Look-Ahead Bias in Multi-Timeframe Backtests

How joining higher-timeframe linear-regression signals to lower-timeframe execution manufactures backtest profits

αλλο research
Independent Researcher
allobiancofiore@gmail.com

September 2026

Abstract

Multi-timeframe trading strategies read a signal from slow bars and trade on fast ones, so the slow series must be joined onto the fast timeline. The common way of doing this, keyed on each slow bar's opening time, gives every fast decision the final value of a slow bar that is still forming. We show that for a least-squares trend slope, and for any linear filter of closes, the naive value equals the causal value plus a fixed multiple of the price move still to come, and we derive the expected size of the resulting leak in closed form. On E-mini Nasdaq-100 futures from 2020 to 2023, a 15-minute regression model traded on one-minute bars earns a Sharpe ratio of 2.06 after costs with the naive join and 0.33 with a causal one; the naive version also profits on random walks. The closed-form prediction, registered before it was tested, matches the measured leak within the registered band on Nasdaq-100 and S&P 500 futures, on their random walks and for a moving-average crossover. Because the leak changes the signal on only 0.8–5.2% of minutes, a truncation test on the signal detects it in 0.56% of checks, against 83% on the raw slope. No causal variant has a tradeable edge. The study was pre-registered; two of its five main criteria failed, and we report why.

Keywords: look-ahead bias, backtesting, multi-timeframe analysis, linear regression, trend following, intraday futures, pre-registration

JEL classification: C52, C58, G14, G17

All results use in-sample data (2020–2023); data from 2024 onwards were sealed before the study and never read. The pre-registrations, code and the ledger of every test are available from the author on request. Market data: Databento.

1. Introduction

Many intraday trading strategies combine two timeframes. A trend is read from slow bars, say 15 minutes long, and trades are timed on fast bars of one minute. The slow chart is meant to filter out noise; the fast chart is meant to improve entries. A natural choice for the slow signal is the slope of a least-squares line through the last N closes. It is smooth, easy to interpret and offered by charting platforms as a standard "linear regression" indicator.

Backtesting such a strategy requires a step that looks harmless in code: the slow series must be joined onto the fast one. At 10:07 the one-minute strategy needs "the 15-minute slope". But the 15-minute bar that covers 10:00 to 10:15 is still forming at 10:07. Its close, and every slope that includes it, will not be known until 10:15. A join that matches each fast bar to the slow bar with the same *opening* time hands the 10:07 decision the 10:15 value. This is look-ahead bias, and it is easy to introduce by accident. pandas labels resampled intraday bars by their opening time, TradingView stamps bars by their opening time, and Freqtrade stamps them by their opening date. A join on any of these stamps, written without care, gives each fast bar the final values of an unfinished slow bar.

The problem is known in practice. TradingView and Freqtrade both warn about it in their documentation, and Sobreiro, Martinho, Martins and Vardasca (2026) recently documented it in a machine-learning study of Bitcoin. There, merging higher-timeframe features by opening time raised the classifiers' ROC-AUC from about 0.60 to between 0.73 and 0.81. What is missing is an account of the leak in the terms that matter to anyone who trades on a backtest: how it works, how much profit it manufactures, and why the usual checks do not find it.

This paper provides that account. We study a deliberately minimal model, the sign of a least-squares slope on 5-, 15- or 30-minute bars traded on one-minute bars, on E-mini Nasdaq-100 (NQ) and E-mini S&P 500 (ES) futures from 2020 to 2023. We compare the naive join with three causal alternatives. The design, the decision rules and a later addendum were registered before the corresponding results were computed. Data from 2024 onwards were sealed and never read.

We make five contributions.

First, we derive the leak exactly. For a least-squares slope, and for any other linear filter of closes, the naive value equals the causal value that a live trader sees plus a fixed multiple of the price move still to come before the slow bar closes. From this identity we derive a closed-form expression for the expected size of the leak. The prediction, registered before it was tested, falls within the registered band in every setting on both futures markets, on random walks built from their returns, and for a second trend signal.

Second, we measure the leak in trading profit after costs. With 15-minute bars and a 50-bar window, the naive NQ strategy earns $393,000 over four years on one contract, with a Sharpe ratio of 2.06. The same strategy with a causal join earns $68,000, with a Sharpe ratio of 0.33. Across settings the

naive join produces Sharpe ratios of up to 5.35 on NQ and 4.42 on ES, and up to 7.58 for a moving-average crossover.

Third, we show that the leak is mechanical. On random walks, where no strategy can have an edge, the naive join still earns between $90,000 and $182,000 in the same setting, while the causal joins earn nothing systematic. The naive join's advantage shrinks minute by minute as the join is delayed and disappears exactly when the delay reaches a full slow bar.

Fourth, we explain why common checks miss it. The naive and causal signals disagree on only 0.8% to 5.2% of minutes, so statistics averaged over all minutes barely move. Those minutes, however, are exactly where the signal changes sign and a trend-following strategy trades. A truncation test that compares signals detects the leak in 0.56% of checks; the same test on the raw slope value detects it in 83%.

Fifth, we report the study as registered. Two of the five criteria registered for the main study failed, because they relied on an averaging statistic that dilutes a leak concentrated in a few minutes. We report those failures, explain them, and label every analysis added after the results were seen.

Once the join is causal, nothing is left to trade. No causal variant, in either market or with either signal, reaches the significance bar we set for the project as a whole. This is consistent with evidence that short-term trend following has earned little since about 2009 on futures with small ticks relative to their volatility, a group that includes equity-index futures (Kurth, Eisler, Rej and Bouchaud, 2026). In our data, almost all of the naive backtest's performance comes from information about the future.

The paper is organised as follows. Section 2 reviews related work on look-ahead bias, backtest overfitting and trend signals. Section 3 describes the alignment problem, defines the four joins and derives the leak and its expected size. Section 4 describes the data and design, and Section 5 presents the results. Section 6 turns them into practical checks, Section 7 discusses limitations, and Section 8 concludes.

2. Related literature

This paper draws on three literatures: work on look-ahead bias and data leakage, work on false discoveries in backtests, and work on trend signals. Each contributes a piece of our argument, and none measures the multi-timeframe alignment leak in trading terms.

2.1 Look-ahead bias and data leakage

Look-ahead bias is among the oldest recognised errors in empirical finance. Banz and Breen (1986) showed that the standard COMPUSTAT database introduced look-ahead and ex-post selection biases into studies of accounting data. Portfolios formed from it earned returns significantly different from those formed from a database free of both biases, enough to change the conclusions of a hypothesis test. The general lesson, that data must enter an analysis only once they would have been available, has since become standard in the design of backtests (López de Prado, 2018).

Computer science treats the same problem under the name *leakage*. Kaufman, Rosset, Perlich and Stitelman (2012) define it as the introduction of information about the prediction target that would not legitimately be available at prediction time. Noting that leakage has been counted among the "top ten data mining mistakes", they propose to avoid it by strictly separating what is learned from what is predicted. Fonseca (2026) formalises look-ahead freedom as temporal non-interference: the future must not influence the present. His framework separates the time a datum refers to from the time it becomes available, and he gives a checker that proves pipelines of windows, resampling operations and joins free of leakage. He also shows that commonly used empirical detectors miss leaks that his checker catches.

The multi-timeframe case has only recently been documented. Sobreiro, Martinho, Martins and Vardasca (2026) build classifiers for Bitcoin on features from 15-minute, 4-hour, daily and 3-day bars. They report that merging the slower bars by their opening time raised cross-validated ROC-AUC from 0.57–0.61 to 0.73–0.81, and that shifting each slower bar forward by one period before the merge removes the effect. They note that this form of the bias had not been discussed explicitly in the literature and that it is difficult to detect from classification metrics alone. Their backtests use only the corrected features, so the size of the bias is reported in classification accuracy rather than trading profit. Practitioners have documented the same error in software. TradingView warns that its higher-timeframe request function returns future data on historical bars when look-ahead is enabled (TradingView, 2026a). Freqtrade warns that a plain merge of a slower timeframe can cause look-ahead bias because its bars are stamped by their opening time (Freqtrade, 2026).

We build on this work in three ways. We explain the mechanism exactly for linear signals, we predict the size of the leak, and we measure it in net trading profit on futures, together with the reasons it escapes common checks.

2.2 False discoveries in backtests

A second literature studies how backtests produce results that do not survive. Its main concern is selection: when many strategies or specifications are tried, the best of them looks significant by chance. Lo and MacKinlay (1990) analyse the resulting data-snooping biases in tests of asset pricing models. White (2000) proposes a bootstrap "reality check" that accounts for the full set of rules searched. Sullivan, Timmermann and White (1999) apply it to technical trading rules and find that the best rule's apparent superiority on the Dow Jones Industrial Average does not carry over to the following ten years.

Later work turns these ideas into tools for evaluating strategies. Bailey, Borwein, López de Prado and Zhu (2014) show how easily backtest overfitting produces strategies that fail out of sample. They later estimate the probability of such overfitting through combinatorially symmetric cross-validation (Bailey, Borwein, López de Prado and Zhu, 2017). Bailey and López de Prado (2014) propose a deflated Sharpe ratio that corrects for the number of trials and for non-normal returns. Harvey and Liu (2015) derive haircuts for Sharpe ratios under multiple testing. Harvey, Liu and Zhu (2016) argue

that a new factor should clear a *t*-statistic of about 3 rather than 2. Harvey (2017) examines the incentives that favour the publication of positive results and proposes reporting standards to counter them. Arnott, Harvey and Markowitz (2019) collect these lessons into a protocol for backtesting.

We use this literature in our design, through a project-wide ledger of tests, a search-adjusted significance bar and pre-registration. Our main subject, however, is a different kind of error. The alignment leak requires no search: it biases a single backtest of a single rule, and a correction for multiple testing does not remove it. A deflated Sharpe ratio computed from a leaked backtest would still be inflated by the leak.

