Overnight Returns in 184 F&O-Eligible Indian Equities: Descriptive Evidence, an Invalid Long Strategy, and Its Correction

Subhash Sadacharam, Independent Researcher

Working paper — draft of 12 September 2026.

Abstract

"Buy Today, Sell Tomorrow" (BTST) — buy near the close, hold overnight, sell at the open — is common folklore among Indian retail and quant traders. This paper takes that folklore seriously and tests it directly. Stated at the level of precision it can support: in a selected, partially point-in-time F&O-eligible Indian equity panel, close-to-open returns are positive and broadly distributed; straightforward long-only attempts to monetize the pattern do not appear robust after realistic implementation constraints. Using six years of one-minute data (2020–2025) on 184 NSE stocks that were F&O-eligible when this panel was built, we find the close-to-open ("overnight") return earns +17.97 basis points per day pooled. Under two-way (stock × date) cluster-robust inference, the conservative standard this paper relies on throughout, this gives a t-statistic of 9.48; a naive per-observation Newey-West HAC estimate, which does not account for cross-sectional correlation across stocks on the same day, gives a considerably larger t-statistic of 81.04, but we do not treat that larger figure as the paper's primary evidence. The premium is positive for all 184 sampled stocks, with the per-stock mean overnight return ranging from 2.50 bps/day (HDFCBANK) to 36.56 bps/day (HINDCOPPER) and a median of 18.23 bps/day — stable in every calendar year from 2020 to 2025, and robust to value-weighting, winsorization, liquidity screening, and a point-in-time correction for stocks that joined the F&O segment mid-sample (16.95 bps corrected; two-way clustered t=10.05, alongside a supplementary HAC t=80.66). A calendar-time block bootstrap confirms it (95% CI [12.95, 22.16] bps). Two precise, deliberately un-blurred nuances qualify this: roughly a third to a half of the raw premium is market-beta exposure (unit-beta-adjusted 6.93 bps/day; rolling-beta-adjusted 9.13 bps/day, using a corrected beta estimator described below), while the cross-sectional ranking skill that identifies which stocks will outperform survives factor neutralization far better — about 87% of a persistence ranking's edge over random selection (9.24 to 8.09 bps/day) remains after removing beta, liquidity, and volatility jointly. This decomposition, not the execution analysis that follows, is this paper's primary contribution. All risk adjustment in this paper is against a single NIFTY-beta factor; it is not a claim of a risk-adjusted excess return net of a fuller set of standard equity risk factors (size, value, momentum, quality), which this paper does not construct.

BTST folklore is silent on execution, however, and this paper asks simply whether any harvesting method delivers a return that is both economically material and statistically robust net of realistic costs and correct real-time implementability, rather than applying a single bright-line pass/fail threshold. By that standard, no long-only strategy tested here is robust after realistic implementation constraints. Buying every eligible stock every night is catastrophic under realistic position sizing and fees (−63.2% CAGR out-of-sample). A model-conditioned improvement — forced nightly liquidation of a top-decile selection, chosen by a simple 60-day trailing-mean persistence rule that a seven-feature LightGBM model and a linear ridge regression both fail to beat on identical cross-validation folds — does much better but still lands at best near breakeven (−0.5% CAGR retail, −9.5% entity, at ₹20,00,000 starting capital), and 0 of 30 cost-sensitivity parameter combinations we tested are both net-positive and statistically significant. A third

discipline (“Hold-Through”) initially appeared to solve the problem convincingly, but this rested on a look-ahead bug in the backtest, caught only through a simulated adversarial referee review and corrected (summarized in Section 5.5; full account in Appendix A). Once corrected, Hold-Through returns −18.1% CAGR (retail) and −22.7% (entity) on the original discovery window and −25.1% on the 2026 window — worse than forced liquidation in every case, negative across every capital level from ₹5,00,000 to ₹1,00,00,000, every cost scenario tested, and both windows.

One tested component needs careful, separate characterization rather than a verdict borrowed from its point estimates: a bottom-decile short overlay — a strategy built independently and walk-forward-legitimately from the start, entirely separate from the flawed Hold-Through construction — shows a positive point estimate in isolation (CAGR +2.4% discovery window, +12.2% 2026; Sharpe 0.24 and 0.99 respectively), but a formal significance test computed for this revision (an HAC-adjusted t-test and a block-bootstrap confidence interval on its actual daily net returns) finds it is not statistically distinguishable from zero at conventional levels in either window (HAC t=0.40, p=0.69 discovery; t=0.83, p=0.41, 2026; both 95% bootstrap confidence intervals span zero). This is weak evidence in the discovery sample and is not independently established as implementable. A further, unfavorable qualification, established for this revision and stated plainly rather than left implicit: unlike this paper's core model, whose specification was frozen (commit 746d590, 10 September 2026) before any 2026 data existed anywhere in the project, the short-sleeve scripts that produced the 2026 figures above were themselves authored on 11–12 September 2026, after the 2026 raw data files were already present in the working directory (the raw pulls completed by early afternoon on 11 September, the merged file shortly after). The short sleeve's 2026 number is therefore not an out-of-sample or forward test of any kind, genuine or otherwise; it is an in-sample result relative to when the code that produced it was written, on the same data used to specify it, and we describe it that way throughout rather than implying it carries the validating weight of the ex-ante spread discussed below. Combined 50/50 with the now-corrected long sleeve it improves on the long-only result and cuts drawdown, but the combined portfolio remains net-negative in both windows (−8.1% discovery, −7.9% 2026). We also note, as a standing qualification rather than a one-time disclosure, that many specifications were explored over the course of this research program, and no formal multiple-testing correction has been applied to any individual result reported in this paper. The paper's conclusion, at the level of precision it can support, in this panel: no long-only strategy tested is robust after realistic implementation constraints; a modest short-side effect is present in point-estimate terms but is not established as statistically distinguishable from zero and is not independently established as tradeable. The underlying overnight premium documented in this panel is, separately, real, large, and pervasive by every dependence-robust standard we could apply here. A genuinely ex-ante predictive test — a same-day top-minus-bottom-decile return spread, ranked purely on information available before the day's open, with no conditioning on any subsequent day's information — replicates on genuine, previously-unobserved 2026 data (11.13 bps/day, HAC t=4.17, against 14.90 bps/day, t=9.78, in the original discovery window): a later-sample replication under a frozen specification, and the only claim in this paper that legitimately qualifies as forward validation. A separate, purely descriptive regression coefficient (the “continuing-selection” coefficient, +36.17 bps discovery vs. +35.59 bps 2026) shows a similar pattern in the later sample, but because it conditions on an ex-post classification of whether a stock is also selected the following day, this replication should not be read as validation of a live, implementable predictive rule. Every harvesting method we tested — blanket capture, forced liquidation, and the corrected Hold-Through discipline, under any cost assumption, capital level, or parameter choice — fails to deliver a return that is both economically material and statistically robust. We report the Hold-Through

false-positive and its correction (summarized in Section 5.5, given in full in Appendix A) because how it was caught — through deliberate adversarial review rather than internal confidence, however extensive — is itself a substantive methodological contribution, and because this project's own internal five-track audit (Appendix C), while thorough, is not a substitute for independent external validation, which would require public code/data release, a data dictionary, file checksums, a pinned environment, automated leakage tests, and third-party replication, none of which this draft yet provides.

Keywords: overnight returns, close-to-open return pattern, BTST, bid-ask spread estimation, transaction costs, market microstructure, Indian equities, F&O eligibility, look-ahead bias, backtest validation

JEL Classification: G12, G14, G15

1. Motivation

“Buy Today, Sell Tomorrow” — BTST — is a piece of trading folklore repeated constantly in Indian retail and quant circles: buy a stock near the close, carry it overnight without ever taking delivery into your demat account under the ordinary settlement cycle, and sell it the next morning at the open, on the belief that stocks tend to gap up overnight more often than they should. It is folklore in the precise sense that matters for a research paper — widely believed, rarely tested with the same rigor a documented academic finding would demand, and just as often invoked to justify a strategy that does not survive contact with real costs as one that genuinely does. This paper takes BTST at its word and asks, in order: is the underlying belief about overnight returns even true in this market; if it is, does it survive the scrutiny a skeptical referee would apply; and, having tried every increasingly sophisticated way we could construct to actually collect it, does any of them survive contact with realistic 2026 Indian trading costs and, just as importantly, with correct real-time implementability.

A stock's daily return decomposes into two pieces: the gap from yesterday's close to today's open, and the move from today's open to today's close. Cooper, Cliff and Gulen (2008) and Lou, Polk and Skouras (2019) show that in U.S. markets these two pieces behave very differently — one carries almost the entire long-run equity premium, the other often does not — and that the split persists across individual stocks, broad indices, and futures alike. Berkman, Koch, Tuttle and Zhang (2012) document a closely related U.S. pattern, a systematic overnight rise followed by an intraday reversal, attributed to retail-order-flow-driven attention effects concentrated at the open. A separate strand asks whether the split predicts future returns rather than simply persisting: Heston, Korajczyk and Sadka (2010) find intraday return continuation at intervals that are exact multiples of a trading day, and Akbas, Boehmer, Jiang and Koch (2022) show that a more pronounced overnight-rise/daytime-reversal pattern in a given stock predicts higher future cross-sectional returns for that stock, attributed to a tug of war between overnight noise traders and daytime arbitrageurs. Haghani, Ragulin and Dewey's “Night Moves” study is the flagship modern reference documenting this pattern at the individual-stock level in developed markets; the closest cross-sectional precedent to this paper's own decomposition approach is Bogousslavsky (2021), which decomposes intraday and overnight legs across a range of well-known cross-sectional anomalies without pursuing an execution-cost analysis of any of them. This paper extends that line of work to a cross-sectional emerging-market context with a real execution and cost analysis that, as far as we are aware, none of these papers attempt for individual equities: Kamath (2026, SSRN) studies an economically related question in India but at the index-futures level (NIFTY), not cross-sectionally across individual stocks.

This paper's primary contribution is Section 4: establishing that, in this panel, the close-to-open premium exists, is pervasive across stocks and years, and survives the most demanding dependence-robust inference

standards we could apply, alongside a precise account of how much of it is a market-beta effect and how much is genuine cross-sectional ranking skill. A secondary, cautionary component, Section 5, then asks whether any of several increasingly sophisticated ways of trying to harvest that premium can deliver an economically significant, real-time-implementable, cost-adjusted return. One of the harvesting disciplines tested there appeared, for a time, to have solved that problem, before a deliberate, adversarial re-examination of its information timing found the apparent success rested on a look-ahead bug. That discovery and correction is documented in full in one place, Appendix A (summarized in Section 5.5); it is a secondary methodological finding, not this paper's headline result, and we refer to it only briefly elsewhere.

