Point-in-Time Backtesting of Momentum-Trend Equity Strategies: A Formal Bias Taxonomy, ATR Trailing Stop Analysis, and Investor Experience Metrics

Xavier Fonseca *

Posted Date: 4 June 2026

doi: 10.20944/preprints202606.0436.v1

Keywords: systematic trading; backtesting bias; look-ahead bias; point-in-time correctness; ATR trailing stop; momentum investing; backtest overfitting probability; investor experience metrics

 doi:10.20944/preprints202606.0436.v1

Article

Point-in-Time Backtesting of Momentum-Trend Equity Strategies: A Formal Bias Taxonomy, ATR Trailing Stop Analysis, and Investor Experience Metrics

Xavier Fonseca

Academy for AI, Games and Media, Breda University of Applied Sciences; santosfonseca.f@buas.nl

Highlights

What are the main findings?

- Three-class taxonomy of look-ahead bias as natural-filtration violations in equity backtests.
- Lemma 1: the ATR trailing stop satisfies a monotonic non-decreasing ratcheting property, with closed-form per-trade loss bound.
- Theorem 2: Sharpe ratio and maximum consecutive negative-year run (MCNYR) are independent.
- Across 14 ATR configurations on 18 years of NASDAQ-100 PIT data, MCNYR is invariantly 1.
- Stop-multiplier-insensitive plateau: 12 configs at annual Sharpe 0.805 ± 0.007, CAGR within 0.20 pp.

What are the implications of the main findings?

- Natural-filtration compliance distinguishes reproducible backtests from those inflated by leakage.
- MCNYR, time-under-water and rolling Calmar are mathematically independent of Sharpe.
- A wide performance plateau supports region-based parameter defence against PBO = 0.9329 overfitting.

Abstract

Systematic trend-following strategies applied to equity markets are widely studied, yet most reported performance statistics are non-reproducible in live trading. This paper makes three contributions. First, we introduce a formal taxonomy of look-ahead bias organised around point-in-time correctness: a strategy is point-in-time correct if, for every decision time t, its information set lies in the natural filtration $\mathcal{F}_t$. Three bias classes — universe-membership contamination, price-data forward leakage, and stop-exit sequencing violations — are characterised as filtration breaches. Second, we formalise the Average True Range (ATR) trailing stop as a stochastic recurrence and codify its monotonic non-decreasing ratcheting property (Lemma 1), providing a structural per-trade loss bound. Third, we exhibit a closed-form construction (Theorem 2) of two return sequences with identical Sharpe ratio but arbitrarily divergent maximum consecutive negative-year run, establishing investor-experience metrics as independent optimisation objectives. We complement these contributions with an 18-year empirical study (2008–2025) on the NASDAQ-100 with reconstructed point-in-time index constituency (Class I compliant) and an acknowledged residual Class II exposure, applying combinatorially symmetric cross-validation (CSCV) to a 14-configuration ATR-multiplier grid. The grid exhibits a stop-multiplier-insensitive CAGR-flat region across $k \in [3.5,7.0]$ (CAGR 10.28–10.39%, net of Dutch progressive tax) and a uniform maximum consecutive negative-year run of 1 across all 14 configurations. The correlation-matrix eigenvalue spectrum of the grid is dominated by a single mode ($\lambda_1 = 13.91$ of 14), yielding effective independent-test count $M_{\text{eff}} = 1.09$ — a near-

 doi:10.20944/preprints202606.0436.v1

degeneracy that persists in a parallel grid with the regime classifier disabled (Appendix B), establishing the ATR multiplier as a structurally near-redundant parameter for this strategy class. The associated PBO = 0.933 is the expected consequence of this near-degeneracy under CSCV's maximum-selection rule. The plateau-level performance survives Bonferroni correction at both $M = 14$ and M_{eff}. The combined evidence supports a region-based interpretation of robust strategy parameters in preference to single-point optimisation.

Keywords: systematic trading; backtesting bias; look-ahead bias; point-in-time correctness; ATR trailing stop; momentum investing; backtest overfitting probability; investor experience metrics

MSC: 91G10; 91G70; 91G80; 62P05; 62M10

---

1. Introduction

Momentum investing — the practice of buying past winners and avoiding or shorting past losers — represents one of the most replicated anomalies in empirical asset pricing [1–3]. Since the seminal work of Jegadeesh and Titman [1], strategies that rank securities by intermediate-horizon returns and hold the top decile have generated statistically significant risk-adjusted returns across markets, time periods, and asset classes [4,5]. Systematic implementations combining momentum with trend-following rules and dynamic stop-loss mechanisms have attracted substantial practitioner and academic attention, with reported annualised CAGR ranging from single-digit to triple-digit figures depending on the study [6,7].