2.3 Trend signals

The signals we study come from the trend-following tradition. Moskowitz, Ooi and Pedersen (2012) document time-series momentum across futures markets at monthly horizons. Levine and Pedersen (2016) show that time-series momentum and moving-average crossovers are equivalent in their general forms, and that both belong to the class of linear filters. Any such filter can be written as a weighted sum of past returns, which they display as a "trend signature". Our analysis of the regression slope relies on this linearity. Bruder, Dao, Richard and Roncalli (2011) review econometric trend estimators, including linear regression, for use in momentum strategies.

Evidence on short-horizon trend signals is weaker. Brock, Lakonishok and LeBaron (1992) report profitable moving-average and trading-range rules on the Dow Jones index, and Sullivan et al. (1999) show that the best of them failed after the sample period. In their survey, Park and Irwin (2007) find that early studies reported profits from technical rules in foreign-exchange and futures markets but not in stocks, and that many studies suffered from problems in their testing procedures. At the intraday horizon, Gao, Han, Li and Zhou (2018) find that the first half hour's return on an S&P 500 exchange-traded fund predicts the last half hour's. Most relevant to our setting, Kurth, Eisler, Rej and Bouchaud (2026) show that short-term trend following has earned little since about 2009 on futures whose tick is small relative to their volatility, a group that includes equity-index futures. They attribute this to the way high-frequency market makers withdraw liquidity in front of predictable directional orders.

This evidence gives a prior: a causal short-horizon trend signal on NQ or ES should have little or no edge. A backtest that shows a large edge from such a signal should therefore be examined for errors before it is believed. This paper describes one such error, shows how large it is, and explains why it is easy to miss.

3. The alignment problem

A multi-timeframe model computes its signal on slow bars and acts on fast ones. Before the model can act, the slow signal has to be placed on the fast timeline, and this step decides what information each decision can see. This section explains how the step goes wrong, defines four ways of

performing it, and shows that for linear trend signals the resulting error can be written down exactly and its expected size predicted.

3.1 Two times for every bar

Every bar has two times. Its *reference time* is the moment it opens. Its *availability time* is the moment it closes, when its close, high, low and volume are final and anything computed from them becomes known. For a bar of length τ opened at S, the availability time is $S + \tau$. Fonseca (2026) formalises the same distinction for data pipelines in general. Look-ahead freedom requires that every datum enters a decision only after its availability time, whatever its reference time.

Data platforms stamp bars with one time or the other, and the defaults differ (Table 1). A join that treats a reference time as an availability time hands each fast decision the final values of the slow bar it falls inside. With a 15-minute slow bar, a decision at 10:07 then receives a value that will not exist until 10:15.

Table 1. Bar time-stamp conventions and their default behaviour when a higher-timeframe series is joined to a lower-timeframe one.

Environment	A bar is stamped with	What the default join delivers	Source
pandas `resample`	Left edge (open) for minute and hour frequencies; bins closed on the left	A backward as-of join on the stamp assigns every fast bar the slow bar that contains it, including that bar's final close	pandas development team (2026)
TradingView Pine Script	Opening time (`time`); closing time is a separate variable (`time_close`)	`request.security` with look-ahead enabled returns the slow bar's final values on historical bars; the documented safe form offsets the series by one bar (`expr[1]`)	TradingView (2026a, 2026b)
TradeStation EasyLanguage, MultiCharts	Closing time (`Time`)	Stamps equal availability times. The risk arises when data or code move between this convention and an open-stamped one	TradeStation (2026)
Freqtrade	Opening time (`date`)	The documentation warns that a plain merge of a slower timeframe "can implicitly cause a lookahead bias", and provides a helper that shifts the slow series by one bar	Freqtrade (2026)

None of these conventions is wrong in itself. Each is internally consistent and documented. The leak arises at the boundary, when a series stamped by reference time is joined as though the stamp were its availability time. Sobreiro et al. (2026) meet exactly this case in a machine-learning study of Bitcoin: open-time stamps combined with a backward as-of join raised their classifiers' ROC-AUC from about 0.60 to 0.73–0.81.

3.2 The model

We study a deliberately minimal model, so that any difference between its variants can come only from the join. Let $C_1, \ldots, C_N$ be the closes of the last N slow bars, each of length τ minutes, indexed from oldest to newest. The slope of an ordinary-least-squares line through them is a fixed weighted sum of those closes,

$$b = \sum_{k=1}^{N} w_k C_k, \qquad w_k = \frac{k - (N+1)/2}{S_{xx}}, \qquad S_{xx} = \frac{N(N^2-1)}{12} \tag{1}$$

and the model's trend signal is $D = \text{sign}(b)$. The model trades on 1-minute bars in the direction of D. It has no entry filter, stop or target.

A least-squares slope is a standard trend estimator (Bruder, Dao, Richard and Roncalli, 2011) and is offered by charting platforms as a "linear regression" or "least-squares moving average" indicator. For this paper its important property is linearity. Levine and Pedersen (2016) show that every linear trend filter, including moving-average crossovers and time-series momentum, is a weighted sum of past prices or returns. Everything derived below for the slope therefore holds for that whole class; only the weights change. We use $\tau \in \{5, 15, 30\}$ minutes and $N \in \{20, 50, 100\}$.

3.3 Four ways to join

A 1-minute bar that opens at u closes at $T = u + 1$ minute, and T is the decision time. When T falls strictly inside a slow bar $[S, S + \tau)$, that bar is still forming: its eventual close, C_{final}, is not yet known, and the current price is c_T. Let $r = S + \tau - T$ be the number of minutes the forming bar still has to run. We compare four joins (Table 2).

Table 2. The four joins. "Leak" is the largest amount of future data a decision can use.

Join	The decision at T uses	Typical origin	Leak
J0, naive	The slope over the N slow bars ending with the forming bar, at that bar's final close C_{final}	A join keyed on the slow bar's opening time (Table 1)	Up to $\tau - 1$ minutes
J1, developing bar	The same N bars, with the forming bar entered at the current price c_T	What a live chart displays before the bar closes	None
J2, last closed bar	The slope over the N slow bars ending with the last bar whose close is at or before T	A backward as-of join on closing times	None
J3, rolling one-minute	The slope over the last $\tau \cdot N$ one-minute closes	No join at all	None

J0 is the join that an open-time stamp produces. The three alternatives are all causal and differ only in how they treat the forming bar. J1 includes it at its current price, J2 ignores it until it closes, and J3 avoids slow bars altogether. J1 is the natural comparison for J0 because the two use exactly the same window; J2 is the conventional correct implementation, and the one a practitioner would use in place of J0.

Figure 1 shows the difference on one decision. On 20 May 2022, at T = 10:46, the 15-minute bar opened at 10:45 had fourteen minutes left to run. J2 used the close of the bar that ended at 10:45. J0 used the close at 11:00, which came after a decline of more than 100 points. With $N = 50$, the causal slope pointed up and the naive slope pointed down.

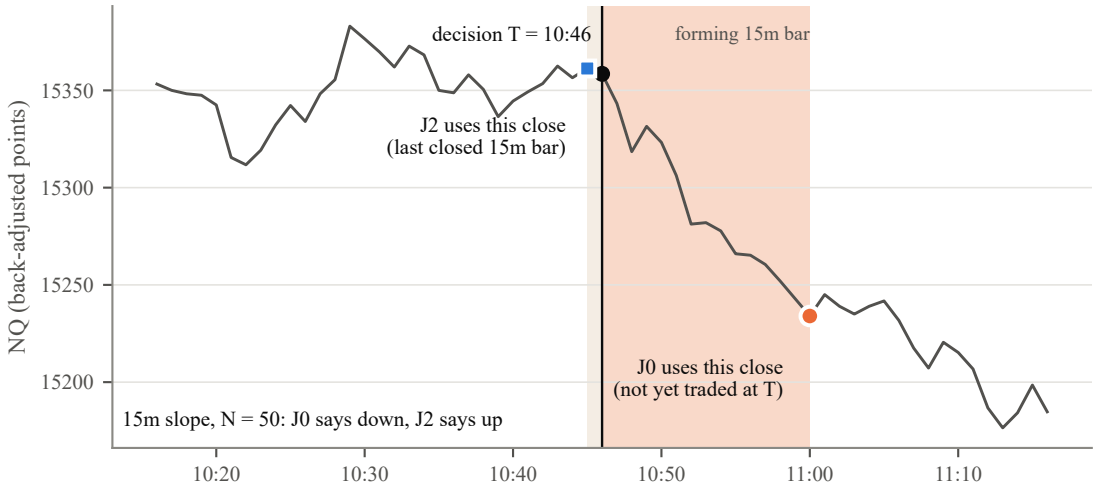


Figure 1. One decision on real NQ data, 20 May 2022. At T = 10:46 the 15-minute bar that opened at 10:45 is still forming. The causal join (J2) uses the close of the last completed 15-minute bar (blue square). The naive join (J0) uses the close at 11:00 (orange dot), which lies 14 minutes in the future. With $N = 50$ the causal slope points up and the naive slope points down.

3.4 The naive slope contains the future

Because J0 and J1 use the same N bars and differ only in the value entered for the newest one, equation (1) gives an exact identity:

$$b_{J0} = b_{J1} + w_N F, \qquad F = C_{\text{final}} - c_T, \qquad w_N = \frac{6}{N(N+1)} \tag{2}$$

The naive slope equals the slope a live trader sees plus a positive multiple of F, the price move from the decision to the end of the forming bar. F is part of the very move the strategy is about to trade. Three properties follow directly from (2).

The leak needs no predictability in prices. Even if price is a martingale, so that b_{J1} carries no information about F, the sign of $b_{J1} + w_N F$ is positively related to F. A backtest using J0 is therefore long more often before rises and short more often before falls, on any price path.