The contribution is evidence within a liquid Indian equity universe that was F&O-eligible when this panel was built, not a claim about Indian equities generally, and because the universe is not fully point-in-time and corporate-action screening is a threshold-based approximation rather than a full adjustment pipeline (Section 3), the claim that the premium is broad and persistent should be read as applying to this specific, imperfectly-reconstructed panel, not as a fully general claim about the Indian equity market during this period. We chose F&O eligibility because it identifies a liquid, currently-investable population suitable for the implementation analysis in Section 5, and we do not demonstrate that restricting to it is innocuous with respect to the phenomenon under study. Within that scope, the paper makes four claims. First, in this panel, the close-to-open premium is real, large, and pervasive by any dependence-robust standard we could apply. A genuinely ex-ante predictive test — a same-day top-minus-bottom-decile return spread using only information available before the day's open, with no conditioning on any later day's information — replicates on genuine 2026 data (Section 5.3) — a later-sample replication under a frozen specification; this is the only sense in which anything in this paper is forward validated. A separate descriptive regression coefficient used throughout Section 5 also replicates in the 2026 sample, but because it conditions on an ex-post classification of next-day selection, it does not itself constitute a predictive test or forward validation (Section 4.5 states this distinction explicitly). Second, every harvesting method we tested — naive, moderately sophisticated, and initially the most sophisticated of all — fails to deliver a return that is both economically material and statistically robust once implemented correctly and costed realistically. Third, one of those methods appeared, for a substantial part of this project's life, to be the exception, and its apparent success rested on a specific, non-obvious, and initially highly convincing look-ahead bug in an invalid long strategy; that discovery and correction is consolidated in Appendix A (summarized in Section 5.5), presented there as a cautionary, audit-style account rather than a main line of argument. Fourth, a genuinely walk-forward-legitimate short overlay — built independently of, and never subject to, the flawed long-strategy construction — shows a positive point estimate in isolation but is not statistically distinguishable from zero under a formal test of its own daily returns (Section 5.7), and is not large enough, even taken at face value, to rescue the overall portfolio. Zaremba and Nikorowski (2018) document that most documented emerging-market anomalies die once real frictions are applied; this paper's central strategy result does not even reach the point where transaction costs are the deciding factor — it fails a more basic test of real-time implementability first, and only afterward, and independently, also fails on costs.

2. Institutional Background

Three features of Indian market structure matter directly for how this paper is built. We describe them here, before the data section, because each one shapes a specific methodological choice made later.

2.1 Session structure

NSE's continuous trading session runs 09:15–15:30 IST, preceded by a pre-open call auction (roughly 09:08–09:12) and followed by a closing-auction window (15:40–15:50). One-minute historical data from the Kite Connect API carries only the continuous session — the auction prices are not available at this granularity — so this paper's overnight and intraday returns are built from the continuous session's first and last one-minute bars, not the auction prices. Read "overnight" throughout as shorthand for a close-to-open, non-trading-period return that spans the exchange close, the pre-open call auction, and the opening transition of continuous trading combined, not a clean isolation of any single one of the three. This matters directly for BTST as it is literally practiced: a real BTST order is typically filled at or near the opening auction print, not at the continuous session's first one-minute bar price this paper uses throughout, so even this paper's purely descriptive overnight/intraday decomposition (Section 4) is built from prices that are not literally executable auction or fill prices. BTST exactly as the folklore describes it would therefore face even more implementation friction than anything modeled in this paper, not less.

2.2 The January 2023 T+1 transition

India moved from T+2 to T+1 settlement in a rollout that began 25 February 2022, ordered from the smallest stocks by market capitalization upward. The exchange's entire futures-and-options-eligible universe, the population this paper studies, moved in one final batch on 27 January 2023, unlike the staggered treatment applied to the broader, non-derivative market — a simultaneous, panel-wide policy change with no contemporaneous control group, rather than a staggered rollout that would support difference-in-differences identification. Using the staggered, non-F&O rollout as a quasi-natural experiment, Bhanu, Nath and Patnaik (2024) find that India's T+1 transition improved market liquidity overall, concentrated in smaller-capitalization stocks. A stock-fixed-effects panel regression around the F&O universe's own single transition date, run as a robustness exercise, gives the same point estimate under two different clustering choices but opposite significance conclusions, and a placebo test at other dates produces an even larger effect than the true event date — confirming this design cannot separate a T+1-specific effect from any other panel-wide shock, and we do not lean on it for any substantive claim.

2.3 Cost structure: three distinct layers

Capturing an overnight return requires holding a position across the close, which in Indian equities means a delivery trade rather than an intraday one for the overnight leg specifically, and this paper treats the resulting cost as three genuinely distinct layers rather than one. The first layer is fees: securities transaction tax, stamp duty, exchange transaction charges, SEBI turnover fees, GST on exchange and regulatory charges, and a flat depository participant charge per scrip on the sell leg, all built here from Zerodha's officially published 2026 rate card, with individual retail accounts (zero delivery brokerage) treated separately from company/LLP/HUF/trust accounts, which are not zero-brokerage under Zerodha's terms. The second layer is market impact — the price concession a real order of nontrivial size moves against itself. The third layer is bid-ask spread: a cost invisible to a fee-only model built from official rate cards, and, as Section 5 discusses, unusually difficult to estimate correctly for a return series built around large overnight price jumps. All three layers matter to this paper's conclusions, but, as Section 5 shows, none of them turns out to be the reason the paper's central execution discipline ultimately fails — that reason is more basic than any cost layer.

3. Data and Return Construction

The starting universe is 189 NSE-listed, futures-and-options-eligible stocks, chosen because F&O eligibility identifies a liquid, currently investable universe suitable for the implementation analysis in Section 5, not because F&O eligibility is itself part of the phenomenon under study. We obtained one-minute OHLCV data from 1 January 2020 to 31 December 2025 via Kite Connect for 185 of them (four were not retrieved, and no retry log survives to say why). One symbol, TMCV — Tata Motors' commercial-vehicle demerger entity, listed only since November 2025 — was dropped for insufficient history. The working panel is 184 stocks, 274,263 stock-days. A separate, genuine 2026-01-01 to 2026-09-11 extension for all 184 of the same names (31,832 stock-days, 173 trading days), pulled live and merged with zero missing, provides the forward-holdout window used throughout Section 5; the model, its hyperparameters, and every execution-discipline definition tested against it were fixed before this window existed — including the discipline whose apparent success later proved to be a bug, a point Appendix A returns to directly.

Overnight and intraday log returns are defined as $\text{overnight}(t) = \ln(\text{open}(t)/\text{close}(t-1))$ and $\text{intraday}(t) = \ln(\text{close}(t)/\text{open}(t))$, using the continuous-session open and close described in Section 2.1. Days with an absolute overnight return exceeding 8% are flagged and cross-checked against NSE's corporate-actions record and yfinance's split history; a flag only counts as a confirmed corporate action, and gets excluded, if it falls within two days of a recorded action. Of 603 flagged stock-days across the panel, 29 were confirmed corporate actions under this window — the rest, including RELIANCE's 23 March 2020 flag (the COVID-19 crash, not a corporate action), are kept as genuine data. The two-day window is a design choice, not a load-bearing one: varying it from 0 to 5 days moves the panel-wide overnight mean by at most 0.12 bps (17.94 to 18.06 bps). The panel splits temporally into an in-sample period (2 January 2020 – 29 December 2023, 182,634 stock-days) and an out-of-sample period (1 January 2024 – 31 December 2025, 91,629 stock-days); every specification choice in this paper, including liquidity quartile boundaries and machine-learning hyperparameters, is fixed using in-sample data only, and the 2026 window described above is fixed and untouched relative to every one of those choices as well.

4. The Overnight Premium Is Robust

4.1 The core result

Table 1 reports the core decomposition: a large, positive overnight return and a smaller, negative intraday return, both significant in each sample period. This is the number BTST folklore is actually about, and it holds up under every check we could construct. Nothing in this section is affected by the look-ahead bug discussed in Appendix A (summarized in Section 5.5): none of these figures involve a forward-looking selection decision of any kind, only the retrospective, backward-looking decomposition of already-realized daily returns.

Leg	Period	Mean (bps/day)	t-stat	p-value	n
Overnight	In-sample	+19.60	65.86	≈ 0	182,634
Overnight	Out-of-sample	+14.73	49.74	≈ 0	91,629
Intraday	In-sample	−11.75	−23.20	4.25e-119	182,634
Intraday	Out-of-sample	−10.96	−18.92	8.30e-80	91,629

Two-way (stock × date) cluster-robust inference, following Cameron, Gelbach and Miller (2011), which no longer treats correlated same-day observations across stocks as independent evidence, gives a stock-clustered t of 40.78, a date-clustered t of 9.69, and a two-way clustered t of 9.48 (p≈0). This two-way clustered figure is the primary, conservative inference standard this paper relies on throughout, because it does not treat 184 stocks moving together on the same date as 184 independent pieces of evidence. A naive per-observation Newey-West HAC estimate, which does not correct for that cross-sectional correlation, gives a considerably larger t-statistic of 81.04, but we do not treat that larger figure as the paper's primary evidence — it is reported only as a secondary, supplementary statistic. Every other t-statistic reported elsewhere in this paper for a leg-level, subsample, or adjusted figure (Table 1's per-leg t-stats, the beta-adjustment t-stats in Section 4.3, and similar) is, unless stated otherwise, the same naive per-observation HAC type as the 81.04 figure above, not independently re-computed under two-way clustering; the pooled, full-history two-way clustered t=9.48 (and, for the point-in-time-corrected sample discussed below, t=10.05) is this paper's reference point for how much these HAC figures should be discounted, rather than a claim that every subsidiary statistic has itself been re-clustered. A calendar-time block bootstrap of the pooled overnight mean (2,000 resamples, blocks of 10 consecutive trading dates, each block carrying every stock's observations for those dates) gives an observed mean of 17.93 bps and a 95% CI of [12.95, 22.16], with no resample mean at or below zero. Per-stock Wilcoxon signed-rank tests are significant in 182 of 184 stocks, with a positive median in all 184. All 184 sampled stocks post a positive full-sample mean overnight return; we report this purely as a distributional fact about the panel, not as an additional statistical test: the per-stock mean overnight return ranges from 2.50 bps/day (HDFCBANK, the weakest) to 36.56 bps/day (HINDCOPPER, the strongest), with a median of 18.23 bps/day and an interquartile range of 13.87–21.44 bps/day. We describe the breadth of the pattern across the panel this way, rather than with a sign-count significance calculation, because stock signs are not independent draws in this panel — Section 4.3 shows much of the raw premium is a shared market-beta effect, which would push many stocks' signs to move together rather than independently, making any such calculation uninformative as an inferential test regardless of how it is framed.