A recurrent concern in this literature is the difficulty of distinguishing genuine alpha from backtesting artefact. Bailey et al. [8] formalise this problem through the concept of backtest overfitting: the probability that a strategy selected as best in-sample from a grid of N configurations ranks below the median out-of-sample. Under combinatorially symmetric cross-validation (CSCV), they show that this probability — the Probability of Backtest Overfitting, or PBO — approaches 1 as N grows, regardless of whether any individual configuration has genuine predictive power. The practical implication is that most reported backtested strategies are, in expectation, overfit. Bailey and López de Prado [9] further develop the Deflated Sharpe Ratio (DSR) as a correction for selection bias, non-normality, and backtest overfitting, providing a statistical test that controls the false-positive rate across multiple configurations.

Beyond overfitting, a separate class of problems arises from biases that inflate performance irrespective of parameter selection: look-ahead bias and survivorship bias. Look-ahead bias occurs when information not available at the time of a historical trading decision is used in the backtest. Survivorship bias occurs when the instrument universe consists only of those that survived to the end of the sample. Both biases are well known in principle but surprisingly difficult to eliminate in practice, particularly in multi-asset momentum strategies where universe membership, price adjustments, and intraday timing are all time-varying [10,11]. No rigorous taxonomy of look-ahead bias classes for equity trend-following strategies — with each class characterised as a specific filtration violation — appears to exist in the prior literature. López de Prado [11] discusses several forms of data leakage in machine-learning applications to finance — including overlapping training windows, embargoed cross-validation, and feature-engineering with future information — but these treatments are method-specific rather than constituting a unified filtration-based taxonomy organised by the locus of the breach (universe, price series, within-day computation). The taxonomy introduced here is intended to fill this latter gap.

A further gap concerns stop-loss mechanism evaluation. The ATR trailing stop, introduced by Wilder [12], is widely used in systematic strategies, but its mathematical properties are rarely stated formally. In particular, the ratcheting property — whereby the stop level can only move in the

 doi:10.20944/preprints202606.0436.v1

direction favourable to the position holder — is treated as an obvious engineering feature rather than a provable consequence of the recurrence.

Finally, the equity backtesting literature overwhelmingly reports performance using terminal statistics — CAGR, Sharpe ratio, maximum drawdown — without acknowledging that these measures are insufficient to characterise the practical experience of a live investor. An investor who endures three consecutive years of negative returns may be forced to withdraw capital at the worst moment, irrespective of the terminal outcome. We exhibit a deterministic construction (Theorem 2) of two return sequences with identical Sharpe ratio but arbitrarily divergent maximum consecutive negative-year run, providing a mathematical basis for treating investor-experience metrics as independent first-class optimisation objectives.

This paper makes the following specific contributions:

- We introduce a formal definition of point-in-time correctness and a three-class taxonomy of look-ahead bias (Section 3). Each class is formalised as a specific violation of the natural-filtration condition.
- We formalise the ATR trailing stop as a stochastic recurrence and codify its monotonic non-decreasing ratcheting property (Lemma 1, Section 4).
- We exhibit a closed-form construction of return sequences with arbitrary Sharpe / MCNYR pairings (Theorem 2, Section 5), establishing the mathematical independence of investor-experience metrics from Sharpe.
- We complement these theoretical contributions with an 18-year empirical study (Section 6, Section 7). On a NASDAQ-100 point-in-time universe (2008–2025), a 14-configuration ATR-multiplier grid exhibits a stop-multiplier-insensitive performance plateau, structurally consistent loss periods, and a uniform MCNYR of 1 — supporting Theorem 2 in practice. CSCV analysis (PBO = 0.9329) combined with Bonferroni-corrected significance of the plateau Sharpe argues for region-based parameter defence in preference to single-point optimisation.

The structure of the paper is as follows. Section 2 reviews the relevant literature on momentum and trend-following, backtesting bias, stop-loss mechanisms, and investor-experience research. Section 3 introduces the point-in-time correctness framework and the three-class bias taxonomy. Section 4 formalises the ATR trailing stop and proves Lemma 1. Section 5 develops the investor-experience metrics and proves Theorem 2. Section 6 describes the empirical methodology. Section 7 presents the empirical results. Section 8 discusses limitations and implications. Section 9 concludes.

2. Related Work

2.1. Momentum and Trend-Following

Cross-sectional momentum, formalised by Jegadeesh and Titman [1], remains one of the most extensively documented anomalies in finance. Fama and French [2] characterised momentum as a residual factor not explained by the three-factor model; Carhart [13] extended the Fama–French model by adding a momentum factor (UMD), documenting that mutual-fund performance persistence is largely explained by momentum exposure. Cross-sectional momentum profits have been documented across 40+ countries [4], multiple asset classes [14], and over 200 years of historical US data [15,16].

Trend-following conditions position direction on the sign and slope of a price-trend indicator and is related to but distinct from cross-sectional momentum. Moskowitz, Ooi, and Pedersen [14] formalise time-series momentum in futures markets. Practitioner implementations combining trend-confirmation filters with fundamental growth criteria and systematic entry and exit rules produce hybrid strategies that are neither pure momentum nor pure trend-following.