The leak is concentrated where the signal changes sign. When $|b_{J1}|$ is large relative to $w_N |F|$, adding the second term leaves the sign unchanged and the two joins agree. Only when the causal slope is close to zero can the future move decide the sign. Those are exactly the moments at which a sign-following strategy trades.

The leak shrinks with the window and vanishes at the bar's close. The weight w_N falls roughly as $1/N^2$: it is 0.0143 for $N = 20$, 0.0024 for $N = 50$ and 0.0006 for $N = 100$. The typical size of the slope falls more slowly, so shorter windows leak more. When $r = 0$, the decision coincides with the bar's close, $F = 0$, and J0 equals J1.

The identity is not specific to regression. For any linear filter $S = \Sigma_k\, v_k\, C_k$ of slow-bar closes, the same argument gives $S_{J0} = S_{J1} + v_{newest} \cdot F$, where v_{newest} is the filter's weight on the newest close. A crossover of simple moving averages of lengths $f < s$ has $v_{newest} = 1/f - 1/s$, for example 0.075 for $f = 10$ and $s = 40$. That weight is much larger than the regression weights above, so the corresponding leak is larger. We test this generalisation in Section 5.7.

3.5 How large is the leak?

Equation (2) also makes the size of the leak predictable. Let $D_{J0} = \text{sign}(b_{J0})$ and $D_{J1} = \text{sign}(b_{J1})$, and measure the leak in the signal as the average difference in the move captured,

$$L = \mathrm{E}\left[(D_{J0} - D_{J1}) F \right]$$

in index points per decision.

Proposition 1. *If b_{J1} and F are independent, zero-mean normal variables with standard deviations σ_b and σ_F, then*

$$L = \sqrt{2/\pi}\, \frac{w_N \sigma_F^2}{\sqrt{\sigma_b^2 + w_N^2 \sigma_F^2}} \tag{3}$$

The proof is in Appendix C. The first step is that $\text{sign}(b_{J1})$ is independent of F and contributes nothing. The second is that, for jointly normal variables, the expected value of F given b_{J0} is proportional to b_{J0}, so the expectation reduces to the mean absolute value of a normal variable.

Equation (3) has a simple reading. When the causal slope varies much more than the leaked term ($\sigma_b \gg w_N\, \sigma_F$), the leak is approximately $\sqrt{(2/\pi)} \cdot w_N\, \sigma_F^2 / \sigma_b$. It grows with the weight on the newest bar and with the square of the move still to come, and it falls as the signal becomes more decisive. The formula applies to any linear filter once w_N is replaced by that filter's weight on the newest close.

Real prices violate both assumptions: returns are fat-tailed, their volatility clusters, and the slope may carry some information about the next few minutes. The proposition is therefore a benchmark rather than an exact description. In applying it we allow σ_b and σ_F to vary with the minutes left in the forming bar and with the volatility regime, and evaluate (3) within each such group (Section 4). Section 5.6 compares the prediction with the leak measured on two futures markets, on data from which any predictability has been removed, and for two kinds of trend signal.

4. Data and design

The design has three aims. The model should be simple enough that the join is the only thing that varies. Costs should be realistic enough that the results mean something in trading terms. And every analysis should be fixed before its result is seen.

4.1 Data

We use the front-month E-mini Nasdaq-100 (NQ) and E-mini S&P 500 (ES) futures traded on the Chicago Mercantile Exchange. The source is per-contract one-second data from Databento, aggregated to one-minute bars. A continuous series is formed by taking, on each day, the contract with the highest volume on the *previous* trading day, so the choice of contract never uses same-day information. Prices are difference back-adjusted, which preserves point changes across a roll. Contract values are $20 per index point for NQ and $50 for ES; both have a tick of 0.25 points.

Times are US Eastern. A trading day runs from 18:00 to 17:00 on the following calendar day and carries the later date. The sample is 1 January 2020 to 31 December 2023, which gives 1,032 trading days with a regular session. Bars from November and December 2019 are read only to warm up rolling windows. Data from January 2024 onwards form an out-of-sample period that was sealed before the study began: the code asserts that no bar at or after 1 January 2024 is loaded, and none was.

Slow bars of $\tau \in \{5, 15, 30\}$ minutes are built from all one-minute bars, overnight included, and aligned to the clock. Because every τ divides the hour and the trading day starts on the hour, no slow bar spans two trading days. A slow bar's availability time is its nominal close, its opening time plus τ, even when its final minute has no trade.

4.2 Outcomes

We measure each join in three ways: as a forecast, as a trading strategy, and as a leak in the signal itself.

Forecast. Every one-minute bar in the regular session that opens at or after 09:30, with its decision no later than 15:45, is an event: about 381,700 per configuration. The position is entered at the open of the next one-minute bar and exited at the close $h \in \{5, 15, 30\}$ minutes later, and never after 15:59. The outcome is $s = D \times (\text{exit} - \text{entry})$ in index points. Events on the same day overlap, so standard errors are clustered by trading day (Petersen, 2009).

Trading strategy. Each minute from 09:31 to 15:59 the strategy holds one contract in the direction of D as known at the previous minute's close, entering at the open. It is flat overnight. Costs are $4 commission per round trip plus one tick of slippage on each side, charged on every change of position: 0.70 points per round trip for NQ and 0.58 points for ES. We report net profit, the

annualised Sharpe ratio of daily net profit, and the maximum drawdown. Because J0 and J2 hold the same position on most minutes, their difference in daily net profit isolates the effect of the join. We call this the *paired leak* and test its mean with an ordinary *t*-statistic on daily values.

Signal leak. For every decision with at least one minute left in the forming bar ($r \geq 1$), we record b_{J1}, F and the two signs, and compute the measured leak L of Section 3.5. The predicted leak $\hat{L}$ evaluates equation (3) separately within groups of decisions, then averages the results weighted by group size. Groups are defined by the minutes remaining, r, and by the tercile of a volatility measure: the mean daily high–low range over the previous 20 trading days, lagged one day. Within each group σ_b and σ_F are root mean squares, which matches the zero-mean assumption of Proposition 1.

4.3 Controls

Lag sweep. We delay the naive join by $k = 0, 1, \ldots, \tau - 1$ minutes, so that the decision at T uses the J0 value from $T - k$. If the naive join's advantage comes from the forming bar's close, it should shrink steadily as k grows. At $k = \tau - 1$ the lagged join is identical to J2, which the code checks.

Random walk. To remove any possible predictability while keeping realistic price moves, we randomly permute all one-minute close-to-close changes across the sample and rebuild prices on the original timestamps, with each bar opening at the previous close. The result has the unconditional return distribution of the original market but no structure in time. Every analysis is repeated on five such series per market. Permutation also removes volatility clustering, a limitation we return to in Section 7.

Truncation test. A causal feature must not change when future data are deleted. For 40 randomly chosen decision times (20 for the moving-average signal), we rebuild every join from data truncated at the decision and compare its sign with the full-sample value. After seeing the results we added the same comparison for the raw slope value; this addition is reported as post hoc.

A second signal and a second market. To test whether the results depend on regression or on NQ, we repeat the analysis for a crossover of simple moving averages of slow-bar closes, with lengths $(f, s) \in \{(5, 20), (10, 40), (20, 80)\}$, and for the regression model on ES. Neither combination had been tested before in this project.

4.4 Pre-registration and inference

The hypothesis, the joins, the configurations, the controls and the decision rules were written down and committed to version control before any result was computed. The main study was registered on 22 September 2026, and an addendum covering Proposition 1, the moving-average signal and ES was registered on 23 September, before those tests were run. Analyses added after results were seen are labelled *post hoc* throughout. Two pre-registered criteria failed, and we report them as failures in Section 5.1.

This study is the third hypothesis tested on these data within one research project, after two earlier pre-registered studies of regression-trend signals that found no edge. To guard against the multiple-testing problem (Harvey, Liu and Zhu, 2016), we keep a single ledger of every test run in the project. Any claim of a *tradeable* edge must reach a t-statistic of $\sqrt{(2 \ln N_{\text{tests}})}$, approximately the largest $|t|$ expected among N_{tests} tests with no true effect. With 936 tests at the time of the main study the bar is 3.70; with the addendum's 99 further tests, 1,035 in total, it is 3.73. The bar does not apply to the leak itself: it measures a coding error, not a search for profitable rules. We report the leak's t-statistics against the same bar anyway, as a conservative reference.

5. Results

We first report the pre-registered verdict, including the two criteria that failed. We then show the effect of the join on trading results, how it behaves as the join is delayed, that it survives on data with no predictability, and why standard checks fail to reveal it. Sections 5.1 to 5.5 use NQ; Sections 5.6 to 5.8 add the formula test, the second signal and ES.

5.1 The pre-registered verdict

Table 3 lists the five criteria registered for the main study. Three pass: no causal cell shows a tradeable edge; the naive join's advantage falls steadily with lag and ends at the causal value; and the naive join fails the truncation test while the causal joins pass it. Two fail. The naive join clears the significance bar with a mean above the cost of a round trip in only 11% of its 27 event-level cells, against the required two-thirds. On the random walks, the share of naive cells with $t \geq 3.70$ is 11–19% per seed, against the same two-thirds requirement. Because a pre-registered claim either meets its own test or does not, the hypothesis *as specified* is not supported.

Table 3. Pre-registered criteria of the main study and their results (NQ, 2020–2023; significance bar 3.70).