The estimate is insensitive to reasonable alternative constructions: value-weighting each day by traded value gives 20.20 bps (t=8.36); winsorizing at the 1st/99th percentiles gives 18.97 bps (t=112.96); excluding the least liquid quartile of stocks entirely gives 17.58 bps (t=69.47). None of these moves the estimate by more than about 2 bps or changes its significance.

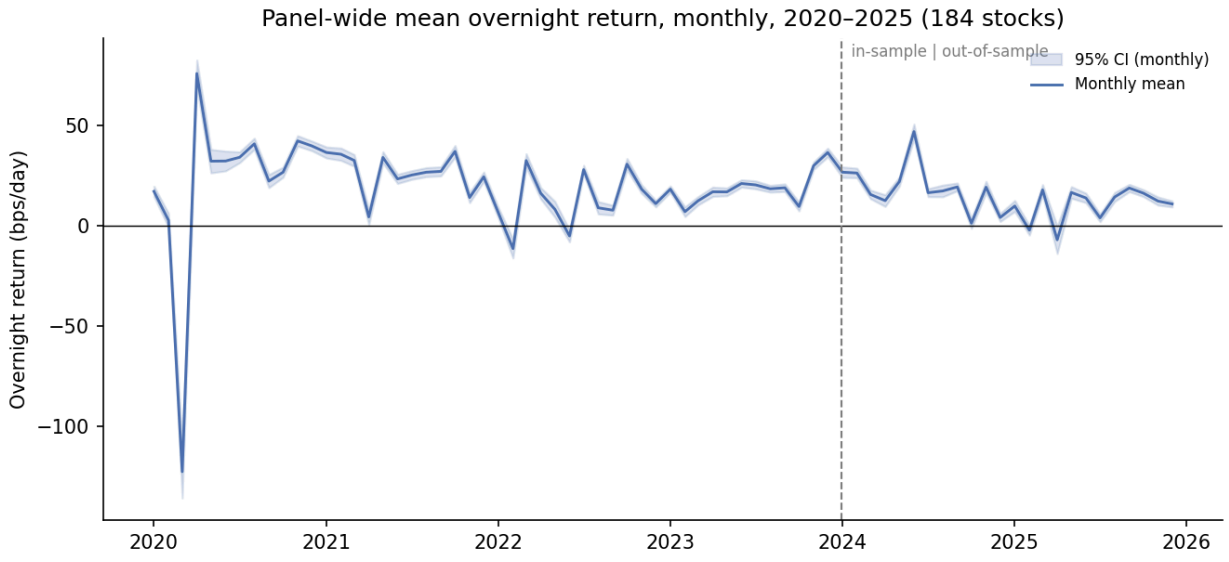


Figure 1. Panel-wide mean overnight return by month, 2020–2025, with a 95% confidence band. The dashed line marks the in-sample/out-of-sample boundary.

4.2 Breadth, stability, and a point-in-time correction

The result is not driven by a handful of extreme stocks: every one of 184 stocks shows a positive mean overnight return, the median stock premium is 18.23 bps/day (interquartile range 13.87–21.44 bps), the weakest stock (HDFCBANK, 2.50 bps) is still positive, and the strongest (HINDCOPPER, 36.56 bps) is roughly fifteen times larger. Splitting by calendar year, the premium is positive and significant every year from 2020 (19.93 bps) through 2025 (10.60 bps), peaking at 26.90 bps in 2021, with two-way clustered t-statistics ranging from 2.40 (2020) to 10.97 (2023) — every year clears the conservative standard, though not with equal statistical force.

A concern worth taking seriously and correcting rather than merely disclosing: the F&O-eligibility label used to build the universe is applied retrospectively, and 61 of the 184 panel stocks were not yet F&O-eligible for part of 2020–2025 (most joined in a cluster of exchange reviews across 2021). Recovering each stock's actual eligibility start date from archived NSE records and cross-referenced circulars, then re-estimating the pooled premium using only genuinely eligible stock-days, drops the sample from 274,263 to 253,019 stock-days (a 7.75% reduction) and moves the headline from 17.97 to 16.95 bps/day — a real, 5.70% correction. Applying the same two-way (stock × date) cluster-robust standard used as this paper's primary inference standard throughout, the PIT-corrected sample gives $t=10.05$ ($p\approx0$) — computed specifically for this corrected sample rather than left as an exception where only the supplementary per-observation HAC statistic (81.04 to 80.66, essentially unchanged) is reported. The correction changes no conclusion in this paper, under either standard; we report the corrected figure here, rather than only in an appendix, because a reader should not have to hunt for it. This point-in-time correction fixes exactly one issue — entry-date bias for the 61 stocks it identifies — and Section 7 states plainly what it does not fix: former F&O constituents, historical membership changes, the 4 originally-unretrieved symbols, and delistings or suspensions all remain unresolved.

4.3 Market adjustment and what survives it: two precise, un-blurred facts

A related concern is whether the pooled premium is substantially NIFTY 50's own overnight move propagating through nearly every constituent stock. Subtracting the index's own overnight return one-for-one (a unit-beta adjustment) reduces the premium from 17.97 to 6.93 bps/day ($t=42.68$, $p\approx0$). A more precise test uses each stock's own rolling 120-day beta to NIFTY, estimated from overnight-return covariance and re-estimated daily from strictly trailing data: on the beta-eligible subsample (19.16 bps base), the adjusted premium is 9.13 bps/day ($t=60.82$, median estimated beta 0.833). This corrects an earlier version of this calculation, which used a covariance estimator with a mismatched variance denominator (a ddof=0/ddof=1 inconsistency between numerator and denominator that inflated every rolling beta estimate by roughly 0.8%, giving a median beta of 0.840 and an adjusted premium of 9.04 bps); the correction is immaterial to this section's conclusion — the adjusted premium remains large and highly significant either way — but the corrected figures are the ones used throughout this paper. These two facts should be stated plainly and separately, not blurred into one number: a meaningful fraction of the raw premium, roughly a third to a half depending on method, is market-beta exposure, and a substantial, market-beta-adjusted component remains, significant at any conventional threshold.

The second, complementary fact concerns not the raw premium but the cross-sectional ranking skill used throughout Section 5: how much of a persistence ranking's edge over a same-sized random selection

survives once beta, liquidity, and volatility are removed jointly. A beta-matched random benchmark shows beta composition accounts for only about a fifth (21%) of the top-decile's advantage over an unconditional random draw (edge falls from 8.29 to 6.52 bps/day, t=6.20, once beta is matched exactly). A more complete neutralization — residualizing the persistence signal itself against beta, liquidity, and realized volatility jointly before ranking on the residual — shows the same pattern more starkly: average selected beta falls from 1.104 to 0.974, essentially to the panel median, while the edge over uniform random selection shrinks only from 9.24 to 8.09 bps/day, an 87% retention. Ranking skill, in other words, survives factor neutralization far better than the raw level survives market adjustment. Both facts are true at once and should be read together: the level of the overnight return is substantially a beta phenomenon; which stocks will rank where is substantially not. This distinction matters again in Section 5, where the ranking skill turns out to be real but the ability to monetize it does not follow automatically.

4.4 Regime stability

Splitting the panel by regime shows two genuine complications rather than uniform robustness, and we report them as such. During the five-month COVID crash window (February–June 2020), the mean overnight return is statistically indistinguishable from zero (−0.85 bps, t=−0.34) — the panel's overall average is carried by the other 94% of the sample by trading-day count. A market-direction split shows a large raw asymmetry (+37.05 bps on NIFTY up days versus −4.98 bps on down days), but this asymmetry shrinks by 106% under beta adjustment (to +7.65 versus +11.82 bps, both individually positive and significant) — the direction-dependence is essentially fully a market-beta effect, not a separate, unexplained state-dependence. Appendix A finds a related, and considerably starker, down-market weakness in the corrected Hold-Through strategy's actual day-level returns, worth reading against this same beta-driven asymmetry.

4.5 A methodological note: ex-ante prediction versus ex-post classification

Section 5 repeatedly builds a selection-continuation regression: a coefficient describing stock-days that are selected by the model on both day t and day t+1. Because this construction recurs throughout Section 5 and is easy to misdescribe, we state the relevant distinction plainly here, once, before it is needed. Three distinct objects are involved. (a) An ex-ante ranking, formed at the close of day t−1 using only information available through that moment. (b) The subsequent return, realized at day t, which the ranking in (a) is used to predict. (c) An ex-post classification of whether the same stock is also selected by the ranking rule on day t+1 — a classification that itself requires data through day t's own close, the same moment day t's own return becomes realized. Only (a) combined with (b) constitutes a predictive test: a ranking built from information available before the fact, checked against an outcome realized after it. Classification (c) is not available at any point before day t's return is already known, and a regression or descriptive statistic that conditions on (c) — however useful as a description of which stock-days happen to share a property — must never be described in this paper as a prediction, a validation of a real-time rule, or a live trading signal. Section 5.3 reports a genuinely ex-ante test in this sense (a same-day top-minus-bottom-decile spread, ranking at t−1 against realized return at t) and is the only construction in this paper to which the language of forward validation is applied. The continuing-selection coefficient used elsewhere in Section 5, by contrast, conditions on (c) and is reported strictly as a descriptive regression about which stock-days share a property, not as a predictive or forward-validated result, however similar its magnitude looks across samples.

5. Harvesting the Premium: A Secondary, Cautionary Investigation

Section 4 establishes that the overnight premium is real; that decomposition is this paper's primary contribution. This section is secondary and cautionary: it works through successive attempts to actually collect the premium, in increasing order of sophistication, presented in the order they actually occurred rather than cleaned up in hindsight, and asks whether any of them delivers a return that is both economically material and statistically robust net of realistic costs at realistic capital — a distinct question from whether the underlying overnight effect itself is statistically real, which Section 4 already establishes. Naive methods fail quickly and clearly. A moderately sophisticated method does better but still falls well short of an attractive result. A more sophisticated method appeared, for a substantial stretch of this project, to succeed convincingly, complete with what appeared at the time to be a forward-validated finding on genuine 2026 data — and that appearance turned out to rest on a bug, corrected once found; Appendix A gives the full, consolidated account of that discovery and correction (Section 5.5 summarizes it), presented there as a cautionary, audit-style account rather than a main line of argument. One further component, a bottom-decile short overlay built independently and correctly from the start, shows a positive point estimate but is not statistically distinguishable from zero under a formal test of its own daily returns (Section 5.7). We report this full arc, briefly elsewhere and in full only in Appendix A, because a paper that showed only the final, corrected numbers would understate how easily a backtest of this kind can produce a confident, wrong, and genuinely convincing positive result.