2.2. Backtesting Bias and Overfitting

Bailey, Borwein, López de Prado, and Zhu [8] provide the foundational framework for quantifying backtest overfitting probability via CSCV. Their key result states that the expected out-

 doi:10.20944/preprints202606.0436.v1

of-sample performance rank of the best in-sample configuration declines sharply with the number of configurations tested. Bailey and López de Prado [9] extend this to the Deflated Sharpe Ratio. Harvey, Liu, and Zhu [17] apply related methods to the factor zoo, arguing that most published equity factors are false discoveries. These contributions establish that selection bias and overfitting are pervasive and quantifiable, but do not address the separate problem of structural look-ahead bias embedded in the data pipeline itself.

Survivorship bias in equity datasets has been studied since Brown, Goetzmann, and Ross [18], who showed that mutual-fund databases containing only surviving funds overstate average performance by 0.4–1.4% per annum. Henker et al. [19] document survivorship bias of similar magnitude in individual-stock momentum backtests. Look-ahead bias from use of contemporaneous index membership is documented as well [20–24]. However, systematic decomposition of look-ahead bias into distinct classes — each formalised as a violation of a specific filtration condition — is, to our knowledge, absent from the prior literature.

2.3. Stop-Loss Mechanisms

The Average True Range indicator was introduced [12] as part of a volatility system. Subsequent applications in systematic trading use ATR multiples as trailing-stop distances; the ratcheting property is standard practice but has not been formally proved. Fixed-percentage and ATR-based stops have been compared on US equity momentum strategies [25–27], finding ATR stops produce better risk-adjusted performance. Zakamulin [28] provides a comprehensive empirical comparison of moving-average and stop-based filters.

2.4. Investor Experience and Capital Withdrawal

Sharpe [29] originally introduced the Sharpe ratio as a per-unit-of-risk reward measure; it has since become the dominant scalar performance summary. Young [30] proposed the Calmar ratio as a return-over-drawdown alternative. Berk and Green [31] formalise capital flows into and out of mutual funds as a function of recent performance, providing a theoretical basis for the claim that consecutive losses can trigger forced withdrawals. Agarwal and Naik [32] document persistence of hedge fund withdrawals after drawdown periods. Sapp and Tiwari [33] and Kinniry et al. [34] further document that retail and institutional investors exhibit return-chasing behaviour that imposes a behavioural cost beyond the Sharpe ratio. These findings collectively argue that maximum consecutive negative-year run (MCNYR) and time-under-water (TUW) capture aspects of investor experience that are not summarised by Sharpe alone.

3. Point-in-Time Correctness and a Three-Class Taxonomy of Look-Ahead Bias

3.1. The Natural Filtration Condition

Let $(\Omega, F, \mathbb{P})$ be a probability space carrying all market information relevant to a backtest, and let $\{\mathcal{F}_t\}_t \geq 0$ denote the natural filtration, where $\mathcal{F}_t$ is the σ-algebra of events observable up to and including time t. A backtest evaluates a deterministic mapping from market data to trading decisions; let d_t denote the trading decision at time t (a vector of position changes) and let $\mathcal{I}_t$ denote the information set used to compute d_t.

Definition 1 (Point-in-Time Correctness). *A backtest is point-in-time correct if, for every decision time t, the information set $\mathcal{I}_t$ satisfies $\mathcal{I}_t \subset \mathcal{F}_t$ — that is, the inputs to the decision at time t are measurable with respect to the σ-algebra of events observable at or before t.*

A backtest that violates this condition produces decisions informed by future market data and therefore overstates achievable performance. The remainder of this section enumerates three distinct ways in which the condition is commonly violated.

 doi:10.20944/preprints202606.0436.v1

3.2. Class I – Universe Membership Contamination

Definition 2 (Class I Bias). *A backtest exhibits universe-membership contamination if, for a decision at time t, the eligible universe U_t used to select securities includes the symbol identifiers of an index roster that was published or determined after time t.*

The canonical case arises when a NASDAQ-100 or S&P 500 backtest uses the current index roster as the universe at every historical date. Securities subsequently added to the index are erroneously included in past decisions; securities subsequently removed (for delisting, bankruptcy, or fundamental deterioration) are erroneously excluded. The latter case produces survivorship-type bias [18,35–37]. Mitigation requires a point-in-time index-constituency dataset, reconstructed from quarterly or monthly index publications.

3.3. Class II – Price Data Forward Leakage

Definition 3 (Class II Bias). *A backtest exhibits price-data forward leakage if, for a decision at time t, the price series p_t used as input contains adjustments — for splits, dividends, or corporate actions — that occurred after time t.*

Modern data vendors typically distribute prices as a back-adjusted series in which historical bars are retroactively scaled by future split and dividend ratios. A naive backtest that consumes these series treats the back-adjusted price as the price observable at time t, when in fact the adjustment depended on events not yet realised. The magnitude of this leakage depends on the proportion of universe securities undergoing such actions during the sample. Mitigation requires either raw unadjusted prices with explicit handling of corporate actions, or a vendor-provided point-in-time adjusted series.

