flat across sixty-three steps would put the nine-day decay at \$59–\$76 depending on the hourly estimate used, and theta is not constant over nine days of a one-to-two-month option.) The nine-day 34× therefore does not describe thirty-four tolls' worth of *tradeable dispersion*; it largely describes decay, which accrues to whoever is short regardless of any signal. What the gradient legitimately establishes is the comparative claim—the same fixed toll is amortised over a much larger price change at longer horizons—and that claim is if anything strengthened by the decay term, since decay is the one component of the move that is not a forecasting problem. Slow expressions of any option signal therefore deserve the first look, not the last. For depth imbalance specifically, the implied achievable value at the hourly clock — IC × ceiling $\approx 0.104 \times 10.1 \approx 1.05$ before estimation error and implementation — sits below a deployable threshold, and the required standalone IC ($\approx 0.15$) exceeds anything the positioning literature sustains at daily horizons. The natural fallback is execution: time the fills of trades one was doing anyway, where the signal bears no toll. Muravyev and Pearson (2021) give that fallback a strong precedent. The next section shows why it fails at this horizon anyway.

## 6 Result 3: the theta wall

**Design (pre-registered; four tests—quintile monotonicity, the sell rule, the buy mirror, and the unhedged variant—giving Bonferroni $|t| > 2.50$; $t$-statistics are Newey–West at 10 lags on the daily mean series, which removes the overlap induced by multiple entries per day).** For a representative premium-seller's population (out-of-the-money puts in the commonly traded moneyness belt, one-to-two-month tenor, snapshots early enough to permit a same-day wait; $n = 93{,}165$), compare filling *now* at the bid against waiting one snapshot and filling at the next bid, delta-adjusted for the intervening underlying move; the fixed decision rule "wait iff imb $> 0.33$" against always-now; the mirror test for buys on the ask; and the unhedged variant. The tradeable side, not the mid, is the object: order-book imbalance predicting mid drift is partly the mechanical microprice adjustment, which a crossing trader cannot capture.

| | Mean effect ($/contract) | NW $t$ | Verdict |
|---|---|---|---|
| Wait-to-sell rule, hedged, conditional | **−1.19** | −4.31 | significant, *wrong way* |
| Q5−Q1 imbalance spread (sells) | +0.24 | 1.31 | n.s. (whole belt: +0.46, $t$ 5.2) |
| Wait-to-buy rule, hedged, conditional | +1.57 | 6.77 | significant — see text |
| Wait-to-sell, unhedged, conditional | −6.85 | −6.09 | significant, worse |

Table 3: The execution test on the tradeable side. Waiting to sell loses even conditional on favorable imbalance.

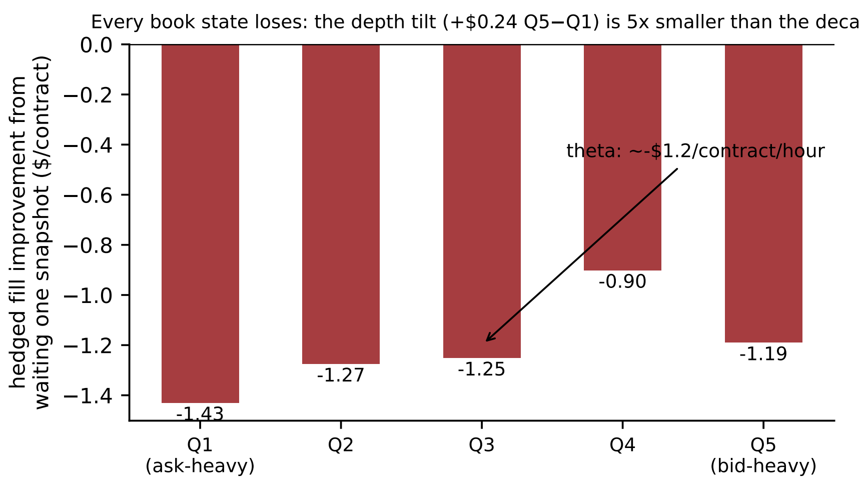


Figure 3: Hedged fill improvement from waiting one snapshot, by imbalance quintile, sells. Every quintile is negative: the level is theta, and the depth tilt cannot climb it.