5.1 Naive blanket capture fails

The simplest possible implementation of BTST — buy every eligible stock every night, sell every morning, at the continuous-session bar prices described in Section 2.1 rather than literal auction fills — is catastrophic once realistic position sizing and fees are applied: spreading a fixed ₹20,00,000 starting capital across all 184 names nightly produces an average position size of roughly ₹10,870, at which the flat ₹15.34 per-scrip depository charge becomes a large fraction of the position (an average cost of 54.48 bps, more than double the 23.78 bps cost at a ₹1,00,000 position size). The result is a −63.2% CAGR (−86.1% total return) over the tested window, despite the underlying gross premium being genuinely positive. This is a position-sizing artifact layered on top of a real effect, not evidence the premium itself is spurious — but it is the first, sharpest illustration in this paper that the gross premium and a collectable strategy are two different things.

5.2 The Strategy A/B/C/D quartet: where the return actually comes from

A cleaner diagnostic decomposes a full buy-and-hold position into its constituent legs across the full 2020–2025 history: Strategy A (overnight-only, buy at close, sell at next open), Strategy B (buy-and-hold, close-to-close), Strategy C (intraday-long, buy at open, sell at close), and Strategy D (intraday-short, the mirror image of C, using the actual MIS intraday cost structure rather than the delivery cost structure, since a same-day round trip carries securities transaction tax on the sell leg only and no depository participant charge). Table 2 reports the full decomposition, gross, on both the 184-stock panel and NIFTY 50 itself.

Strategy	Panel CAGR (gross)	NIFTY 50 CAGR
A: overnight-only	+57.0% (Sharpe 4.05, MaxDD −29.6%)	+31.9% (Sharpe 2.49)
B: buy-and-hold (close-to-close)	+17.6% (Sharpe 0.94, MaxDD −40.0%)	+13.8% (Sharpe 0.81)

C: intraday-long	−25.1% (Sharpe −1.64, MaxDD −82.5%)	−13.8% (Sharpe −0.95)
D: intraday-short	+33.5% gross / +22.1% net (Sharpe 1.80, MaxDD −11.5%/−12.4% net)	+15.9% (Sharpe 1.10)

The decomposition identity, (1+overnight)(1+intraday) = (1+full), holds to numerical precision throughout, and the message is unambiguous: essentially all of the long-run positive return in the full buy-and-hold panel is overnight-sourced. Strategy A alone (+57.0%) comfortably exceeds full buy-and-hold (+17.6%); the intraday leg alone is sharply negative (−25.1%) and drags the combined result down from what the overnight leg achieves on its own. Strategy D, shorting the intraday leg using MIS-appropriate costs, is a useful confirmatory check rather than a standalone finding: at +33.5% gross / +22.1% net CAGR its net returns are positive in isolation under the fee-only cost model, but it is dominated by the far simpler Strategy A, which earns more with a cleaner cost structure and without needing to short anything. We read Strategy D as confirming the intraday leg is genuinely negative and genuinely shortable, not as a separate strategy worth pursuing on its own merits.

5.3 Persistence beats machine learning: a rejected robustness check

Every selection rule used in this paper's backtests ranks stocks by a simple 60-day trailing mean of each stock's own overnight return. Before settling on that rule, we tested whether a more sophisticated model does better, following the broader empirical asset-pricing literature's finding that tree-based methods are among the strongest general-purpose cross-sectional return predictors (Gu, Kelly & Xiu, 2020). A seven-feature LightGBM model — trailing overnight-return mean, calendar-days-since-close, prior-day intraday return, 20-day realized volatility, prior-day overnight return, a volume trend, and trailing liquidity, selected via a purged, embargoed, walk-forward cross-validation scheme within the in-sample period only — does not beat the one-feature persistence rule. On the identical five purged CV test folds, persistence's mean information coefficient is 0.1026 (fold SD 0.0269) against LightGBM's 0.0821 (fold SD 0.0462); persistence wins on 4 of 5 individual folds and is more stable fold-to-fold. Out-of-sample, the same ordering holds: 0.1227 for persistence versus 0.1088 for LightGBM, and a top-minus-bottom decile spread of 18.58 bps/day for persistence versus 14.90 bps/day for LightGBM. A ridge regression on the identical seven features does slightly better than persistence in both periods (in-sample IC 0.1074, out-of-sample 0.1326) — a small, genuine linear improvement — while the nonlinear, tree-based model does worse in both. The corrected reading is specific: nonlinearity does not help here, and neither, on net, do the six additional features beyond persistence itself; a small amount of genuine incremental information exists in a linear combination of them, but it is modest relative to persistence's own edge and does not change any of this paper's economic conclusions. We state this plainly because it reframes everything that follows: the model driving every backtest in this paper is fundamentally a persistence ranking, and the machine-learning comparison is a rejected robustness check, not the mechanism.

The top-minus-bottom-decile spread reported just above (18.58 bps IS, 14.90 bps OOS, t=9.78) is a genuinely ex-ante predictive test in the sense made precise in Section 4.5: the ranking is formed from data through the close of day t−1, and the spread is measured against the same day t's realized return, with no conditioning on any later day's information. Applying the identical construction, with no other change, to the genuine 2026 fresh-holdout window gives a top-decile gross return of +14.34 bps/day, a bottom-decile gross return of +3.21 bps/day, and an ex-ante top-minus-bottom spread of 11.13 bps/day (HAC t=4.17, p=3.05e-05, n=173 trading days) — smaller than the original discovery-window spread but a genuine, statistically significant replication on data the model never saw at the time it was specified. This is the

only result in this paper to which we apply the language of forward validation, precisely because it is built entirely from information available before the fact, checked against an outcome realized after it, with no ex-post classification of any kind — a later-sample replication under a frozen specification, not a claim of any stronger kind.

Because this is the single result this paper leans on most heavily, its implementation is specified precisely below, as implemented in `panel_2026_exante_spread_check.py`, rather than described in idealized terms. Table 2a records exactly what the code does at each of the six points a referee would want pinned down.

Element	Exactly what the code does
(a) Return leg predicted	TARGET = "overnight_ret", the same-day close-to-open log return ln(open(t)/close(t−1)) — not intraday_ret and not total_ret.
(b) Rank-formation cutoff	The ranking column `pred` is computed as final_model.predict(FEATURES) on a feature set built from data through close(t−1) only; the model itself is the one trained and frozen in panel_holdthrough_corrected.py, reused via runpy rather than retrained.
(c) Order timing required	To realize the predicted return, a position would need to be established at or immediately after close(t−1) — i.e., before the open of day t — in order to be held into overnight(t).
(d) Bar/price convention	Continuous-session first/last one-minute bars, per Section 2.1's project-wide convention: overnight_ret is inherited unchanged from the upstream panel (ln(open/prev_close) on continuous-session open and close), not the exchange pre-open or closing-auction print.
(e) Spread weighting	Equal-weighted: for each day, top_bps and bottom_bps are the simple arithmetic .mean() of overnight_ret across the top and bottom decile of names that day (k = max(1, round(10% × names/day))); the spread is not value-weighted or dollar-neutral.
(f) HAC t-stat aggregation	Computed on the daily portfolio-level spread series — one (top-minus-bottom) observation per trading day (n=173 for the 2026 window) — via OLS with HAC(maxlags=5) covariance, not on the underlying pooled stock-day observations.

5.4 Model-conditioned forced liquidation: still near breakeven at best

A model-conditioned improvement on blanket capture — rank all eligible stocks nightly by the persistence rule, take the top decile, and force-liquidate every position at the open regardless of whether the same stock is selected again the next night — substantially improves on blanket capture but still does not clear a bar of economic attractiveness. Against a panel-average out-of-sample gross overnight return of +14.73 bps/day, the top-decile selection earns +22.98 bps/day gross, but a full ₹20,00,000 compounding portfolio backtest under the Section 2.3 fee model returns only −0.5% CAGR for a retail account and −9.5% for a non-individual account (Table 3). A full cost-sensitivity sweep across five selection thresholds, three position sizes, and both account types (30 combinations) finds zero cases that are both net-positive and statistically significant at 5% — a clean null across the realistic parameter range, not a marginal miss at one specific choice of parameters.

Account	Avg. cost/position	Ending capital (₹20L start)	CAGR	Sharpe	Max drawdown
Retail	23.53 bps	₹19,80,312	−0.5%	0.00	−21.0%
Entity	27.29 bps	₹16,41,286	−9.5%	−0.92	−30.7%

This is the bar the paper's final, most sophisticated harvesting attempt needed to clear: not merely positive, but positive enough, and robustly enough, to be worth the operational complexity of an active strategy over simply holding the panel or the index. Section 5.5 describes an attempt that, for a while, appeared to clear that bar decisively.

5.5 The Hold-Through discipline: an apparent success, and how it was found to be an illusion

A model-conditioned discipline that held a continuing selection straight through its intraday session, rather than force-liquidating and rebuying it nightly as Section 5.4 does, initially appeared to solve the harvesting problem decisively: +75.0% CAGR at ₹20,00,000 retail capital on the discovery window, a result that appeared to replicate almost exactly on the genuine 2026 forward-holdout window (+41.2% CAGR), and that survived a full battery of stress tests we ran at the time. A subsequent, deliberately adversarial re-examination of the backtest's information timing — modeled on the scrutiny a skeptical referee would apply, not a fresh statistical test — found this rested on a look-ahead bug: the decision to hold a position through its intraday session rather than force-liquidate it was conditioned on whether the same stock is also selected the following day, a fact that is not knowable until that day's own close, the same moment the discarded intraday return itself becomes realized. Corrected, the finding reverses outright: −18.1% CAGR (retail) / −22.7% (entity) on the discovery window and −25.1% on 2026, negative at every capital level, cost scenario, and window tested — worse than the already-unattractive forced-liquidation alternative in every comparison. Appendix A gives the full, consolidated account: the discipline exactly as first specified and its original (uncorrected) results, the referee-style re-examination and the precise mechanism of the bug, an event-time specification table for the corrected mechanic, and the corrected results in full.