3.4. Class III – Stop-Exit Sequencing Violations

Definition 4 (Class III Bias). *A backtest exhibits stop-exit sequencing bias if, for a position entered at time t_{entry} on calendar day D, the operative stop level at any subsequent time $t > t_{entry}$ on the same day D incorporates price data from times t′ on day D that were not yet observable at time t — that is, $t' \in [t_{open}, t_{eod}]$ with $t' > t$.*

Class III bias most commonly arises in vectorised implementations of trailing-stop calculations. A correct implementation iterates over minute bars sequentially, updating the stop level using only price data from bars at or before the current time. A vectorised implementation may, by accident, compute the stop level on the entry bar using a daily summary statistic (e.g., a max operation across all bars of the day) that incorporates post-entry price action. The empirical magnitude of Class III bias is implementation- and configuration-dependent: in strategies dominated by hard-stop binding or by short holding periods, the effect can be negligible; in strategies where the trailing stop is the operative exit mechanism over multi-day positions, the effect can be substantial. We leave empirical quantification of Class III bias on equity momentum strategies — both via known production bug forensics and via calibrated synthetic constructions — as an open question for future work.

3.5. Summary

The three bias classes correspond to three distinct loci of filtration violation: the universe definition (Class I), the input price series (Class II), and the within-day stop computation (Class III). Each class is independent in the sense that a backtest can satisfy any two while violating the third. The combined elimination of all three is necessary, but not sufficient, for point-in-time correctness in the sense of Definition 1; additional violations (e.g., look-ahead in volatility estimation, in

 doi:10.20944/preprints202606.0436.v1

fundamental-data alignment, or in regime-classifier inputs) are possible but lie outside the scope of this paper.

4. The ATR Trailing Stop as a Stochastic Recurrence

4.1. Setup

Let P_t denote the price of a security at time t, and let ATR_t denote the Average True Range over a fixed lookback window. For a long position entered at time t_{entry} at price P_{entry}, the ATR trailing stop with multiplier $k > 0$ is defined as the following recurrence on the operative stop level S_t:

Definition 5 (ATR Trailing Stop Recurrence).

$$S_{t_{entry}} = P^{low}_{entry} - k \cdot ATR_{t_{entry}}$$
$$For\ t > t_{entry}: S_t = max\,(S_{t-1}, P_t - k \cdot ATR_t) \quad (5)$$

The position is closed at time t if $P_t \leq S_t$.

4.2. The Ratcheting Lemma

Lemma 1 (ATR Stop Ratcheting). *Let* $\{S_t\}_t \geq t_{entry}$ *be the sequence of stop levels defined by the recurrence in Definition 5. Then for every* $t \geq t_{entry}$, *we have* $S_t \geq S_{t-1}$, *i.e., the sequence is monotonically non-decreasing.*

Proof of Lemma 1. By construction, $S_t = \max(S_{t-1}, P_t - k \cdot ATR_t)$. Since S_t is the maximum of $S_{\{t-1\}}$ and another value, we have $S_t \geq S_{t-1}$. □

Corollary 1 (Downside Protection). *Once an ATR trailing stop has been raised to level* S^*, *the position cannot subsequently be exited at any price below* S^*.

4.3. Discussion

Lemma 1 formalises a standard practitioner heuristic. While the proof is a direct consequence of the recurrence, the property has two structural implications worth recording explicitly. First, the stop level provides monotonically non-decreasing downside protection during any open trade — a guarantee that adaptive volatility-based stops, computed from time-varying conditional variance estimates, cannot provide without explicit max-clipping over prior stop levels. Second, the maximum drawdown of any individual trade closed at the trailing stop is bounded by $P_{entry} - S_{final}$, providing a closed-form upper bound on per-trade loss.

5. Investor Experience Metrics

5.1. Definitions

Let $r_1, r_2, \ldots, r_t$ denote the annual net-of-tax return sequence of a strategy over T calendar years.

Definition 6 (Maximum Consecutive Negative Year Run, MCNYR). *The maximum consecutive negative year run is*

$$MCNYR = \max\{\, b - a + 1 : 1 \leq a \leq b \leq T,\ r_t < 0 \text{ for all } t \in [a, b]\}. \quad (6)$$

with the convention $MCNYR = 0$ *when the set over which the maximum is taken is empty (i.e., when* $r_t \geq 0$ *for all* t*).*

Definition 7 (Time-Under-Water, TUW). *Let* $V_t = V_0 \cdot \prod_{s=1}^{t}(1 + r_s)$ *denote the portfolio value at year-end t and* $M_t = max_{s \leq t} V_s$. *The fractional time-under-water is*

$$TUW = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{V_t < M_t\}. \quad (7)$$

 doi:10.20944/preprints202606.0436.v1

Definition 8 (Rolling Calmar Ratio). *For a window of length w years and terminal year T, the rolling Calmar ratio is*

$$R - Cal_{T,w} = \frac{CAGR_{T-w:T}}{|MaxDD_{T-w:T}|} \ . \tag{8}$$

5.2. The Sharpe-MCNYR Independence Theorem

The Sharpe ratio is the standard scalar performance summary, but it depends on returns only through their first two empirical moments. Two return sequences with identical mean and identical variance can therefore have arbitrarily different patterns of consecutive losses. The following construction makes this explicit.