#	Criterion	Result	Pass
1	J0: event-level $t \geq 3.70$ and mean ≥ 0.70 pt in at least 2/3 of 27 cells	11% of cells	No
2	No J2 or J3 cell with $t \geq 3.70$ and mean ≥ 0.70 pt	0 cells	Yes
3	J0 mean falls with lag (at most 2 reversals) and ends within 2 standard errors of J2 (15 min, $N = 50$, $h = 15$)	0 reversals; ends at J2	Yes
4	Random walk: J0 $t \geq 3.70$ in at least 2/3 of cells in every seed; no causal cell with $\|t\| \geq 3.70$	J0 11–19% of cells; causal 0	No
5	Truncation test fails for J0 and passes for J1–J3	J0 differs in 0.56% of checks; J1–J3 in 0%	Yes

The failures are informative about *where* the leak lives, not about whether it exists. Criteria 1 and 4 score every minute of the session over a hold of 5 to 30 minutes. As Section 5.5 shows, the naive and causal signals disagree on only 0.8–5.2% of minutes, so most of each cell's variance comes from price moves common to all joins. Even so, the naive join has the largest event-level t-statistic in 26 of the 27 cells (the exception is 5 minutes, $N = 100$, $h = 30$), with a maximum of 7.81. No causal cell

exceeds an absolute t of 2.11 (Appendix Table A3). With 15-minute bars, $N = 50$ and a 15-minute hold, the naive join gains 0.45 points per event ($t = 2.54$) and J2 gains 0.13 ($t = 0.71$). The criteria asked an averaging statistic to detect an effect concentrated in a few minutes. The measures below look at those minutes directly.

5.2 Backtests

Figure 2 and Table 4 show the regime strategy for 15-minute bars and $N = 50$, net of costs, on one contract. With the naive join the strategy earns $392,896 over four years, with a Sharpe ratio of 2.06 and a maximum drawdown of $25,571. The same strategy with the conventional causal join, J2, earns $68,262, with a Sharpe ratio of 0.33 and a drawdown of $52,780. The other causal joins lie close to J2 (J1: $100,354, Sharpe 0.50; J3: $107,508, Sharpe 0.54).

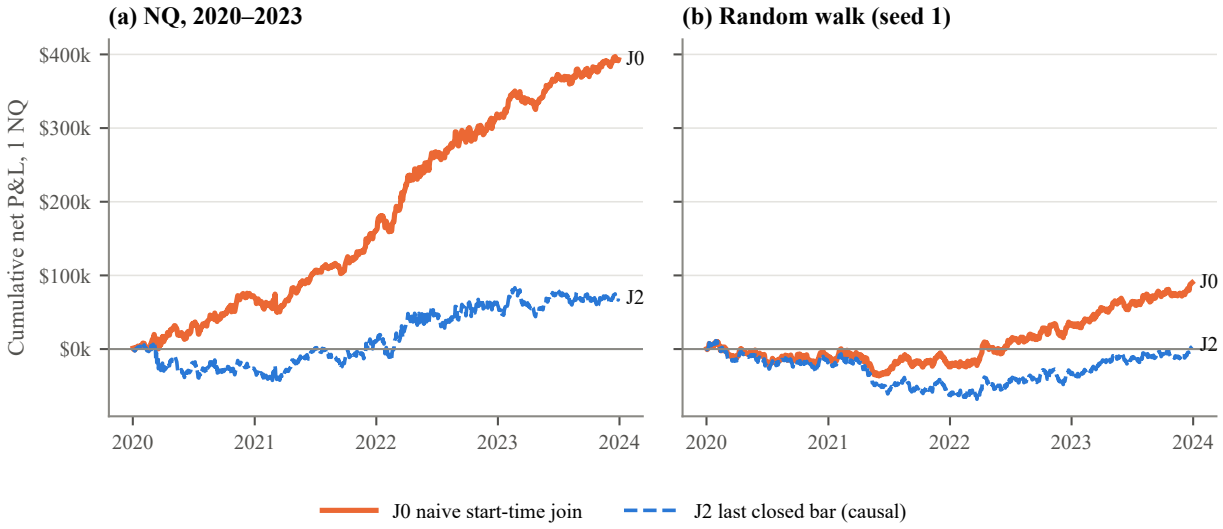


Figure 2. Cumulative net profit of the 15-minute regression model ($N = 50$) on one NQ contract, after 0.70 points per round trip. (a) NQ, 2020–2023. (b) The same model on a random walk built from permuted NQ one-minute returns (seed 1), where no strategy can have an edge. Only the naive join makes money in both panels. J1 and J3 track J2 closely and are omitted for clarity (Table 4).

Table 4. Regime strategy, 15-minute slope, $N = 50$, one NQ contract, net of 0.70 points per round trip, 2020–2023. Random-walk columns give the range over five seeds.

	NQ, 2020–2023					Random walk, 5 seeds	
Join	**Net P&L**	**Sharpe**	**Max DD**	**Daily PF**	**Round trips**	**Net P&L**	**Sharpe**
J0 naive	+$393k	2.06	−$26k	1.41	1,576	+$90k to +$182k	0.69 to 1.40
J1 developing bar	+$100k	0.50	−$49k	1.09	2,219	−$16k to +$75k	−0.12 to 0.57
J2 last closed bar	+$68k	0.33	−$53k	1.06	1,597	−$28k to +$101k	−0.22 to 0.76
J3 rolling 1-minute	+$108k	0.54	−$39k	1.09	1,603	−$17k to +$101k	−0.13 to 0.76

The two equity curves differ in shape as well as level. The naive strategy is profitable in every calendar year, with annual profits between $70k (2020) and $156k (2022). The causal strategy loses $28k in 2020, gains $38k and $51k in 2021 and 2022, and is nearly flat in 2023. A steady, rising

equity curve from a rule with no free parameters is itself a warning sign.

Across all nine combinations of bar length and window (Appendix Table A1), the naive Sharpe ratio ranges from 0.12 to 5.35. The highest value, for 5-minute bars and $N = 20$, corresponds to $958,111 of net profit. The highest Sharpe ratio of any causal join in the grid is 0.88 (J1, 30-minute bars, $N = 50$). Over four years that is a t-statistic of about 1.8, well below the bar of 3.70.

5.3 The lag sweep

If the naive join's advantage comes from knowing the forming bar's close, it should decline as the join is delayed and disappear exactly when the delay reaches a full bar. Figure 3 confirms this. For 15-minute bars, mean continuation over the next 15 minutes falls from 0.454 points with no delay to 0.130 points with a 14-minute delay, the same value as J2. The decline has no reversals for any bar length at the 5- and 15-minute horizons; at 30 minutes there is a single small reversal. It is close to linear in the share of the bar that is leaked, as equation (2) implies: each additional minute of delay removes one minute of the future move F.

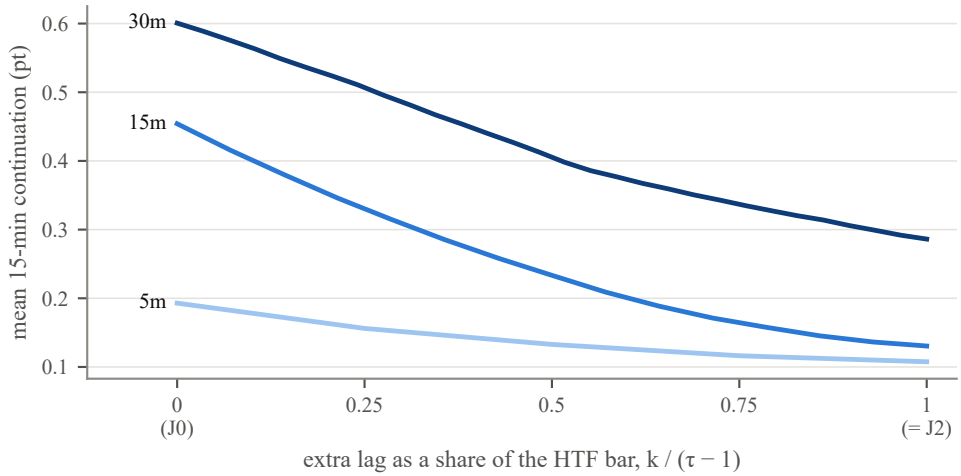


Figure 3. Mean 15-minute continuation per event of the naive join as it is delayed by k minutes, $N = 50$, for 5-, 15- and 30-minute bars (light to dark). The horizontal axis is the delay as a share of the slow bar; at 1 the delayed join is identical to J2.

5.4 The leak on data with no predictability

On the random walks, where no strategy can have an edge, the naive join still profits. For the 15-minute model with $N = 50$ it earns between $90k and $182k net across the five seeds, while J2 earns between −$28k and +$101k (Table 4, Figure 2b).

To measure the leak directly we use the paired leak defined in Section 4.2: the daily difference in net profit between J0 and J2. This comparison was added after the results were seen and is therefore post hoc. On NQ the paired leak is positive in all nine settings, worth between $19k and $227k per year on one contract, with t-statistics from 3.6 to 20.8 (Figure 4, Appendix Table A2). On the random

walks it is of the same order: 37 of the 45 combinations of setting and seed exceed a t of 3.70, with a maximum of 23.0. The only negative value ($t = -0.05$) is for 30-minute bars with $N = 100$, where the leak is smallest. In every bar length, the leak falls as the window lengthens, as equation (2) predicts.