5.6 Market impact and cost layering: no scale or cost assumption rescues the corrected result

Because the corrected long sleeve (Section 5.5) is already negative on a fee-only basis, no cost layer applied on top of it can do more than confirm the result gets monotonically worse; none can reverse it. A calibrated square-root market-impact model leaves every capital level negative under every scenario tested, and three separate bid-ask spread estimators — ranging from a minute-bar estimate of 3.68 bps to daily Corwin-Schultz estimates of 70.78–71.98 bps, depending on method — each leave the strategy negative as well, from −22.0% to −72.2% CAGR across the discovery and 2026 windows and the full cost grid. Appendix B specifies the impact-model calibration and all three spread estimators in full, including why the daily and minute-bar estimates diverge as much as they do and which one this project regards as the more credible proxy.

5.7 The bottom-decile short: a discarded diagnostic and a separate, independently-built strategy

This subsection deliberately separates two distinct objects that must not be conflated: a discarded statistical diagnostic, and the actual short-sleeve strategy this paper reports. They share a research question — can the bottom-decile selection, the mirror image of the top-decile ranking, be shorted to add a second, negatively correlated sleeve? — but they are otherwise unrelated constructions, built at different times for different purposes, and only one of them was ever reported as a strategy.

The INITIAL BOTTOM-DECILE CONTINUATION DIAGNOSTIC was a statistical regression, not a trading rule: it tested whether stocks continuing in the bottom decile show a mirror-negative effect, conditioning on the stock still being in the bottom decile on day t+1. It found a strikingly strong negative effect (−46.44 bps on the discovery window, −62.30 bps on the 2026 holdout, both highly significant and correctly signed for a short), but that conditioning uses day t+1's ranking, which requires data through day t's own close — the same look-ahead timing flaw described for the Hold-Through mechanic in Appendix A.2. Once this was recognized, the diagnostic was correctly discarded as unusable for a live trading rule. It was never itself reported as a strategy in this paper, has no CAGR, Sharpe, or drawdown figures anywhere in this draft, and was not "repaired" into the strategy reported below — it was abandoned.

The ACTUAL SHORT-SLEEVE STRATEGY reported throughout this paper (with the +2.4%/+12.2% CAGR figures below) is a completely separate, independently-built construction, using only same-day, real-time-available information: it ranks strictly on information available before today's open and takes its intraday short position for the day using only that information, with no continuation decision of any kind — because Indian cash equities cannot be shorted overnight or held short across multiple days (only via derivatives, which this paper does not pursue), the short overlay was always confined to a single intraday MIS leg. This strategy was built from scratch after the diagnostic above was discarded, not derived from or repaired from it, and was walk-forward-legitimate from the moment it was first specified; it was never subject to the look-ahead flaw that affected either the discarded diagnostic above or the Hold-Through mechanic of Section 5.5.

As an isolated position, this independently-built bottom-decile intraday short shows a positive point estimate: +2.4% CAGR (Sharpe 0.24, MaxDD −13.6%) on the discovery/OOS window and +12.2% CAGR (Sharpe 0.99, MaxDD −8.0%) on the 2026 window, at half of a ₹20,00,000 total capital allocation. A formal significance test computed for this revision, run directly on the sleeve's daily net returns, does not support calling this an established, statistically distinguishable-from-zero edge: an HAC-adjusted t-test (5 lags) gives a mean of +1.32 bps/day (t=0.40, p=0.69) on the discovery window and +4.87 bps/day (t=0.83, p=0.41) on the 2026 window, and a calendar-time block bootstrap (2,000 resamples, 10-day blocks, the same method used elsewhere in this paper) gives 95% confidence intervals of [−4.22, +7.23] bps/day and [−6.78, +15.82] bps/day respectively — both spanning zero, with 32.95% and 20.15% of resamples non-positive. This is the one piece of this project's active-strategy work whose point estimate survives every walk-forward and look-ahead check applied to it, but the point estimate itself is weak evidence rather than an established result, and we report this honestly rather than describing it as statistically positive. Combined 50/50 with the corrected (now-negative) long Hold-Through sleeve, the merged portfolio substantially improves on the long-only result and cuts drawdown — from −19.9%/−26.7% CAGR long-only (half capital) to −8.1% (discovery) and −7.9% (2026) combined, with maximum drawdown falling to −20.9% and −8.5% respectively — but the combined portfolio remains net-negative in both windows. The short sleeve's point estimate is directionally helpful at the margin, but it is neither statistically established on its own terms nor large enough to make the combined book profitable.

A separate, unfavorable qualification about the 2026 figures above must be stated plainly rather than left to a footnote. This paper's core model was frozen — specification, hyperparameters, and every execution-discipline definition fixed — at commit 746d590 (10 September 2026), strictly before any 2026 data existed in this project, which is what licenses calling the 2026 window a genuine forward holdout for that model and for the ex-ante spread in Section 5.3. The short-sleeve scripts (`panel_bottomdecile_continuation_test.py`, `panel_shortoverlay_backtest.py`, and the corrected/combined scripts that followed them) carry no comparable freeze: file timestamps show they were first written on 11 and 12 September 2026, after the 2026 raw data pulls for all 184 names had already completed (by roughly 13:44 IST on 11 September) and after the merged 2026 daily file existed (from roughly 15:15 IST the same day). "Walk-forward-legitimate," as used elsewhere in this subsection, describes only the mechanic's information timing — that each day's decision uses no information from a later day — and remains true of the short sleeve on that narrow, mechanical definition. It does not mean, and should not be read to imply, that the 2026 figure is an out-of-sample or forward test in the sense the ex-ante spread of Section 5.3 is: the code that specifies this strategy was written with the 2026 data already sitting in the working directory, so the +12.2% CAGR and the HAC/bootstrap statistics on the 2026 window are in-sample results relative to when that code was authored, on the same data used to build

it, not a held-out replication. We did not find evidence of the reverse — no version of these scripts predates the 2026 data pull — and we report this honestly rather than allowing the short sleeve's 2026 number to be read as carrying the same validating weight as Section 5.3's genuinely ex-ante spread.

5.8 Horizon decay: a standing structural offset, not a temporal reversal

A further test, unaffected by the Hold-Through mechanic bug since it uses the same descriptive continuing-selection regression rather than a compounding backtest, asks whether the overnight/intraday offset documented throughout this paper decays or reverses at longer horizons. Run on the 2026 genuine holdout using the momentum-controlled continuing-selection specification at forward horizons of 1, 5, 20, and 60 trading days, the overnight-leg coefficient stays significant and grows with horizon ($t=2.27$ at $h=1$ up to $t=3.07$ at $h=20$, easing only slightly to $t=2.82$ at $h=60$); the intraday-leg coefficient stays negative and of roughly offsetting magnitude throughout, its own significance weakening at $h=60$ ($t=-0.90$); the combined total-return coefficient is consistently positive but never reaches conventional significance at any tested horizon ($t\approx0.25$–0.90). The two legs are pulling in opposite directions at every horizon tested — a standing, non-decaying structural offset, not a pattern that resolves itself over time, and a useful reminder that the underlying overnight/intraday split documented in Section 4 is a persistent structural feature even though, as Section 5.5 shows, no execution discipline built on top of it survived being made real-time-implementable and correctly costed.

This result is worth contrasting directly with two related findings in the literature. Lou, Polk and Skouras (2019) document a month-ahead temporal reversal between overnight and intraday returns in U.S. data — today's overnight-minus-intraday spread predicts next month's intraday-minus-overnight spread, a genuine sign-flip over time. This paper's horizon-decay test finds something different: not a reversal over time, but a standing offset that holds at every horizon simultaneously, with overnight persistence the more robust of the two legs throughout. More strikingly, Chen et al.'s Taiwan study (Pacific-Basin Finance Journal, 2023) decomposes momentum into an intraday (IMOM) and overnight (OMOM) component and finds the opposite sign assignment from this paper's: in their Taiwanese cross-section, the intraday leg is the continuation (momentum) leg and the overnight leg is the reversal leg, essentially a mirror image of what this paper documents for Indian equities. Whether this reflects a genuine cross-market difference in the mechanism generating overnight and intraday returns, or a difference in what each paper's model is actually selecting on, is a question this paper's evidence cannot resolve, but the contrast is worth stating plainly rather than leaving implicit.

6. Discussion

Return to the BTST claim this paper opened with. It is right about where the edge lives: overnight capture, not intraday holding, is where essentially all of this panel's long-run positive return originates (Section 5.2), and that overnight return is real by every dependence-robust standard applied in Section 4. It is wrong that this edge is straightforwardly exploitable. Naive blanket capture is catastrophic (Section 5.1); a reasonably sophisticated model-conditioned discipline, force-liquidated nightly, does not clear a bar of economic attractiveness (Section 5.4); and the paper's most sophisticated attempt rested, for a time, on a look-ahead bug rather than a genuine edge, a full account of which is given once, in Appendix A (Section 5.5 summarizes it). Even a well-motivated, carefully stress-tested execution discipline can rest on an unrealizable assumption about the timing of available information, and catching that requires deliberately adversarial scrutiny of the exact information timing at every decision point, not merely internal confidence, however extensive the stress-test battery applied to the wrong mechanic.

This paper's central empirical claim sits against Zaremba and Nikorowski's (2018) broader finding that most documented emerging-market anomalies do not survive contact with real trading frictions. This result does not even reach the point where that finding would apply: the paper's most promising harvesting method failed a more basic test — correct real-time implementability — before transaction costs became the deciding factor at all (Appendix A, whose methodological lesson — that a passed stress-test battery does not itself prove a mechanic is real-time-realizable — is not re-derived here). Once the mechanic is corrected, costs are consistently negative but not decisive the way they appeared to be for the false-positive version of the result; the corrected fee-only baseline is already negative, and every cost scenario tested only deepens that loss, never reverses it, because there is no positive result left to reverse. Haghani, Ragulin and Dewey's individual-stock overnight-drift work in developed markets remains the flagship modern reference for the underlying phenomenon; this paper extends it to a cross-sectional emerging-market setting, but where we had hoped to also extend it with a genuine, cost-surviving execution discipline, what we can report instead is a documented account of why one specific, carefully constructed discipline does not survive scrutiny. Kamath (2026, SSRN) studies a related question in India at the NIFTY-futures, index level, not cross-sectionally across individual stocks, and does not attempt the execution-discipline analysis this paper builds. Bogousslavsky (2021) remains the closest cross-sectional precedent for decomposing intraday and overnight legs across a set of anomalies in the canonical literature, but does not pursue an execution-cost analysis for any of them; this paper's contribution relative to that line of work is specifically the descriptive decomposition of Section 4, with the execution analysis of Section 5 as a secondary, cautionary complement.