Theorem 2 (Closed-Form Sharpe-MCNYR Counterexample). *Let $\mu \in \mathbb{R}$ and $\sigma > 0$ be a target empirical mean and standard deviation, and let $T \geq 2$ and $k \in \{1, 2, \ldots, T-1\}$ be the target horizon and target MCNYR. Define $p = \frac{k}{T}$, $q = 1 - p$, and*

$$a = \sigma \cdot \sqrt{\frac{q}{p}} \ , \ b = \frac{p}{q} \cdot a, \tag{9}$$

$$d = a - \mu, \ g = b + \mu \, . \tag{10}$$

Define the deterministic sequence r^B by $r_t^B = -d$ for $t = 1, \ldots, k$ and $r_t^B = g$ for $t = k+1, \ldots, T$. Then r^B has empirical mean μ, empirical standard deviation σ (population, ddof = 0), and $MCNYR(r^B) = k$ provided $d > 0$ (equivalently $\mu < \sigma \cdot \sqrt{\frac{q}{p}}$).

Proof of Theorem 2. The mean condition $p \cdot (-d) + q \cdot g = \mu$ is equivalent, after substituting $d = a - \mu$ and $g = b + \mu$, to $q \cdot b = p \cdot a$. Setting $b = \frac{p}{q} \cdot a$ satisfies this exactly. The variance condition $p \cdot (-d - \mu)^2 + q \cdot (g - \mu)^2 = \sigma^2$ becomes $p \cdot a^2 + q \cdot b^2 = \sigma^2$. Substituting $b = \frac{p}{q} \cdot a$:

$$p \cdot a^2 + q \cdot \left(\frac{p}{q} \cdot a\right)^2 = p \cdot a^2 \cdot \left(1 + \frac{p}{q}\right) = p \cdot a^2 \cdot \left(\frac{1}{q}\right) = \sigma^2, \tag{11}$$

which yields $a^2 = \sigma^2 \cdot \frac{q}{p}$, i.e., $a = \sigma \cdot \sqrt{\frac{q}{p}}$ as in required. $MCNYR(r^B) = k$ follows because the first k entries are all equal to the negative value $-d$ (assuming $d > 0$) and the remaining $T - k$ entries are all positive. The condition $d > 0$ reduces to $a > \mu$, i.e., $\mu < \sigma \cdot \sqrt{\frac{q}{p}}$. For any μ, σ, T this condition is satisfied for k sufficiently small relative to T. □

Corollary 2 (Sharpe-MCNYR Independence). *For any μ, σ, T such that the construction in Theorem 2 admits valid $k \in \{1, \ldots, k_{max}(T, \mu, \sigma)\}$, there exist return sequences with Sharpe ratio $\frac{\mu}{\sigma}$ and any prescribed MCNYR value in $\{1, \ldots, k_{max}\}$. The Sharpe ratio therefore provides zero information about MCNYR.*

5.3. Numerical Verification

The construction in Theorem 2 was verified numerically; the verification table is reported in Appendix A. For target parameters $(\mu = 0.10, \sigma = 0.20, T = 18)$ and $k \in \{1,2, \ldots, 9\}$, the constructed sequences exhibit, in every case, empirical mean 0.10000, empirical standard deviation 0.20000 (ddof = 0), and Sharpe ratio $(\mu/\sigma) = 0.5000$, with MCNYR equal to the prescribed k. The Sharpe ratio is invariant across the prescribed MCNYR values, as required by Corollary 2.

5.4. Empirical Relevance

Theorem 2 establishes that Sharpe and MCNYR are mathematically independent. A separate question is whether realistic strategies empirically exhibit MCNYR variation that is decoupled from Sharpe. Section 7 reports on this empirically: across a 14-configuration grid spanning a 7.3× range of

 doi:10.20944/preprints202606.0436.v1

the operative parameter, annualised Sharpe varies from 0.741 to 0.831 (a 12% spread on annual returns) while MCNYR is uniformly equal to 1. The grid therefore exhibits the property predicted by Theorem 2 in a realistic strategy setting.

6. Empirical Methodology and Data

6.1. Strategy

We study a systematic equity strategy that combines a published-style trend-confirmation filter with a price-action entry rule and an ATR trailing stop. The strategy is described at category level to preserve commercial confidentiality. Its main components are:

1. A multi-criteria trend-confirmation filter applied to each candidate security.
2. A point-in-time price-action entry rule triggered when the trend filter is satisfied, and a confirmation condition holds.
3. An ATR trailing stop with multiplier k, ratcheting as defined in Section 4.
4. A hard maximum loss percentage applied at the position level.

Position management policy: at the end of each quarterly rebalancing period, all open positions are closed at the period-end price (with transaction costs and slippage applied) and the next period's basket is constructed independently from the point-in-time eligible universe. This force-close-at-rebalance policy follows the convention used in published momentum studies [1,4,7] and ensures that period returns are well-defined and independent of cross-period stop-out timing. A continuous-rebalancing variant in which positions carry over across rebalancing dates is more representative of live trading but introduces path-dependencies orthogonal to the bias-correction framework studied here; we defer that variant to future work.