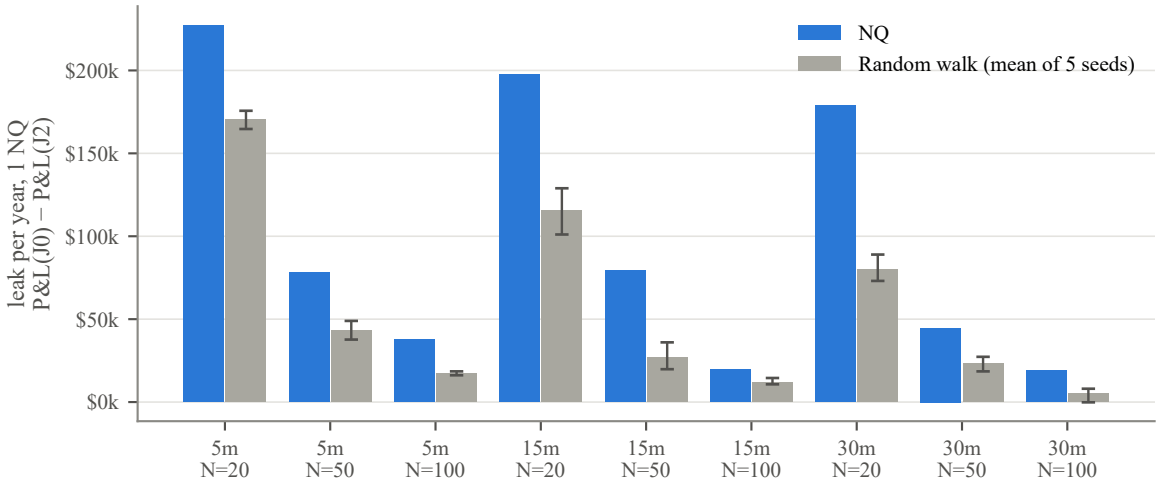


Figure 4. Annual profit of the naive join in excess of the causal join (J0 − J2), one NQ contract, net of costs, by bar length and window. Blue: NQ. Grey: mean over five random walks, with the range across seeds. Paired t-statistics are in Table A2. Post hoc.

The leak is larger on NQ than on the permuted series in every setting, by a factor of 1.3 to 3.8 relative to the random-walk average. We have not tested why. A plausible explanation is that permutation removes volatility clustering: in real markets the minutes in which the trend is ambiguous tend to be volatile, so the future move the naive join sees is larger there. If so, the random walk understates the leak rather than exaggerating it.

5.5 Why standard checks miss it

The leak is large in profit terms but small in most other statistics, which is why it survives the usual checks.

First, it changes the signal rarely. Across the nine settings, J0 and J2 disagree on only 0.8% to 5.2% of regular-session minutes. Any statistic averaged over all minutes, such as a hit rate, an information coefficient or the event-level means of Section 5.1, moves only slightly. A regime strategy, however, trades only when the signal changes sign, and those are precisely the minutes the leak affects (Section 3.4). This is the mechanism behind the observation of Sobreiro et al. (2026) that the bias "is difficult to detect from classification metrics alone".

Second, a truncation test that compares signals rarely fires. In the pre-registered test, the naive join's *sign* changed when future data were deleted in only 2 of 360 checks (0.56%). For the moving-average signal of Section 5.7 the figure was 3 of 180. The causal joins never changed. When we compared the raw slope value instead (post hoc), the naive join changed in 83% of checks. By equation (2), the remaining checks are exactly those where $F = 0$: the last minute of a slow bar, where

nothing is left to leak, or minutes after which the price ends the bar exactly where it was at the decision. The lesson is general: a threshold hides a leak everywhere except where it is crossed, and the crossings are where trades occur. Look-ahead tests should therefore compare raw features, not the signals derived from them.

This mechanism also explains two reports from outside academic finance. A Freqtrade user reported that the framework's built-in look-ahead analysis returned "No Bias" for a strategy that used an unfinished one-hour candle through an as-of join (Freqtrade issue #12507, 2025). Fonseca (2026) shows formally that differential detectors can miss planted leaks.

5.6 Testing Proposition 1

Proposition 1 turns the mechanism of Section 3.4 into a quantitative prediction. We registered a test of it before computing any of the quantities involved. For each setting, the measured leak L must lie between 0.67 and 1.5 times the predicted leak $\hat{L}$ in at least seven of the nine settings, and the rank correlation between L and $\hat{L}$ across settings must be at least 0.80. The rule had to hold on the random walks of both markets, where the proposition's assumptions hold by construction, and on at least one of the two real markets.

The prediction passes on every dataset, in every setting (Figure 5). On the random walks the ratio of measured to predicted leak lies between 0.96 and 1.17 for NQ and between 0.94 and 1.11 for ES, and the rank correlation is 1.00 for both. On the real markets the prediction is less exact but still close: the ratio lies between 0.70 and 1.33 for NQ and between 0.87 and 1.22 for ES, with rank correlations of 0.93 and 0.98. For 15-minute bars and $N = 50$ on NQ, for example, the formula predicts 0.242 points per decision and the measured leak is 0.277.

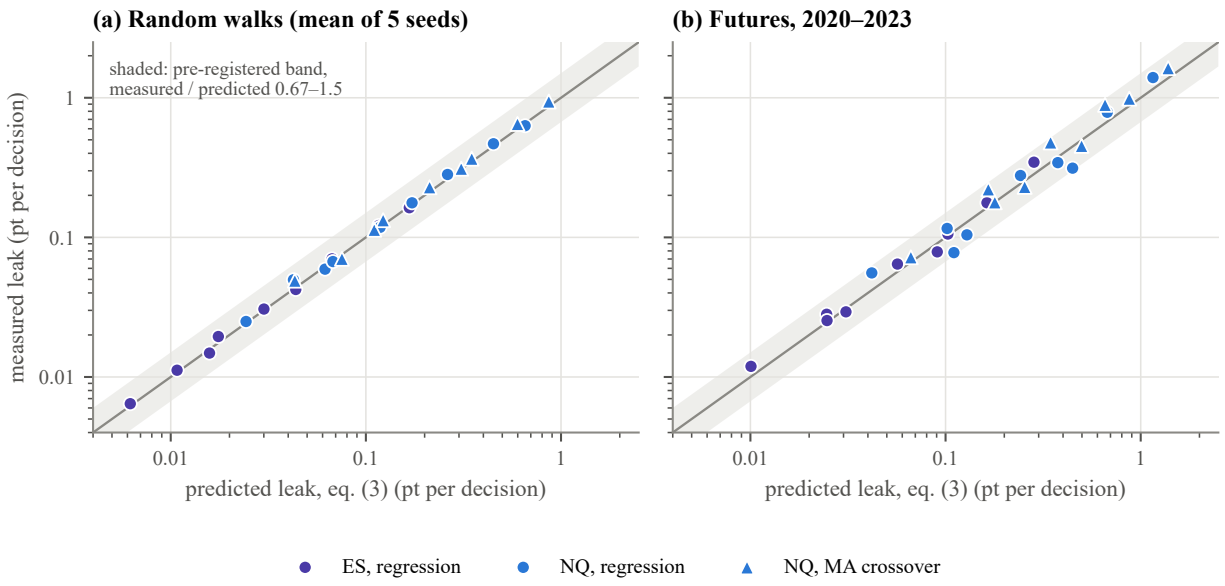


Figure 5. Leak predicted by equation (3) against the measured leak, in index points per decision, for every setting. (a) Random walks, means over five seeds. (b) NQ and ES futures, 2020–2023. Circles: regression; triangles: moving-average crossover. The shaded band is the registered range, 0.67–1.5 times the prediction.

The departures on real data run in both directions and show no simple pattern across settings. That is what one would expect when the proposition's assumptions hold only approximately. Real returns are fat-tailed, their volatility clusters, and the causal slope is not exactly independent of the next few minutes. The formula nevertheless ranks the settings almost perfectly. As a descriptive check that was not part of the registered rule, it also ranks the *profit* leak of Section 5.4 reasonably well: the rank correlation between $\hat{L}$ and the paired leak is 0.62 for NQ and 0.68 for ES. This is notable because the formula describes the signal, not the trading rule built on it.

5.7 Beyond regression

Section 3.4 showed that equation (2) holds for any linear filter of closes, with the filter's weight on the newest close in place of w_N. To test this we repeated the analysis for a crossover of simple moving averages of slow-bar closes, $D = \text{sign}(\text{SMA}_f - \text{SMA}_s)$. This signal is a staple of trend-following practice and one of the two filter families Levine and Pedersen (2016) study. Its weight on the newest close, $1/f - 1/s$, is 0.150, 0.075 and 0.0375 for $(f, s) = (5, 20)$, (10, 40) and (20, 80), an order of magnitude larger than the regression weights for comparable windows.

All three registered conditions hold (Table 5). The paired leak exceeds the significance bar of 3.73 in all nine settings on NQ, with t-statistics from 4.86 to 24.65 and annual values from \$39k to \$301k on one contract. With 5-minute bars and (5, 20) the naive join produces a Sharpe ratio of 7.58. No causal variant reaches a t-statistic above 2.02. On the random walks, 42 of the 45 combinations of setting and seed exceed the bar. The formula predicts the crossover's leak as well as it predicts the slope's,

with ratios between 0.93 and 1.13 on the random walks and between 0.90 and 1.39 on NQ (rank correlation 0.97). In the truncation test the causal joins again never changed, while the naive join's sign changed in only 3 of 180 checks.

Table 5. Moving-average crossover on NQ, 2020–2023: naive and causal Sharpe ratios, paired leak and formula ratio by setting.

Bars	(*f*, *s*)	J0 Sharpe	J2 Sharpe	Max causal *t*	Leak per year	Paired *t*	$L / \hat{L}$
5 min	(5, 20)	7.58	0.62	1.78	+$301k	24.6	0.91
5 min	(10, 40)	2.62	0.09	0.28	+$118k	13.4	1.00
5 min	(20, 80)	1.47	0.43	1.23	+$49k	8.4	1.09
15 min	(5, 20)	5.73	0.47	0.95	+$232k	16.6	1.13
15 min	(10, 40)	2.61	0.15	0.91	+$114k	11.3	1.39
15 min	(20, 80)	2.16	0.85	2.02	+$59k	7.9	1.33
30 min	(5, 20)	5.34	0.51	1.04	+$206k	14.6	1.18
30 min	(10, 40)	3.20	0.97	1.97	+$95k	9.1	1.36
30 min	(20, 80)	1.16	0.33	1.14	+$39k	4.9	0.90

The alignment leak is therefore a property of how the signal is joined, not of regression. Regression is simply the case whose weights are easiest to write down.