7. Limitations

- This is a single-market study. Every result is specific to 184 stocks that were F&O-eligible when this panel was built, with successfully retrieved continuous histories — not a claim about Indian equities generally, and not a claim that a comparable exercise in another emerging market would find the same magnitudes or even the same sign pattern. Because the universe is not fully point-in-time (Section 4.2 corrects one specific entry-date bias but leaves several others unresolved, below) and corporate-action screening is a threshold-based approximation rather than a full adjustment pipeline (next bullet), the claim that the premium is broad and persistent across the sampled universe should be read as applying to this specific, imperfectly-reconstructed panel, not as a fully general claim about the Indian equity market during this period.
- The genuine forward-holdout window (Section 5.5, Section 5.6, Appendix A, Appendix B) covers roughly nine months of 2026 trading. This is real out-of-sample evidence for both the underlying overnight premium (Section 4) and the corrected Hold-Through result (Appendix A.3), not a discovery-window artifact in either direction, but nine months is a short window against which to certify any conclusion's long-run robustness; and, as Section 5.7 states plainly, the short sleeve's own 2026 number does not carry this same forward-holdout status, since its code was written after the 2026 data already existed.
- Every cost estimator used in this paper — the fee model, the market-impact calibration, and both the daily and minute-bar bid-ask spread estimators — is model-based, not verified against actual exchange order-book fills or a live execution log. This did not end up being the deciding factor for this paper's central strategy conclusion (Appendix A.3 shows the corrected result is negative under every cost scenario tested, from the cheapest to the most expensive), but the underlying cost estimates themselves

remain proxies, not measurements, and a reader should not treat the specific spread figures used here (3.68–71.98 bps depending on estimator, detailed in Appendix B) as exchange-verified numbers.

- Survivorship is only partially addressed, not resolved. The point-in-time F&O-eligibility correction in Section 4.2 fixes exactly one issue — entry-date bias for the 61 stocks that joined the F&O segment mid-sample — and does not establish a full historical point-in-time universe. Several separate, open problems remain entirely unresolved by that correction: former F&O constituents that left the segment are not reconstructed; historical membership changes among stocks never captured in this panel at all are unknown; 4 of the original 189-symbol target universe were never successfully retrieved and no retry log survives to explain why, so we cannot characterize whether their absence is random or informative; and delistings or suspensions affecting any of the 184 currently-covered names, before or during the sample, are not separately identified or corrected for. The direction and magnitude of the net bias from these unresolved problems combined is unknown.
- The 8%-threshold corporate-action screen (Section 3, already sensitivity-tested there across window widths) is a screen, not a full security-level corporate-action adjustment pipeline. It catches large, discrete jumps but does not individually audit every stock for splits, bonuses, rights issues, spin-offs, or vendor-supplied price adjustments whose return impact falls below that 8% threshold; some number of such smaller, unaudited adjustments could still be present in the data. Because this screen is a threshold-based approximation rather than a full adjustment pipeline, and the universe itself is only partially point-in-time (previous bullet), this paper's claim that the premium is broad and persistent should be understood as a claim about this specific, imperfectly-reconstructed panel, not a fully general claim about the Indian equity market during this period.
- This paper's risk adjustment throughout (Section 4.3) uses only a single NIFTY-beta factor, in both a unit-beta and a rolling-beta form. It does not construct or control for a fuller set of standard Indian equity risk factors (size, value, momentum, quality, or similar), which would be necessary to support any claim of a risk-adjusted excess return net of standard risk exposures, as opposed to the narrower claim this paper actually makes: return in excess of a NIFTY-beta-adjusted benchmark. Building a fuller factor model is unattempted future work, not something this draft resolves.
- The down-market weakness documented for the corrected Hold-Through strategy (Appendix A.3) and for the raw premium's beta-driven asymmetry (Section 4.4) is disclosed but not fully mechanistically resolved: it is consistent with, but not proven to be, a broader market-wide momentum-crash and beta-exposure dynamic rather than something specific to this model's selections.
- The market-impact model (Section 5.6, detailed in Appendix B) is dual-anchored between an academic square-root-law prefactor and a calibration solved from NSE's own published Impact Cost benchmark, itself built around a much larger order size than this strategy trades; the resulting low/base/high range is a defensible calibration, not a single verified number.
- Open and close prices are continuous-session bars, not the NSE pre-open call-auction or closing-auction prices, which one-minute Kite data does not carry (Section 2.1); an execution-price robustness check across 1-, 5-, and 15-minute definitions confirms the premium's existence and sign throughout but shows roughly a 2 bps range in its exact magnitude across definitions.
- This paper deliberately does not build a beta-hedged, futures-based implementable market-neutral overlay that would account for margin and roll costs — the same derivatives infrastructure Section 5.7 declined to build for the bottom-decile short investigation, for the same reason: it is out of scope for a paper centered on long-only and single-leg-intraday cash-equity disciplines.

- No spread estimator used in this paper — daily or minute-bar — has been validated against real quoted or effective spread data, or against trade-level, order-book data of any kind; this project never had access to either. Both estimators remain proxies built from OHLCV price data.
- The five-track audit described in Appendix C was internal and author-conducted, not independent external validation. Independent validation would require public release of the code and data and third-party replication; neither has been done for this draft. What the audit can and does establish is the absence of the specific, named classes of error it deliberately checked for — look-ahead bias in every remaining backtest mechanic, arithmetic errors in return/cost compounding, inconsistencies in variance/covariance estimation, data-integrity and universe-construction defects, and formula errors in the cost and spread estimators — not a general guarantee against every conceivable defect, including classes of error this audit did not think to check for, and not the kind of assurance a third party re-running the code from scratch would provide. Concretely, and itemized rather than left as a general caveat, what is absent that would upgrade this from an internal assertion to external evidence: (i) publicly released code and data, so a third party can actually run the pipeline; (ii) a data dictionary or manifest describing every field, source, and transformation in the processed panel; (iii) file checksums over the raw and processed data, so a reader can verify they are looking at the same inputs this draft used; (iv) a pinned, published environment/dependency specification, as opposed to the locally-frozen one described below; and (v) automated tests specifically designed to fail if any feature, selection rule, or fill-price assignment in the pipeline were ever changed to use future information — as opposed to the one-time, author-conducted inspection this audit actually was. None of these five exist for this draft. A reader should weigh this paper's conclusions accordingly: with real confidence in the specific checks reported, and appropriate humility about what a single author's internal audit, absent all five of the above, cannot prove.
- The full codebase and processed 184-stock daily panel behind this draft, along with the 2026 forward-holdout extension and the corrected scripts described in Section 5.5 and Appendix A, are frozen locally with pinned dependency versions so this specific draft's numbers stay independently reproducible; as of this draft that snapshot has not yet been pushed to a public host or archived with a DOI, and none of the five artifacts itemized in the previous bullet (public code/data, data dictionary, checksums, a published pinned environment, or automated leakage tests) has been produced. A full, publicly versioned reproducibility archive is future work, not something this draft provides.

8. Summary of Key Results

Result	Figure	Interpretation
Pooled overnight premium	+17.97 bps/day, two-way clustered t=9.48 (primary); supplementary HAC t=81.04	Real, large, panel-wide (descriptive decomposition, Section 4)
Breadth of sign	184/184 positive; per-stock mean 2.50–36.56 bps/day, median 18.23 bps/day (Section 4.2)	Widely shared across the panel
Point-in-time eligibility correction	16.95 bps; two-way clustered t=10.05 (primary); supplementary HAC t=80.66	Survives a real universe correction under either standard
Beta-adjusted premium (corrected)	+6.93 to +9.13 bps/day	Partly, not wholly, a beta effect

Ranking-skill edge after factor neutralization	9.24 → 8.09 bps/day (87% retained)	Ranking skill survives far better than the raw level
Persistence vs. LightGBM, identical CV folds	IC 0.1026 vs. 0.0821, wins 4/5 folds	Persistence, not ML, is the mechanism
Blanket capture, OOS	−63.2% CAGR	Naive execution fails badly
Forced liquidation, ₹20L retail, fee-only	−0.5% CAGR, 0/30 sensitivity cells significant	Model-conditioned but not a robust, cost-surviving result
Hold-Through, original (buggy) mechanic	+75.0% CAGR discovery / +41.2% 2026 — SUPERSEDED	A look-ahead bug produced a false-positive result; see Section 5.5 and Appendix A
Hold-Through, corrected mechanic, ₹20L retail	−18.1% CAGR discovery (retail) / −22.7% (entity) / −25.1% 2026	Negative at every capital level, cost scenario, and window tested
Hold-Through under every tested cost scenario	−16.7% to −72.2% CAGR across the full grid	No cost assumption rescues it; the fee-only baseline is already negative
Bottom-decile short overlay (mechanic walk-forward-legitimate from the start; 2026 figure is in-sample relative to when the code was written — Section 5.7)	+2.4% CAGR (Sharpe 0.24) discovery; +12.2% (Sharpe 0.99) 2026; HAC t=0.40/0.83, bootstrap CI spans zero both windows	Positive point estimate; not statistically distinguishable from zero, not independently established as tradeable, 2026 figure not a forward test
Combined 50/50 long-short portfolio (corrected)	−8.1% CAGR discovery; −7.9% 2026	Real improvement over long-only, still net-negative

9. Conclusion

BTST — buy near the close, hold overnight, sell at the open — is folklore, and this paper set out to test it as rigorously as a documented academic finding deserves to be tested, rather than either dismissing it or taking it on faith. The folklore is right about where the edge lives, in this panel: the close-to-open return in this universe of 184 F&O-eligible Indian equities is large, positive, pervasive across stocks and years, and survives every dependence-robust statistical standard, point-in-time universe correction, and market-adjustment check we could construct (Section 4) — this decomposition, not the execution work that follows, is the paper's primary contribution. It is wrong that this edge is something a real-time trader can straightforwardly turn into a robust, cost-surviving strategy. Naive blanket capture is catastrophic. A reasonably sophisticated model-conditioned version, force-liquidated nightly, does not deliver an attractive result under realistic 2026 costs. A third, more sophisticated discipline appeared, for a substantial part of this project's life, to solve the problem decisively, before a deliberate, adversarial re-examination of its information timing found this rested on a look-ahead bug rather than a genuine edge; Section 5.5 summarizes and Appendix A gives the complete, consolidated account of that discovery and correction, and we do not re-narrate it here.