Within-period exit behaviour. When a position is stopped out by the ATR trailing stop or the hard-loss stop within a rebalancing period, the released capital remains uninvested (held as cash, earning zero return) until the next quarterly rebalance date. No within-period re-entry, basket substitution, or cash redeployment is implemented in this engine. A position stopped out on, for example, the second week of February remains as cash until the April rebalance, at which point a fresh basket is selected. This convention follows the standard academic momentum-portfolio formulation [1,4,14] and produces a conservative estimate of strategy performance relative to a live implementation that could redeploy stopped-out capital; we adopt it for methodological transparency and comparability with the published literature, while acknowledging that production deployments of similar strategies typically implement dynamic capital redeployment that would reduce the cash drag.

6.2. Macro Regime Conditioning

Position sizing and exposure are modulated by a macro regime classifier that takes as inputs a set of standard macroeconomic indicators. The classifier output is held fixed across all configurations in the grid analysis below; only the ATR multiplier varies across configurations. Full disclosure of the classifier construction is deferred to a companion paper for commercial reasons.

6.2.1. Macro Regime Classifier – Category-Level Description

While specific implementation details of the macro regime classifier are withheld for commercial reasons, the category-level structure characterised below is sufficient for assessing the role the classifier plays in the empirical results — specifically, considering the unconditional ablation reported in Appendix B.

Input class: The classifier takes as input a small set (fewer than ten) of standard macroeconomic and market-state indicators publicly available with no lag relative to the decision date. These indicators are drawn from three categories: (i) equity-market volatility measures, (ii) interest-rate term-structure measures, and (iii) cross-sectional market-breadth measures.

 doi:10.20944/preprints202606.0436.v1

Output class: The classifier produces a discrete state from a small finite set ($K \leq 5$ states), interpretable as positioning along a "risk-on / risk-off" axis. Position-sizing in the strategy is conditioned on this discrete state through a deterministic mapping from state to per-position weight cap and to portfolio-level gross-exposure cap.

Calibration class: The state-classification rule is trained on data prior to the backtest sample (pre-2008); the classifier parameters are fixed at the start of the 2008 sample and unchanged thereafter. The classifier is therefore time-invariant from the perspective of the backtest, satisfying the natural-filtration condition of Definition 1.

Variant for robustness: To establish that the empirical findings are not artefacts of the regime classifier, we present in Appendix B a parallel grid run with the classifier output forced to the most aggressive state across all decision dates (the "unconditional variant"). The plateau structure, the uniform MCNYR, and the structural consistency of loss periods all persist in the unconditional variant, with absolute CAGR and Sharpe reduced by a measured amount reflecting the conditioning's positive contribution.

6.3. Data

The empirical study uses one-minute OHLCV data for NASDAQ-100 constituents from January 2008 through December 2025. The data are obtained from a commercial intraday data vendor under licence; redistribution is precluded by the licence terms.

NASDAQ-100 index membership is reconstructed from quarterly index publications, providing a point-in-time eligible universe that satisfies the Class I (Definition 2) non-contamination requirement. Securities are included in the eligible universe of decision date t only if they were members of the published index at t; subsequent inclusions and exclusions are correctly reflected at the next quarterly publication date.

Price adjustments for splits and dividends are applied using a back-adjusted series. We acknowledge that this represents a residual Class II (Definition 3) concern; rigorous mitigation would require a point-in-time adjusted series, which is a known limitation discussed in Section 8.

6.3.1. Bias Compliance Status of This Study

The bias-taxonomy framework introduced in Section 3 enables a structured statement of which compliance criteria are satisfied by the empirical implementation reported below. We provide that statement explicitly rather than leaving it to be inferred.

Class I — Universe-membership contamination (Definition 2): satisfied. The NASDAQ-100 eligible universe is reconstructed from quarterly index publications; securities are included in the universe at decision date t only if they were constituents at t. Subsequent inclusions and exclusions enter the eligible set only at their actual publication dates.

Class II — Price-data forward leakage (Definition 3): residual exposure, not fully mitigated. Prices are sourced as a back-adjusted series. Adjustments for splits and dividends are therefore folded retroactively into the historical bars. The magnitude of this exposure depends on the proportion of NDX-100 constituents undergoing such actions during the sample; on a developed-market large-cap universe this proportion is modest, but the residual filtration violation is real. Full mitigation would require either raw unadjusted prices with explicit corporate-action handling or a vendor-provided point-in-time adjusted series; neither is implemented here. This study should therefore be understood as a demonstration of the bias-taxonomy framework's diagnostic value — specifically, of its capacity to identify and label the residual non-compliance in an otherwise carefully constructed pipeline — rather than as a fully PIT-compliant empirical exemplar.