5.8 A second market

On ES, which had not been used for this signal, the results replicate (Table 6). The naive join produces Sharpe ratios of up to 4.42 (15-minute bars, $N = 20$). The paired leak is worth between $9k and $140k per year on one ES contract and exceeds the bar in eight of nine settings. The ninth, 30-minute bars with $N = 100$, reaches a t-statistic of 3.70 against a bar of 3.73, so we count it as a miss. The registered rule required seven. No causal variant has a t-statistic above 0.82, and several are negative, down to −2.23. On the ES random walks, 35 of 45 seed-setting combinations exceed the bar, and the formula ratio lies between 0.94 and 1.11 there and between 0.87 and 1.22 on ES itself.

Table 6. Regression model on ES, 2020–2023, one contract, net of 0.58 points per round trip: naive and causal results and paired leak by setting.

Bars	*N*	J0 Sharpe	J2 Sharpe	Max causal *t*	Leak per year	Paired *t*	$L / \hat{L}$
5 min	20	4.26	−0.94	−1.63	+$140k	20.6	0.87
5 min	50	1.21	−0.19	−0.11	+$39k	9.2	1.03
5 min	100	0.14	−0.47	−0.80	+$19k	5.8	1.18
15 min	20	4.42	0.23	0.46	+$110k	15.5	1.09
15 min	50	1.66	0.02	0.06	+$44k	8.3	1.14
15 min	100	0.91	0.38	0.76	+$15k	3.8	1.14
30 min	20	4.14	−0.23	−0.15	+$116k	12.1	1.22
30 min	50	1.43	0.41	0.82	+$28k	5.2	1.03
30 min	100	−0.48	−0.78	−1.47	+$9k	3.7	0.95

5.9 What remains once the join is causal

With the leak removed, what does a slow regression slope tell a one-minute strategy? Rewriting equation (1) in terms of returns shows that the slope is a weighted sum of the $N - 1$ returns in its window, with weights proportional to $j(N - j)$ for the j-th return. This is a special case of the trend signatures of Levine and Pedersen (2016). The weight peaks in the middle of the window and is close to zero for the newest return (Figure 6). For $N = 50$ the newest return carries 0.24% of the total weight, against 3.0% at the peak. With 15-minute bars the centre of the weight lies about six hours before the decision, so at a morning entry the signal mostly summarises the previous night's trading. A one-minute execution layer can change *when* a trade is placed, but not what the signal knows.

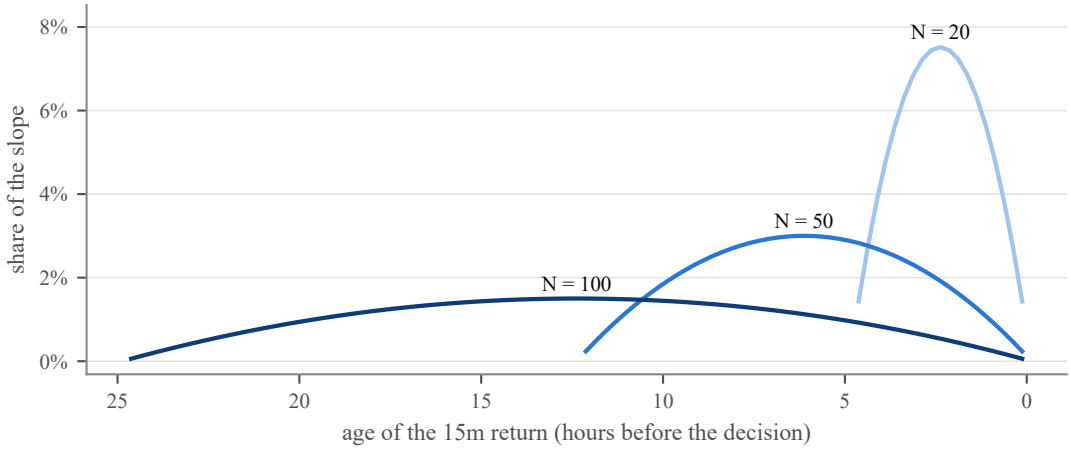


Figure 6. Share of a causal 15-minute least-squares slope contributed by each past 15-minute return, for three window lengths, plotted against the return's age. The most recent return contributes almost nothing.

In our data that knowledge is worth little. Across both markets and both signals, no causal variant reaches the significance bar: the highest t-statistics are about 1.8 for the NQ regression model, 0.82 for ES and 2.02 for the crossover. This is consistent with two other results. Kurth, Eisler, Rej and Bouchaud (2026) find that short-term trend following has earned little since about 2009 on contracts whose tick is small relative to their volatility, and equity-index futures fall in that group; NQ's tick is about 0.13% of its daily standard deviation over our sample. And a pre-registered event study on the same NQ data found only a faint drift in the direction of a regression trend, a median of 0.01–0.04 average true ranges. The one continuation effect that stood out lasted about five minutes and was worth roughly one point before a round-trip cost of 0.70.

A causal multi-timeframe regression model can be backtested honestly. On these data, the honest backtest shows no tradeable edge. Almost everything the naive backtest adds is borrowed from the future.

6. Practical recommendations

Arnott, Harvey and Markowitz (2019) propose a general protocol for backtesting. The results above suggest adding the following checks to it for any strategy that combines two timeframes. None of them is costly. The first two prevent the leak, and in our data each of checks 3 to 6 detected it.

1. **Stamp every series by its availability time.** A bar should carry the time at which its values become known, which is its close. Join the slow series onto the fast one with a backward as-of join on that time, compared with the fast bar's decision time (join J2). Appendix B gives the code in pandas and Pine Script.
2. **Know the default of every tool in the chain.** pandas labels intraday bins by their opening edge; TradingView stamps bars by their opening time, and its documented safe form for higher-timeframe requests is `request.security(…, expr[1], lookahead = barmerge.lookahead_on)`; Freqtrade provides a helper for the same purpose; TradeStation and MultiCharts stamp bars by their close (Table 1). Moving data or code between conventions is the most likely moment for the leak to appear.
3. **Run truncation tests on raw features, not on signals.** Delete all data after a decision and recompute. Compare the *values* of every feature, at decision times chosen inside slow bars as well as at their ends. A test on thresholded signals missed the leak in more than 99% of checks (Section 5.5).
4. **Run a lag sweep.** Delay every slow feature by 1, 2, …, $\tau - 1$ fast bars. A causal edge should barely change; a leaked edge declines steadily towards the causal level as the delay grows (Figure 3).
5. **Compare against the causal version on the same days.** The paired difference in daily profit isolates the effect of the join. It detects the leak with *t*-statistics of 3.6 to 24.6 even when the two versions agree on more than 94% of minutes.
6. **Run the strategy on a permuted series.** Any profit on data with no temporal structure is an artefact of construction.
7. **Budget the leak before trusting the backtest.** For linear signals, equation (3) gives the expected leak from quantities that can be estimated before any backtest is run: the filter's weight on the newest close and the typical sizes of the signal and of the remaining move. If a backtest's edge is of the same order as the predicted leak, the join is the first suspect.
8. **Treat certain results as symptoms.** A high Sharpe ratio from a rule with no free parameters, profits in every calendar year, and an edge that grows as the window shortens are all characteristic of the alignment leak. In our data all three appeared together.

7. Limitations

The study covers one four-year period, 2020–2023, and two closely related markets, the Nasdaq-100 and S&P 500 index futures. The mechanism does not depend on the period, since equation (2) is an identity and the leak appears on random walks. Our conclusion that the *causal* signals have no edge, however, is specific to these markets and years. We deliberately left the sealed 2024–2026 data unused.

We study two linear signals. Equation (2) and Proposition 1 apply to every linear filter of closes, but not directly to non-linear features such as oscillators, rolling highs and lows, or pivot levels. The same alignment error affects them, as Sobreiro et al. (2026) show for oscillators and bands, but their leak has no closed form.

Proposition 1 assumes independent, normally distributed quantities. On real data its predictions are accurate to within about 30–40% rather than exactly. Our random-walk control permutes returns and so removes volatility clustering along with any predictability. The leak is larger on real data than on the permuted series, and we have not tested whether volatility clustering is the reason; a block bootstrap would be the natural next step.

The trading model is deliberately simple: one contract, a position in the direction of the signal every minute of the regular session, and a fixed cost per round trip without market impact. Real strategies filter their trades, which changes how many sign changes they act on and therefore how much of the leak they collect. Our data have one-minute resolution, so leaks shorter than a minute, such as fills at prices inside the decision bar, are outside our scope.

Finally, two of the five pre-registered criteria of the main study failed. The evidence for the size of the leak rests on the pre-registered backtests, the lag sweep and the pre-registered addendum, together with post-hoc paired tests that we have labelled as such.

8. Conclusion

A multi-timeframe model reads its signal from slow bars and trades on fast ones, and the step that joins the two decides what each trade can know. The default way of writing that join, keyed on the time a slow bar opens, gives every fast decision the final value of a slow bar that is still forming. For a linear-regression trend signal the consequence can be written down exactly: the naive slope equals the slope a live trader sees plus a fixed multiple of the price move still to come.

That single term is enough to transform a backtest. On NQ futures a regression model with no demonstrable edge earns a Sharpe ratio of 2.06 with the naive join and 0.33 with a causal one, and the naive version profits even on a random walk. The leak's size is predictable. A closed-form expression, registered before testing, matched the measured leak within the registered band on two

futures markets, on their random walks and for a second trend signal. The leak is also hard to see: it changes the signal on only a few percent of minutes, and a truncation test on the signal misses it almost every time, because those minutes are exactly where a sign-following strategy trades.

A multi-timeframe model can be backtested honestly, but only if every bar is stamped by the time its values become known. Once that is done, the models studied here have no tradeable edge. What the naive backtest showed as skill was information from the future.

References

Arnott, Robert D., Campbell R. Harvey, and Harry Markowitz, 2019, A backtesting protocol in the era of machine learning, *Journal of Financial Data Science* 1(1), 64–74.

Bailey, David H., Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu, 2014, Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance, *Notices of the American Mathematical Society* 61(5), 458–471.