Every imbalance quintile's wait-to-sell improvement is negative (−$0.90 to −$1.43; Figure 3). The mechanism, once named, is obvious: the hedged expected drift of an option price over one hour is its *theta*—≈ $1.2 per contract per hour on this universe, the right order for the moneyness belt and tenor studied—and a premium seller who delays forfeits precisely the decay the position exists to collect. The depth tilt is real (+\$0.24 conditional, +\$0.46 in the wider belt) but five times smaller than the decay it must overcome. The buy-side "gain" (+\$1.57) is the same theta with the sign flipped: available to anyone who delays a purchase, no signal required; the signal-specific increment is again $\approx$ +\$0.3–0.5.

**The general bound.** Let $\theta_\$$ be the instrument's deterministic decay per unit time and $g(h)$ the signal-conditional expected improvement of the tradeable side over delay $h$. A liquidity-taking seller benefits from signal-timed delay iff

$$g(h) \; > \; \theta_\$ \, h.$$

Since $g(h)$ is bounded by the signal's conditional drift forecast (here $\lesssim$ \$0.5 at the strongest conditioning) while the right side grows linearly, the inequality fails at the hourly horizon by a factor of five in the world's most liquid option market—and no improvement in the *accuracy* of the depth signal changes that. The wall is set by theta, not by noise.

## 7 Reconciliation with Muravyev–Pearson

Muravyev and Pearson (2021) show that option price changes are predictable at high frequency and that traders who time marketable executions pay effective spreads under 40% of conventional measures. Our negative result and theirs are both correct, on opposite sides of the boundary the theta wall defines. Their timing operates at second-to-minute horizons, where the decay cost of waiting is small: linearly extrapolating our measured hourly decay gives $\theta_\$ h \approx \$0.02$ per minute—an extrapolation, not a measurement, since hourly snapshots cannot resolve sub-hourly decay—which is negligible against the predictable component of the next quote adjustment, so $g(h) > \theta_\$ h$ can hold. We establish our side of the boundary and the functional form connecting the two; we do not independently verify theirs. At the one-hour delay our snapshot data can measure, decay has grown fifty-fold while the conditional improvement has not, and the inequality fails by a factor of five even conditioning on a favorable book. The practical statement for a premium seller: time executions *within* the minute if infrastructure allows; never *across* hours. The two studies together bracket the usable timescale of option execution timing, and the bound gives the bracket its formula.

## 8 A complementary null: the hour cannot sign trades either

A natural escalation from posted depth is executed direction: sign the interval's trade prints against prevailing quotes (Lee and Ready, 1991) and build a poor-man's signed order flow—96% of traded contract-hours are signable this way, since options print at the touch. Pre-registered tests of print-signed flow at the cell granularity used for positioning research—persistence of signed demand, lead onto adjacent cells, and a skew-axis aggregate—fail at roughly one-tenth of their bars (cell ICs +0.006 to +0.029 against bars of 0.09–0.175), and one diagnostic explains why: the open print and the close print of the *same hour* agree on direction only 59% of the time. At hourly aggregation, initiator information does not survive; signability is not informativeness. The two nulls are complementary: posted depth is informative but unmonetizable; hourly-signed flow is not even informative.

## 9 What survives

Three durable facts. (1) Option order-book imbalance carries genuine, control-surviving, short-horizon information—which makes this a sharper cautionary result than a null: informativeness and monetizability are different properties, separated here by three explicit, measured mechanisms (microprice mechanics; the flat toll; the theta wall). (2) The horizon gradient is a structural fact of option markets with reach beyond this study: strategy evaluation should price friction per unit of horizon, and slow expressions of fast signals deserve the first look. (3) The extracted depth panel retains non-signal value (capacity measurement, monitoring), and the bound $g(h) > \theta_{\$}h$ gives execution research in decaying instruments a stopping rule: measure decay against conditional drift before building anything.