Correcting the mechanic reverses the finding outright: negative CAGR at every capital level tested, for both retail and non-individual accounts, on both the original discovery window and the genuine 2026 holdout, under every cost scenario tested, and worse than the already-unattractive forced-liquidation alternative in every comparison. The one component of this project's active-strategy work that is genuinely walk-forward-legitimate in its mechanic and shows a positive point estimate in isolation — a bottom-decile intraday short overlay, built independently of and never subject to the flawed long-strategy construction — is not itself statistically distinguishable from zero under a formal HAC t-test and block-bootstrap confidence interval computed on its own daily returns (Section 5.7); combined with the corrected long sleeve it improves the portfolio meaningfully but leaves it net-negative in both windows.

As Section 5.7 details, that overlay's own 2026 figure is a separate matter from the mechanic's walk-forward legitimacy: the code that specifies it was written after the 2026 data already existed in this project, so that figure is an in-sample result relative to when it was written, not a forward test.

The paper's final position, stated at the level of precision it can support, in this panel: close-to-open returns are positive and broadly distributed; straightforward long-only attempts to monetize the pattern do not appear robust after realistic implementation constraints. The underlying overnight return premium in this universe is real and statistically overwhelming under a conservative two-way clustered inference standard. A genuinely ex-ante top-minus-bottom-decile spread — ranked only on information available before the day's open, with no conditioning on any later day's information — replicates on genuine, previously-unobserved 2026 data (11.13 bps/day, HAC t=4.17, against 14.90 bps/day, t=9.78, in the original discovery window): a later-sample replication under a frozen specification, and the only result in this paper that legitimately qualifies as forward validation. Every harvesting method this project tested — blanket capture, forced liquidation, and the corrected Hold-Through discipline, under any cost assumption, capital level, or parameter choice — fails to deliver a return that is both economically material and statistically robust, and the one component with a positive point estimate (the short overlay) is not itself statistically distinguishable from zero under a formal test of its own returns, nor is its 2026 figure a forward test in the sense the ex-ante spread above is. A specific, initially highly convincing look-ahead bug very nearly produced the opposite, wrong conclusion for the long sleeve; Section 5.5 summarizes and Appendix A documents its discovery and correction in full, presented there as a cautionary, audit-style account rather than a main line of argument. We report that account, and the results of this project's internal five-track audit (Appendix C) — an author-conducted check against named error classes, not independent external validation, and one still missing public code/data, a data dictionary, file checksums, a published pinned environment, and automated leakage tests, per Section 7 — because how the bug was caught is itself a methodological contribution a reader building similar work may find as useful as the null result. We also note that many specifications were explored over the course of this research program, and no formal multiple-testing correction has been applied to any individual result reported here.

References

Cooper, M. J., Cliff, M. T., & Gulen, H. (2008). Return Differences between Trading and Non-Trading Hours: Like Night and Day. Working Paper, University of Utah. Available at SSRN: https://ssrn.com/abstract=1004081

Berkman, H., Koch, P. D., Tuttle, L., & Zhang, Y. J. (2012). Paying Attention: Overnight Returns and the Hidden Cost of Buying at the Open. Journal of Financial and Quantitative Analysis, 47(4), 715–741. https://doi.org/10.1017/S0022109012000270

Lou, D., Polk, C., & Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. Journal of Financial Economics, 134(1), 192–213. https://doi.org/10.1016/j.jfineco.2019.03.011

Akbas, F., Boehmer, E., Jiang, C., & Koch, P. D. (2022). Overnight returns, daytime reversals, and future stock returns. Journal of Financial Economics, 145(3), 850–875. https://doi.org/10.1016/j.jfineco.2021.09.019

Chui, A., Ranganathan, K., Rohit, A., & Veeraraghavan, M. (2023). Momentum, reversals and liquidity: Indian evidence. Pacific-Basin Finance Journal, 82, 102193. https://doi.org/10.1016/j.pacfin.2023.102193

Heston, S. L., Korajczyk, R. A., & Sadka, R. (2010). Intraday Patterns in the Cross-Section of Stock Returns. The Journal of Finance, 65(4), 1369–1407. https://doi.org/10.1111/j.1540-6261.2010.01573.x

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Sampath, A., & Gopalaswamy, A. K. (2020). Intraday Variability and Trading Volume: Evidence from National Stock Exchange. Journal of Emerging Market Finance, 19(3), 271–295. https://doi.org/10.1177/0972652720930586

Bhanu, A., Nath, G. C., & Patnaik, T. (2024). Unlocking Liquidity through Shortened Settlement Cycle: Empirical Evidence from India. Economics Letters, 239, 111736. https://doi.org/10.1016/j.econlet.2024.111736

Shleifer, A., & Vishny, R. W. (1997). The Limits of Arbitrage. The Journal of Finance, 52(1), 35–55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. Journal of Business & Economic Statistics, 29(2), 238–249. https://doi.org/10.1198/jbes.2010.07136

Corwin, S. A., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. The Journal of Finance, 67(2), 719–760. https://doi.org/10.1111/j.1540-6261.2012.01729.x

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. The Journal of Finance, 39(4), 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Haghani, V., Ragulin, S., & Dewey, D. Night Moves: Overnight vs. Intraday Stock Returns. Working paper / practitioner research, cited here for its documentation of individual-stock overnight drift in developed markets.

Kamath, S. (2026). Overnight and Intraday Returns in Indian Index Futures. Working paper. Available at SSRN.

Zaremba, A., & Nikorowski, J. (2018). Trading costs, market anomalies and the efficiency of emerging market stocks: a note. Economics and Business Letters / related working paper on frictions and emerging-market anomaly survival.

Chen, C.-H., et al. (2023). Intraday and overnight momentum in an emerging market. Pacific-Basin Finance Journal (Taiwan IMOM/OMOM decomposition).

Bogousslavsky, V. (2021). The cross-section of intraday and overnight returns. Journal of Financial Economics, 141(1), 172–194.

Novy-Marx, R., & Velikov, M. (2016). A Taxonomy of Anomalies and Their Trading Costs. Review of Financial Studies, 29(1), 104–147. https://doi.org/10.1093/rfs/hhv063

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. Journal of Financial Economics, 122(2), 221–247. https://doi.org/10.1016/j.jfineco.2015.12.002

Zerodha Broking Ltd. Charges — brokerage, taxes and fees on trading and investing. https://zerodha.com/charges/ (accessed 2026; rates as published, subject to change).

NSE Impact Cost methodology and published Nifty 50 constituent impact-cost benchmarks, cited as the institutional anchor for the market-impact calibration in Section 5.6.

NSE, BSE, CDSL and NSDL joint press releases and circulars on the phased transition to T+1 settlement (2021–2023), cited as institutional/regulatory background for Section 2.2.

Appendix A: The Hold-Through Look-Ahead Bug — Mechanism, Discovery, and Correction

This appendix gives the full, consolidated account summarized in Section 5.5. It is presented as a cautionary, audit-style account of a specific error and its correction, not as a main line of the paper's argument, and it is deliberately written in the order events actually occurred. A.1 presents the Hold-Through discipline and its results exactly as they were understood at the time — a genuinely positive, well-supported finding that we believed, at the time, to be forward-validated by every check we thought to run; Section 4.5 explains why that belief rested on conflating an ex-post classification with a predictive test. A.2 describes a subsequent, deliberately adversarial re-examination of the backtest's information timing that identified a look-ahead bug invalidating A.1's central number. A.3 reports the corrected result.

A.1 The discipline, and the result as it was first found

Section 5.4's forced-liquidation convention has a specific mechanical consequence worth isolating directly: on any night the model selects the same stock it selected the night before, forced liquidation sells it at that morning's open and immediately rebuys it before that evening's close, discarding the entire intervening intraday session and paying a full round-trip cost to do so. What does that discarded intraday session actually look like? On the 4,416 stock-days where the model selects the same stock on consecutive days, the mean intraday return is +18.76 bps/day — sharply positive, against a panel-wide out-of-sample intraday mean of −10.96 bps/day — and the mean overnight return on those same continuing stock-days is +27.17 bps/day. The idea, as originally conceived: hold a continuing selection straight through the session instead of force-liquidating and rebuying it, capturing both legs together, and turn over the portfolio only when the selection actually changes.

As originally implemented, the backtest's decision rule for a position currently held (selected on day d) was: at day d's close, check whether the same stock is also selected for day d+1; if so, hold straight through with no transaction cost that day and capture overnight(d) plus intraday(d); if not, sell at day d's open (capturing overnight(d) alone) and pay the exit cost. Under this rule and the identical fee-only cost model used for forced liquidation, the discipline returned +75.0% CAGR at ₹20,00,000 retail capital (Sharpe 3.61, max drawdown −8.9%, average realized cost per position 11.65 bps) and +70.7% for a non-individual entity account (Sharpe 3.45, MaxDD −9.3%), against forced liquidation's −0.5%/−9.5%. A momentum-controlled regression appeared to rule out conventional short-term price momentum as the explanation: a continuing-selection dummy without controls carried a coefficient of +31.22 bps (t=6.62), and adding prior 5-day return, prior 20-day return, and prior-day intraday momentum as controls did not shrink this toward zero — it grew, to +34.26 bps (t=6.92).

A full stress-test suite, run at the time, appeared to confirm this was not a fragile, single-specification result. A calendar-time block bootstrap (2,000 resamples, 10-day blocks) gave a 95% CI of [+38.4%, +116.4%] for CAGR. Varying the selection threshold (TOP_FRAC) from 5% to 20% moved the CAGR from +98.3% to +49.3% while the momentum-controlled coefficient remained highly significant throughout — the effect did not appear fragile to the exact decile threshold chosen. A concentration check appeared to rule out a handful-of-days artifact: dropping the best 10 days out of 498 retained 72.5% of the full-sample CAGR. Most importantly, a genuine, never-touched-before 2026 forward-holdout window,

pulled and merged after the model, its hyperparameters, and the discipline itself were all frozen, appeared to replicate the result almost exactly: a momentum-controlled continuing-selection coefficient of +35.59 bps (t=5.20) on the 2026 window, against +36.17 bps (t=6.46) on the original discovery window, and a capital sweep from ₹5,00,000 to ₹1,00,00,000 returning CAGRs from +34.0% to +43.1% with no wipeouts at any tested level (headline Rs.20L: +41.2% CAGR, Sharpe 2.05, MaxDD −11.7%). By every check we thought, at the time, to run — stress tests, parameter sensitivity, a genuine forward holdout — this appeared to be a robust, real, and substantial trading edge. One openly disclosed weakness was noted even then: a regime split showed a much weaker, sometimes insignificant effect on down-market days, a pattern flagged as consistent with a broader market-wide momentum-crash dynamic rather than something specific to this model.