Bailey, David H., Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu, 2017, The probability of backtest overfitting, *Journal of Computational Finance* 20(4), 39–70.

Bailey, David H., and Marcos López de Prado, 2014, The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality, *Journal of Portfolio Management* 40(5), 94–107.

Banz, Rolf W., and William J. Breen, 1986, Sample-dependent results using accounting and market data: Some evidence, *Journal of Finance* 41(4), 779–793.

Brock, William, Josef Lakonishok, and Blake LeBaron, 1992, Simple technical trading rules and the stochastic properties of stock returns, *Journal of Finance* 47(5), 1731–1764.

Bruder, Benjamin, Tung-Lam Dao, Jean-Charles Richard, and Thierry Roncalli, 2011, Trend filtering methods for momentum strategies, Lyxor Asset Management White Paper, available at SSRN 2289097.

Fonseca, Xavier, 2026, Look-ahead-freedom as temporal non-interference: A verifiable correctness property for backtesting and agentic trading pipelines, arXiv:2607.04958.

Gao, Lei, Yufeng Han, Sophia Zhengzi Li, and Guofu Zhou, 2018, Market intraday momentum, *Journal of Financial Economics* 129(2), 394–414.

Harvey, Campbell R., 2017, Presidential address: The scientific outlook in financial economics, *Journal of Finance* 72(4), 1399–1440.

Harvey, Campbell R., and Yan Liu, 2015, Backtesting, *Journal of Portfolio Management* 42(1), 13–28.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, …and the cross-section of expected returns, *Review of Financial Studies* 29(1), 5–68.

Kaufman, Shachar, Saharon Rosset, Claudia Perlich, and Ori Stitelman, 2012, Leakage in data mining: Formulation, detection, and avoidance, *ACM Transactions on Knowledge Discovery from Data* 6(4), Article 15, 1–21.

Kurth, Jutta G., Zoltan Eisler, Adam Rej, and Jean-Philippe Bouchaud, 2026, Is trend still your friend? A microstructural account of the demise of short-term trend-following, arXiv:2607.01550.

Levine, Ari, and Lasse Heje Pedersen, 2016, Which trend is your friend? *Financial Analysts Journal* 72(3), 51–66.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3(3), 431–467.

López de Prado, Marcos, 2018, *Advances in Financial Machine Learning* (Wiley, Hoboken, NJ).

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104(2), 228–250.

Park, Cheol-Ho, and Scott H. Irwin, 2007, What do we know about the profitability of technical analysis? *Journal of Economic Surveys* 21(4), 786–826.

Petersen, Mitchell A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22(1), 435–480.

Sobreiro, Pedro, Domingos Martinho, Rui Martins, and Ricardo Vardasca, 2026, Multi-timeframe feature engineering for Bitcoin market prediction: A price-level-agnostic machine learning approach, *Forecasting* 8(3), 40.

Sullivan, Ryan, Allan Timmermann, and Halbert White, 1999, Data-snooping, technical trading rule performance, and the bootstrap, *Journal of Finance* 54(5), 1647–1691.

White, Halbert, 2000, A reality check for data snooping, *Econometrica* 68(5), 1097–1126.

Software documentation

Freqtrade, 2026, Strategy customization, Freqtrade documentation (stable), https://www.freqtrade.io/en/stable/strategy-customization/, accessed 22 September 2026.

Freqtrade issue #12507, 2025, Lookahead bias with merge_asof for HigherTimeFrame calculation not detected by Lookahead-Analysis, GitHub, https://github.com/freqtrade/freqtrade/issues/12507, opened 11 November 2025.

pandas development team, 2026, pandas.DataFrame.resample, pandas API reference (version 3.0), https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.resample.html, accessed 22 September 2026.

TradeStation, 2026, Time (reserved word), EasyLanguage reference, https://help.tradestation.com/10_00/eng/tsdevhelp/elword/word/time_reserved_word_.htm, accessed 22 September 2026.

TradingView, 2026a, Repainting, Pine Script v6 user manual, https://www.tradingview.com/pine-script-docs/concepts/repainting/, accessed 22 September 2026.

TradingView, 2026b, Time, Pine Script v6 user manual, https://www.tradingview.com/pine-script-docs/concepts/time/, accessed 22 September 2026.

Appendix A. Full results

All tables are generated directly from the result files when the paper is built, so that no number is typed by hand.

Table A1. Regime strategy across the full grid, NQ, 2020–2023: net profit and annualised Sharpe ratio (SR), one contract, net of 0.70 points per round trip. J3 with (15 min, $N = 100$) and (30 min, $N = 50$) is the same 1,500-minute regression, so those two entries coincide, which checks the implementation.

		J0		J1		J2		J3	
Bars	***N***	**Net**	**SR**	**Net**	**SR**	**Net**	**SR**	**Net**	**SR**
5 min	20	+$958k	5.35	+$31k	0.16	+$27k	0.14	+$55k	0.28
5 min	50	+$336k	1.77	+$21k	0.11	+$17k	0.08	+$28k	0.14
5 min	100	+$158k	0.77	+$13k	0.06	+$4k	0.02	+$14k	0.07
15 min	20	+$847k	4.55	+$12k	0.06	+$38k	0.19	+$31k	0.15
15 min	50	+$393k	2.06	+$100k	0.50	+$68k	0.33	+$108k	0.54
15 min	100	+$240k	1.24	+$156k	0.80	+$159k	0.80	+$153k	0.78
30 min	20	+$803k	4.29	+$24k	0.11	+$70k	0.34	+$118k	0.59
30 min	50	+$350k	1.82	+$169k	0.88	+$167k	0.85	+$153k	0.78
30 min	100	+$23k	0.12	−$36k	−0.18	−$55k	−0.28	−$34k	−0.17

Table A2. Paired leak, P&L(J0) − P&L(J2), per year on one NQ contract, with the share of regular-session minutes on which J0 and J2 disagree (post hoc).

			NQ		Random walk, 5 seeds	
Bars	***N***	**J0 ≠ J2**	**Leak per year**	**Paired *t***	**Mean leak per year**	**Paired *t***
5 min	20	4.0%	+$227k	20.8	+$170k	20.7 to 23.0
5 min	50	1.6%	+$78k	11.6	+$43k	7.9 to 10.2
5 min	100	0.8%	+$38k	7.0	+$17k	4.6 to 6.4
15 min	20	4.6%	+$197k	16.1	+$115k	13.3 to 16.7
15 min	50	2.1%	+$79k	9.9	+$27k	4.7 to 7.7
15 min	100	1.4%	+$20k	3.6	+$12k	3.3 to 5.1
30 min	20	5.2%	+$179k	13.6	+$80k	8.2 to 11.3
30 min	50	3.0%	+$45k	5.1	+$23k	4.1 to 6.1
30 min	100	1.0%	+$19k	4.5	+$5k	−0.1 to 2.7

Table A3. Event-level mean continuation (points) and day-clustered *t*-statistics for every cell, NQ, 2020–2023.

			J0		J1		J2		J3	
Bars	***N***	***h***	**pt**	***t***	**pt**	***t***	**pt**	***t***	**pt**	***t***
5 min	20	5	0.27	4.79	−0.01	−0.21	−0.01	−0.17	0.00	−0.08
5 min	20	15	0.41	2.46	0.13	0.76	0.16	0.94	0.13	0.79
5 min	20	30	0.75	2.39	0.45	1.43	0.47	1.48	0.42	1.33
5 min	50	5	0.09	1.44	−0.01	−0.09	0.01	0.22	0.00	0.06
5 min	50	15	0.19	1.10	0.10	0.55	0.11	0.61	0.09	0.53
5 min	50	30	0.33	0.95	0.23	0.66	0.22	0.63	0.21	0.60
5 min	100	5	0.03	0.51	−0.01	−0.17	−0.01	−0.22	0.00	−0.04
5 min	100	15	0.10	0.53	0.06	0.30	0.05	0.28	0.09	0.52
5 min	100	30	0.26	0.72	0.22	0.62	0.24	0.67	0.33	0.94
15 min	20	5	0.42	7.39	−0.01	−0.16	−0.01	−0.13	0.00	0.06
15 min	20	15	0.81	4.66	0.06	0.32	0.03	0.15	0.09	0.47
15 min	20	30	0.88	2.56	0.18	0.53	0.08	0.24	0.23	0.67
15 min	50	5	0.20	3.37	0.04	0.74	0.03	0.54	0.05	0.80
15 min	50	15	0.45	2.54	0.19	1.02	0.13	0.71	0.21	1.16
15 min	50	30	0.65	1.82	0.38	1.07	0.36	1.01	0.53	1.50
15 min	100	5	0.16	2.74	0.12	1.97	0.11	1.84	0.12	1.94
15 min	100	15	0.39	2.24	0.31	1.79	0.29	1.65	0.30	1.68
15 min	100	30	0.64	1.85	0.57	1.66	0.56	1.60	0.50	1.45
30 min	20	5	0.45	7.81	0.00	0.02	0.03	0.55	0.04	0.72
30 min	20	15	1.16	6.60	0.09	0.49	0.18	1.00	0.20	1.11
30 min	20	30	1.67	4.83	0.31	0.86	0.47	1.32	0.59	1.65
30 min	50	5	0.22	3.77	0.12	2.11	0.11	1.91	0.12	1.94
30 min	50	15	0.60	3.47	0.36	2.09	0.29	1.62	0.30	1.68
30 min	50	30	0.98	2.89	0.67	1.97	0.54	1.55	0.50	1.45
30 min	100	5	0.01	0.10	−0.03	−0.45	−0.04	−0.70	−0.03	−0.57
30 min	100	15	−0.05	−0.26	−0.13	−0.72	−0.17	−0.95	−0.14	−0.77
30 min	100	30	−0.14	−0.41	−0.25	−0.69	−0.32	−0.89	−0.25	−0.69

Table A4. Proposition 1 by setting: ratio of measured to predicted leak, $L / \hat{L}$. Random-walk values are means over five seeds. The registered band is 0.67–1.5.