Class III — Stop-exit sequencing (Definition 4): satisfied. The backtest engine iterates over minute bars sequentially within each calendar day, updating the ATR trailing stop level using only price data from bars at or before the current decision time. No within-day summary statistics (open-to-close max, etc.) enter the stop computation.

 doi:10.20944/preprints202606.0436.v1

We are explicit about the Class II residual because the framework's value depends on it being applied honestly to its own application. A reader can correctly inspect the results reported below as those of a partially PIT-compliant backtest, with the partial non-compliance bounded in nature (limited to the corporate-action adjustment channel) and localised in source (vendor-side back-adjustment).

6.4. Costs and Taxes

Transaction costs are modelled at 0.20% per round trip (combining commission and slippage) for all NASDAQ-100 trades. The hard maximum loss percentage at the position level is set to 13%.

Annual realised gains are taxed under the Dutch progressive tax schedule applicable to box-1 capital gains: 36% on the first €68,508 of realised gain per year and 50% on realised gain above that threshold. All numerical performance values reported below are net of this tax computation. This treatment ensures that reported returns correspond to investor cash flows after tax.

6.5. The 14-Configuration ATR Grid

We evaluate the strategy across 14 ATR-multiplier configurations:

$$k \in \{1.5,\ 2.0,\ 2.5,\ 3.0,\ 3.5,\ 4.0,\ 4.5,\ 5.0,\ 6.0,\ 7.0,\ 8.0,\ 9.0,\ 10.0,\ 11.0\}$$

All other parameters are held constant across configurations. Each configuration is run as a separate backtest on the full 2008–2025 period inside a frozen evaluation environment (containing the engine code, the constituent roster, the data path, and the tax computation). The grid is asymmetric: the lower end ($k = 1.5$,2.0) captures the tight-stop regime where the trailing stop binds frequently within rebalancing periods; the middle ($k = 3.5$–7.0) captures the regime where the stop is operative but rarely binds; the upper end ($k = 8.0$–11.0) captures the regime where the stop is so wide that it effectively never fires within a quarter at quarterly rebalancing.

6.6. Combinatorially Symmetric Cross-Validation

We apply the CSCV procedure of Bailey et al. [8] to assess the Probability of Backtest Overfitting (PBO) of the grid. The procedure operates on the matrix of per-period returns $R_{i,j}$, where i indexes time periods and j indexes configurations.

The 18-year sample is partitioned into 72 quarterly observations and then aggregated into $S = 12$ contiguous sub-periods of 6 quarters each. For each of the C(12,6) = 924 possible ways to split the 12 sub-periods into 6 in-sample and 6 out-of-sample sub-periods, we identify the in-sample winner — the configuration with the highest annualised Sharpe ratio on the 36 concatenated in-sample quarters — and compute its out-of-sample rank among all 14 configurations on the 36 concatenated out-of-sample quarters. The PBO is defined as the fraction of partitions in which the in-sample winner's OOS rank falls below the median (rank > 7 in our 14-configuration grid).

The sample provides $72/14 \approx 5.1$ observations per configuration. A common heuristic threshold for stable Sharpe estimation across a grid search is ~ 20 observations per configuration; our sample falls below this threshold, a limitation we disclose explicitly and discuss in Section 8.

6.7. Bonferroni-Corrected Significance

In place of the Deflated Sharpe Ratio of Bailey and López de Prado [11], which requires careful estimation of higher empirical moments and the expected maximum Sharpe ratio under multiple trials, we adopt a more conservative and verifiable alternative: the Bonferroni-corrected p-value of the full-sample winner's Sharpe ratio. The procedure is:

1. Compute the t-statistic for the in-sample winner's annualised Sharpe ratio: $t = \mathrm{SR} \cdot \sqrt{T - 1}$, where $T = 72$ quarters.
2. Compute the two-sided p-value under the standard normal approximation.
3. Multiply by $N = 14$ (the grid size) to obtain the Bonferroni-corrected p-value.
4. Compare the corrected p-value to $\alpha = 0.05$.

 doi:10.20944/preprints202606.0436.v1

Following step admits two natural choices for the effective number of independent configurations: the naïve count $M = 14$, which assumes statistical independence across the grid; and the effective count M_{eff} computed by the method of Li and Ji [38] from the correlation-matrix eigenvalue spectrum of the configurations' return series. We report both in Section 7.4 and demonstrate that the significance conclusion is robust to either choice. The strict $M = 14$ value serves as a deliberately conservative upper bound on the achieved p-value; the M_{eff} value reflects the actual statistical structure of the grid as established empirically.

This procedure is well-known to be more conservative than DSR; we treat the trade-off in favour of methodological transparency. The DSR framework remains the more refined alternative, and we acknowledge its precedence in the literature.

6.8. Reproducibility

The empirical results below were generated inside a frozen evaluation environment containing pinned engine code, pinned constituent rosters, and a fixed snapshot of the underlying price data. All 14 ATR-grid runs and the auxiliary analyses (CSCV, full-sample Sharpe ranking, negative-year and time-under-water identification) were produced on a single machine over a single campaign and verified for internal consistency: across all 14 runs, annual returns compound to the reported CAGR within 0.0001 absolute, and the engine's internal reconciliation passes on every run.