A.2 A simulated referee's question, and the bug it found

Before treating this as a finished result, we subjected it to the kind of adversarial scrutiny a skeptical journal referee would apply — not a fresh statistical test, but a line-by-line re-examination of exactly what information the backtest used at each decision point and exactly when that information would actually have been available to a real-time trader. That exercise asked a specific, narrow question about the hold/exit decision described in A.1: at day d's close, the backtest checks whether the stock is also selected for day d+1, and that check requires computing day d+1's ranking feature — which is itself built from data through day d's own close. The question the referee-style review asked was blunt: is that check actually available at the moment the backtest uses it?

It is not. Day d+1's selection depends on data through day d's close — the exact same moment day d's own intraday return becomes realized. The original mechanic effectively let the backtest decide, in advance of day d's intraday session, whether to “skip” that session for a stock that would (unknowably, at that time) turn out not to be selected the next day. No real-time trader can make that determination before day d's intraday session has actually happened, because the input to that determination does not exist yet at day d's open. This is a look-ahead bug: the asymmetric return treatment between “continuing” and “exiting” positions used information that, at the point in time the backtest applied it, could not have been known.

The fix, implemented in a separate, corrected backtest script (`panel_holdthrough_corrected.py`), is direct: every stock selected on day d — continuing or newly entered, with no exception — now captures the full overnight(d) + intraday(d) return that day, unconditionally, because intraday(d) is not something a real-time trader can selectively avoid based on tomorrow's (unknowable) selection. The continuing/exiting/new-entry distinction now affects only transaction-cost timing: a continuing position (selected today and tomorrow) is held with no cost charged that day; an exiting position (selected today, not tomorrow) pays its sell-leg cost at today's close, after already capturing today's full return; a new entry pays its buy-leg cost, exactly as every position already does under forced liquidation. The cost-accounting logic itself was never the buggy part — only the decision about which return to accrue was. Table A1 lays out the exact event-time mechanic for each of the four position types the corrected code distinguishes, so the timing claims above can be checked directly against what the backtest actually does at each decision point, rather than taken on the strength of the prose alone.

Event type	Signal observed (when)	Order submitted (when)	Assumed execution price	Position held (period)	Exit decision (when/how)	Order executed (when)	Cost charged (leg)

New entry (selected day t, not held on day t−1)	Close(t−1), using data through close(t−1)	Economically as of close(t−1)	Close(t−1)	Held through close(t), capturing overnight(t) + intraday(t) in full	At close(t), using data through close(t) (day t+1's ranking, now legitimately known) — determines tomorrow's continuing/exiting label, not today's return	Buy leg executed at close(t−1)	Buy leg, charged on day t
Continuing position (selected on day t and day t+1)	Position already open from a prior entry; day t+1's inclusion resolved only at close(t)	None — no new order; position simply continues	N/A (no trade this day)	Held through close(t) and into day t+1, capturing overnight(t) + intraday(t) in full	Deferred: continuation is confirmed at close(t) using data through close(t)	None this day	None charged on day t
Exiting position (selected on day t, not day t+1)	Position already open (or newly entered same day); exit resolved at close(t)	At close(t)	Close(t)	Held through close(t), capturing overnight(t) + intraday(t) in full before exit	At close(t), using data through close(t): day t+1's ranking excludes this stock	Sell leg executed at close(t)	Sell leg, charged on day t
Position held on the final day of the sample	Position already open (or newly entered that day)	At close(final day) — the code treats every position as exiting on the last day, since there is no day t+1 selection set to check against	Close(final day)	Held through close(final day), capturing overnight + intraday in full for that day	Forced by the end of the backtest window, not by a ranking comparison — mechanically identical to the exiting case because next-day selections are undefined	Sell leg executed at close(final day)	Sell leg, charged on the final day

Table A1. Event-time specification of the corrected Hold-Through mechanic, one row per position type distinguished in `panel_holdthrough_corrected.py`. The final-day row is a mechanical consequence of ending the backtest on a fixed date, not a separate modeling choice: because no day-t+1 selection set exists after the last sample date, every position open on that date is treated identically to an ordinary exiting position and pays the sell leg.

A.3 The corrected result

Under the corrected mechanic, the entire finding reverses. At ₹20,00,000 starting capital, Hold-Through returns −18.1% CAGR (Sharpe −0.83, MaxDD −40.1%) for a retail account and −22.7% (Sharpe −1.10, MaxDD −45.2%) for an entity account on the original discovery window — worse than forced liquidation's −0.5%/−9.5% in both cases. On the 2026 forward-holdout window — a later-sample replication under a frozen specification, per Section 5.3's terminology — the corrected result is −25.1% CAGR (Sharpe −1.32, MaxDD −19.5%), against forced liquidation's −20.3% — again worse. This is not a narrow reversal: a full capital sweep from ₹5,00,000 to ₹1,00,00,000 is negative at every level tested, for both account types, on both windows (discovery-window retail CAGR ranges from −23.5% at ₹5L to −16.7% at ₹1Cr; entity from −40.9% to −17.6%; the 2026 window is similarly negative throughout). A block-bootstrap 95% CI on the discovery-window result is [−36.7%, +5.9%], with only 5.7% of resamples showing a positive CAGR. Varying TOP_FRAC from 5% to 20% keeps the result negative throughout (−23.3%, −18.1%, −17.5%). A regime decomposition of the corrected strategy's actual day-level returns

shows the discipline is net-positive only on market up-days (+54.51 bps/day, roughly +137% annualized, on the discovery window; +73.08 bps/day, roughly +184%, on 2026) and sharply negative everywhere else, including a catastrophic down-day result (−75.85 bps/day, roughly −191% annualized, discovery; −99.39 bps/day, roughly −250%, 2026) — consistent with Section 4.3's finding that the raw premium's up/down asymmetry is substantially a market-beta effect, not a strategy-specific edge. The full cost-layer analysis that follows (market impact and bid-ask spread) is given in Appendix B; it only ever confirms this result gets worse, never reverses it.

Appendix B: Bid-Ask Spread Estimation and Cost-Layer Detail

This appendix gives the calibration and estimator detail summarized in Section 5.6, underlying the headline cost sweeps reported in Appendix A.3. Because the corrected Hold-Through long sleeve is already negative fee-only, before any spread or impact cost is applied, everything in this appendix is a secondary measurement exercise and caveat, not a central part of this paper's argument: it can only confirm the result gets monotonically worse, never reverse it.

We calibrate a square-root market-impact model, impact (bps) = C × daily volatility (bps) × √participation rate, using two independent anchors: the academic square-root law (C=0.34) and a calibration solved directly from NSE's own published Impact Cost benchmark, which requires Nifty 50 constituents to keep impact cost at or below 50 bps for a ₹10 crore basket 90% of the time (C=1.9430, solved against this panel's liquid-subset median daily traded value of ₹410.6 crore), with a simple average of the two (C=1.1415) as a base case. Forced liquidation (Strategy A) is essentially unaffected in relative terms across this same impact grid, since it is already near breakeven or modestly negative at every capital level under the fee-only baseline; the corrected Hold-Through discipline (Strategy B) is uniformly negative across the same grid. A market-impact sweep built on this calibration leaves every capital level negative under every scenario tested (−16.7% to −30.0% CAGR across the grid).

Three bid-ask spread estimates were tried. A daily Corwin-Schultz panel-median estimator gives 70.78 bps round-trip; an overnight-jump-corrected variant of the same estimator gives 71.98 bps; and a considerably smaller minute-bar estimate, which this project regards as the more credible proxy for an intraday-held position, gives 3.68 bps, because it never observes the close-to-open transition and so avoids the overnight-jump bias the daily estimator carries. That minute-bar figure is cross-checked against an alternative proxy estimator built from within-session price data, which is not itself validated against actual quoted or effective spreads from trade-level or order-book data, which this project did not have access to. Under any of the three, the strategy remains negative: −22.0% to −68.9% CAGR at ₹20,00,000 retail on the discovery window, and −28.8% to −72.2% on the 2026 window. Additional flat slippage on top of the 3.68 bps base case only deepens the loss further. No cost assumption tested — from the cheapest to the most expensive — turns Hold-Through into a viable strategy, because the fee-only baseline it would need to survive costs on top of is already negative.

Appendix C: The Five-Track Internal Audit

This appendix gives the full account of the internal audit summarized in Section 5.5 and referenced throughout Section 7's limitations. Finding one look-ahead bug of the kind described in Appendix A, embedded deeply enough to survive an extensive internal stress-test battery, raised an obvious further question: was anything else in this project's pipeline built the same way? We conducted a five-track audit of the full codebase to check. This audit was internal and author-conducted, tested against five specific,

named classes of error — look-ahead timing, compounding conventions, statistical formula correctness, data integrity, and spread-estimator formulas — and is not independent external validation, which would require public code/data release and third-party replication; neither has been done for this draft, and Section 7 restates this limitation.

Within that scope: the model, feature, and cross-validation track found no leakage against the look-ahead class specifically — every feature's strictly-trailing construction, the purged and embargoed walk-forward CV scheme, and the in-sample/out-of-sample split were re-inspected against that one criterion, consistent with the independent determinism check reported in this project's results ledger. The backtest-arithmetic track found the bug described in Appendix A, plus one further, smaller issue: a latent merge bug in the script that combines the Hold-Through long sleeve with the bottom-decile short overlay (Section 5.7) using an inner join that would silently truncate the combined simulation if either sleeve were ever bankrupted before the other's window ended; direct testing confirmed this condition never actually occurred with real data in either window, so it never affected any reported number, but it was fixed for correctness going forward. The statistical-inference track found one further issue, unrelated to look-ahead bias: a rolling-beta calculation (Section 4.3) that computed a covariance estimate and a variance estimate with inconsistent variance denominators, inflating every beta estimate by roughly 0.8% — immaterial to any conclusion, but corrected. The data-integrity/universe track and the spread/impact-formula track found no defects against the specific criteria checked; the universe construction, corporate-action screening, and cost-estimator formulas used throughout this paper were re-inspected against those named error classes and no defect of those kinds was found.

We report this audit not as proof that the pipeline is free of every possible error, and not as external validation of any kind — Section 7 returns to this limitation directly, itemizing five specific artifacts that are not present (public code/data release, a data dictionary/manifest, file checksums, a published pinned environment, and automated tests designed to fail on future-information leakage) — but as evidence that the specific class of error responsible for Appendix A.1's false result was searched for deliberately, by the author, across the rest of the project, rather than left unexamined once the one instance was found and fixed. A reader should weigh this audit accordingly: as a real, itemized internal check against five named failure modes, and not as a substitute for any of the five external artifacts listed above.