Setting	NQ	ES	RW-NQ	RW-ES
Regression				
5 min, $N = 20$	0.91	0.87	1.08	1.05
5 min, $N = 50$	1.14	1.03	0.99	1.11
5 min, $N = 100$	1.33	1.18	1.03	1.04
15 min, $N = 20$	1.17	1.09	1.04	1.04
15 min, $N = 50$	1.15	1.14	1.00	1.02
15 min, $N = 100$	0.70	1.14	1.17	1.04
30 min, $N = 20$	1.21	1.22	0.96	0.98
30 min, $N = 50$	0.70	1.03	1.03	0.97
30 min, $N = 100$	0.81	0.95	0.96	0.94
Moving-average crossover				
5 min, 5/20	0.91	–	1.05	–
5 min, 10/40	1.00	–	1.08	–
5 min, 20/80	1.09	–	1.13	–
15 min, 5/20	1.13	–	1.09	–
15 min, 10/40	1.39	–	1.08	–
15 min, 20/80	1.33	–	0.93	–
30 min, 5/20	1.18	–	1.09	–
30 min, 10/40	1.36	–	1.01	–
30 min, 20/80	0.90	–	1.03	–

Table A5. Moving-average crossover on NQ, full grid: net profit and Sharpe ratio of every join, one contract, net of 0.70 points per round trip.

		J0		J1		J2		J3	
Bars	**(*f*/*s*)**	**Net**	**SR**	**Net**	**SR**	**Net**	**SR**	**Net**	**SR**
5 min	5/20	+$1,351k	7.58	+$83k	0.43	+$120k	0.62	+$169k	0.88
5 min	10/40	+$503k	2.62	+$2k	0.01	+$18k	0.09	+$27k	0.14
5 min	20/80	+$288k	1.47	+$89k	0.44	+$86k	0.43	+$120k	0.61
15 min	5/20	+$1,043k	5.73	−$10k	−0.05	+$93k	0.47	+$47k	0.24
15 min	10/40	+$498k	2.61	+$25k	0.12	+$31k	0.15	+$89k	0.45
15 min	20/80	+$415k	2.16	+$198k	1.00	+$173k	0.85	+$182k	0.89
30 min	5/20	+$950k	5.34	+$64k	0.31	+$105k	0.51	+$89k	0.45
30 min	10/40	+$585k	3.20	+$114k	0.58	+$194k	0.97	+$182k	0.89
30 min	20/80	+$225k	1.16	+$110k	0.57	+$65k	0.33	+$43k	0.22

Table A6. Regression model on ES, full grid: net profit and Sharpe ratio of every join, one contract, net of 0.58 points per round trip.

		J0		J1		J2		J3	
Bars	***N***	**Net**	**SR**	**Net**	**SR**	**Net**	**SR**	**Net**	**SR**
5 min	20	+$460k	4.26	−$129k	−1.10	−$115k	−0.94	−$97k	−0.80
5 min	50	+$136k	1.21	−$51k	−0.43	−$23k	−0.19	−$6k	−0.05
5 min	100	+$17k	0.14	−$66k	−0.53	−$59k	−0.47	−$48k	−0.39
15 min	20	+$476k	4.42	−$41k	−0.34	+$27k	0.23	+$2k	0.02
15 min	50	+$182k	1.66	+$3k	0.03	+$3k	0.02	−$3k	−0.02
15 min	100	+$103k	0.91	+$34k	0.30	+$43k	0.38	+$41k	0.35
30 min	20	+$448k	4.14	−$52k	−0.41	−$29k	−0.23	−$9k	−0.07
30 min	50	+$159k	1.43	+$3k	0.03	+$46k	0.41	+$41k	0.35
30 min	100	−$56k	−0.48	−$96k	−0.81	−$92k	−0.78	−$86k	−0.72

Appendix B. The join in code

The naive and causal joins differ by a single line. The examples below assume one-minute bars indexed by their opening time, as most data vendors deliver them.

pandas.

```
# Naive (J0): 15-minute bins are labelled by their OPENING time, so the bin labelled
# 10:00 contains the close at 10:14 and is matched to every 1-minute bar from 10:00.
htf = m1.resample("15min").agg({"close": "last"})
htf["slope"] = rolling_ols_slope(htf["close"], n=50)        # any rolling OLS-slope helper
naive = pd.merge_asof(m1, htf[["slope"]], left_index=True, right_index=True)

# Causal (J2): stamp each 15-minute bin by its CLOSING time and match it to each
# 1-minute bar's decision time (its own close).
htf.index = htf.index + pd.Timedelta("15min")
m1["decision_time"] = m1.index + pd.Timedelta("1min")
causal = pd.merge_asof(
    m1.reset_index(), htf[["slope"]].rename_axis("htf_close").reset_index(),
    left_on="decision_time", right_on="htf_close", direction="backward")
```

TradingView Pine Script. The documented form that returns only confirmed higher-timeframe values offsets the requested expression by one bar when look-ahead is enabled (TradingView, 2026a):

```
slopeNaive  = request.security(syminfo.tickerid, "15", slope, lookahead = barmerge.lookahead_on)
// J0 on history
slopeCausal = request.security(syminfo.tickerid, "15", slope[1], lookahead =
barmerge.lookahead_on)  // last closed bar
```

TradeStation EasyLanguage and MultiCharts. Bars are stamped by their closing time (TradeStation, 2026), so an as-of join on these stamps is causal. The risk arises when close-stamped data are combined with open-stamped data, or when code is ported from one convention to the other without shifting the stamps.

Appendix C. Proofs

Weights of the least-squares slope. For closes $C_1, \ldots, C_N$ at positions $k = 1, \ldots, N$, the least-squares slope is $b = \Sigma_k\, w_k\, C_k$ with $w_k = (k - (N + 1)/2) \,/\, S_{xx}$ and $S_{xx} = N(N^2 - 1)/12$. The weight on the newest close is

$$w_N = \frac{(N-1)/2}{N(N^2-1)/12} = \frac{6}{N(N+1)}$$

Because $\Sigma_k\, w_k = 0$, the slope can also be written in terms of the returns $r_j = C_{j+1} - C_j$, $j = 1, \ldots, N - 1$:

$$b = \sum_{j=1}^{N-1} r_j \sum_{k>j} w_k, \qquad \sum_{k=j+1}^{N} \left(k - \frac{N+1}{2}\right) = \frac{j\,(N-j)}{2}$$

so the weight on the j-th return is $6\,j(N - j) \,/\, (N(N^2 - 1))$. It is largest in the middle of the window and equals $6/(N(N + 1))$ for the newest return (Figure 6).

Equation (2) and its generalisation. Let $S = \Sigma_k\, v_k\, C_k$ be any linear filter of the last K slow-bar closes, with v_K the weight on the newest. J0 and J1 evaluate S on the same window and differ only in the value entered for the forming bar: C_{final} for J0 and c_T for J1. Hence $S_{J0} - S_{J1} = v_K\,(C_{final} - c_T) = v_K\, F$. For the least-squares slope, $v_K = w_N$, which gives equation (2). For a crossover of simple moving averages of lengths $f < s$, $v_K = 1/f - 1/s$.

Proposition 1. Let $X = b_{J1}$ and $Y = F$ be independent with $X \sim N(0, \sigma_b^2)$ and $Y \sim N(0, \sigma_F^2)$, and let $Z = X + wY = b_{J0}$. Then

$$\mathrm{E}\left[\,(\operatorname{sign} Z - \operatorname{sign} X)\, Y\right] = \sqrt{2/\pi}\;\frac{w\,\sigma_F^2}{\sqrt{\sigma_b^2 + w^2\,\sigma_F^2}}$$

Proof. Because X and Y are independent and $\mathrm{E}[Y] = 0$, $\mathrm{E}[\operatorname{sign}(X)\, Y] = \mathrm{E}[\operatorname{sign} X]\,\mathrm{E}[Y] = 0$. The pair (Y, Z) is jointly normal with mean zero, $\mathrm{Var}(Z) = \sigma_b^2 + w^2\sigma_F^2$ and $\mathrm{Cov}(Y, Z) = w\sigma_F^2$. Hence $\mathrm{E}[Y \mid Z] = w\sigma_F^2\, Z \,/\, (\sigma_b^2 + w^2\sigma_F^2)$, and

$$\mathrm{E}\left[\,Y \operatorname{sign} Z\right] = \mathrm{E}\left[\,\mathrm{E}[\,Y \mid Z\,] \operatorname{sign} Z\right] = \frac{w\,\sigma_F^2\,\mathrm{E}|Z|}{\sigma_b^2 + w^2\,\sigma_F^2}$$

For a normal variable with mean zero, $\mathrm{E}|Z| = \sqrt{(2/\pi)} \cdot \sqrt{\mathrm{Var}(Z)}$. Substituting gives the result. ■

Grouped prediction. In Section 4.2 the formula is evaluated within groups g (minutes remaining × volatility tercile) and averaged with weights n_g/n. If the assumptions hold within each group, the law of total expectation gives $L = \Sigma_g\, (n_g/n)\, L_g$, so the grouped average is the correct prediction even

when σ_b and σ_F differ across groups.

Appendix D. Registration record

Each registration was committed to version control before the corresponding results were computed. The repository and its history are available from the author on request.

Date	Stage	Commit
22 Sep 2026	Main study registered (joins, 108 cells, controls, decision rule)	`5db51d8`
22 Sep 2026	Main study run and reported, including the two failed criteria	`4404054`
23 Sep 2026	Addendum registered (Proposition 1, moving-average crossover, ES)	`5b7da9a`
23 Sep 2026	Addendum run and reported	`cb9d9b8`

The ledger of every test run in the project (1,035 in total) and the corresponding significance bars are part of the same record.