## 10 On the friction basis

We charge the full quoted spread throughout, while citing Muravyev and Pearson (2021) — who show that timed executions pay well under 40% of it — as a central reference. The tension is deliberate and worth making explicit. Writing $\phi$ for the effective-to-quoted ratio, the move-to-toll ratios of Table 2 all scale by $1/\phi$: at $\phi = 0.35$ the hourly figure rises from 1.78× to 5.1× and the nine-day figure from 34× to 98×. Three consequences follow. First, the *gradient* — the ratio of ratios, 19× — is invariant to $\phi$, so the structural claim that friction taxes fast strategies harder survives any execution assumption. Second, the theta wall of §6 does not involve the spread at all: it compares option decay against the signal-conditional drift of the bid, so it too is invariant. Third, the one quantity that does move is the achievable-value estimate of §5 — a lower toll raises the oracle ceiling and the implied achievable figure with it. Readers who believe their execution reaches $\phi \approx 0.35$ should treat that estimate as correspondingly more favourable, and the standalone case as correspondingly less closed, while the two results this paper actually rests on remain untouched.

## 11 Limitations

One underlying; hourly snapshots, so the minute-scale regime where Muravyev and Pearson (2021) operate is unobservable to us and our reconciliation of the two regimes is an extrapolation of the decay line, not a measurement of their side; top-of-book only (hidden and away-from-touch liquidity unobserved); SCREEN basis throughout—no queue modeling, no fill simulation. The theta-wall bound applies to liquidity takers; a maker with queue priority faces different economics that our data cannot address. The premium-seller universe for Result 3 is representative rather than exhaustive; the belt-wide monotonicity (+\$0.46, $t\,5.2$) suggests the tilt generalizes, and the wall's arithmetic (decay versus tilt) generalizes a fortiori.

## A Pre-registration summary and deviations

Three registered documents govern this paper: (A1) the IC design—horizons, five controls, shuffle null, eight-test family and Bonferroni bar, plus committed outcomes for each result-pattern (standalone / execution-input / artifact / nothing); (A2) the execution design—universe, now-versus-wait definition on the tradeable side, the fixed 0.33 threshold, four-test family, and a \$0.50-per-contract economic yardstick; (A3) the print-signing protocol, run under a parent pre-registration for signed-flow research with its bars inherited unchanged. Declared deviations: the Gate-0 bound (Table 2) was computed *before* the depth extraction, on the pre-existing quote panel, deliberately—bounding the driver before spending the extraction; and the execution pre-registration failed to name theta as the expected level of the wait cost, an omission we report rather than repair, since the discovery that the level *is* theta is the paper's central finding.

## References


Bakshi, G., Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies* 16(2).

Cao, C., Hansch, O., Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets* 29(1).

Chordia, T., Roll, R., Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics* 65(1).

Cont, R., Kukanov, A., Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics* 12(1).

Figlewski, S. (1989). Options arbitrage in imperfect markets. *Journal of Finance* 44(5).

Gould, M., Porter, M., Williams, S., McDonald, M., Fenn, D., Howison, S. (2013). Limit order books. *Quantitative Finance* 13(11).

Lee, C., Ready, M. (1991). Inferring trade direction from intraday data. *Journal of Finance* 46(2).

Muravyev, D. (2016). Order flow and expected option returns. *Journal of Finance* 71(2).

Muravyev, D., Pearson, N. (2021). Options trading costs are lower than you think. *Review of Financial Studies* 33(11).

Newey, W., West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3).

Sensoy, A., Omole, J. (2022). Information content of order imbalance in the index options market. *International Review of Economics & Finance* 78, 418–432.

Stoikov, S. (2018). The micro-price: a high-frequency estimator of future prices. *Quantitative Finance* 18(12).