Reproducibility is stratified across the paper's contributions. The methodological framework — the bias taxonomy of Section 3, the recurrence of Section 4, the closed-form construction of Section 5, the CSCV procedure of Section 6.6, and the Bonferroni–Li–Ji significance procedure of Section 6.7 — is fully specified and independently reimplementable. The numerical empirical results of Section 7 and Appendix B depend additionally on three components withheld for commercial reasons: the specific trend-confirmation filter and entry-rule thresholds (Section 6.1), the macro regime classifier (Section 6.2, partially characterised at category level in Section 6.2.1), and the strategy engine source code. Independent reimplementation of the headline empirical numbers is therefore not possible without these components; what is reproducible from the supplementary materials provided is the *verification* of the reported numbers (CSCV partition counts, Li-Ji effective M, Bonferroni p-values, eigenvalue spectra) from the consolidated quarterly and annual return series.

The supplementary materials accompanying this paper include: (i) the consolidated CSV outputs underlying Tables 1, 2, and 3, for both the conditional and unconditional grids; and (ii) the verification scripts that reproduce the headline statistics ($M_e$ff, PBO, Bonferroni p-values, Theorem 2 numerical verification) from the consolidated CSVs. Raw per-trade log files, daily equity series, and figure-generation scripts are available from the author on reasonable request.

7. Empirical Results

7.1. Equity Curve and Performance Plateau

Figure 1 displays the net-of-tax equity curve of the $k = 4.5$ configuration over the full 2008–2025 sample. The maximum portfolio drawdown is -30.4%, reached in early 2009 during the global financial crisis (GFC). A secondary drawdown is visible in 2021. From a \$50,000 initial capital, the portfolio terminates at \$283,000 at end-of-year 2025.

Figure 2 displays the full-sample annualised Sharpe ratio of each of the 14 configurations as a function of the ATR multiplier k. The figure exhibits the central empirical finding of this paper: 12 of the 14 configurations ($k = 2.5$ through $k = 11.0$) cluster within a 0.025 Sharpe range (0.7907 to 0.8153), with mean Sharpe 0.805 ± 0.007 across this region. The tightest configuration ($k = 1.5$) sits as an outlier at Sharpe 0.7411, approximately 0.06 below the cluster mean. A second distinguished configuration ($k = 2.0$) sits above the cluster at Sharpe 0.8305, exceeding the cluster mean by 0.025 — a difference that, as Section 7.4 will show, is not statistically distinguishable from sampling noise on 72 quarterly observations.

 doi:10.20944/preprints202606.0436.v1

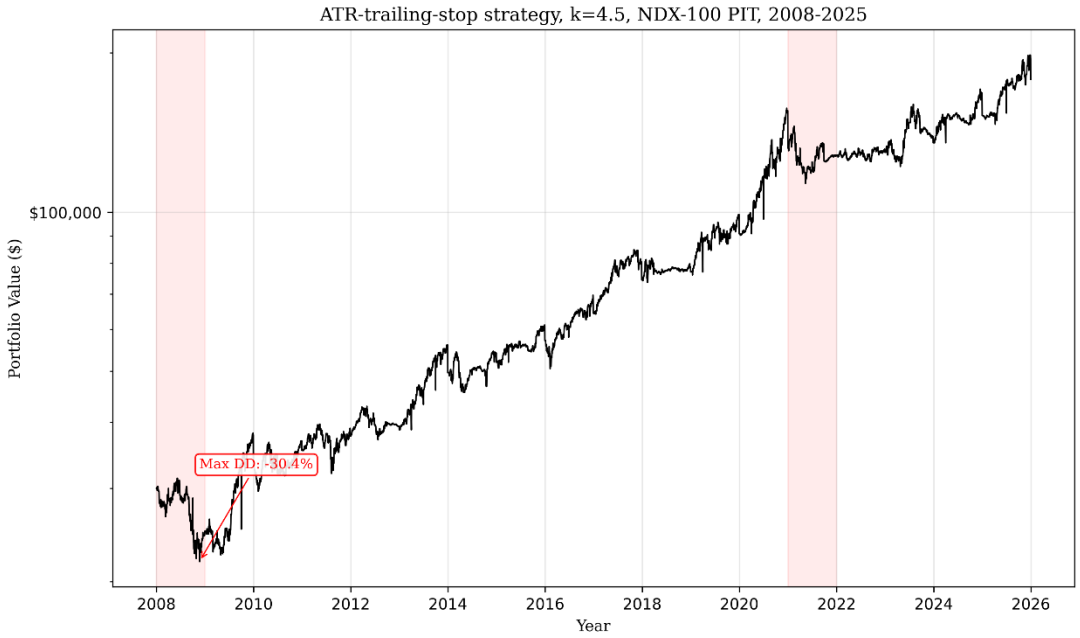

Figure 1. Headline equity curve for $k = 4.5$, log scale, 2008–2025, with GFC and 2021–2022 stress periods shaded.

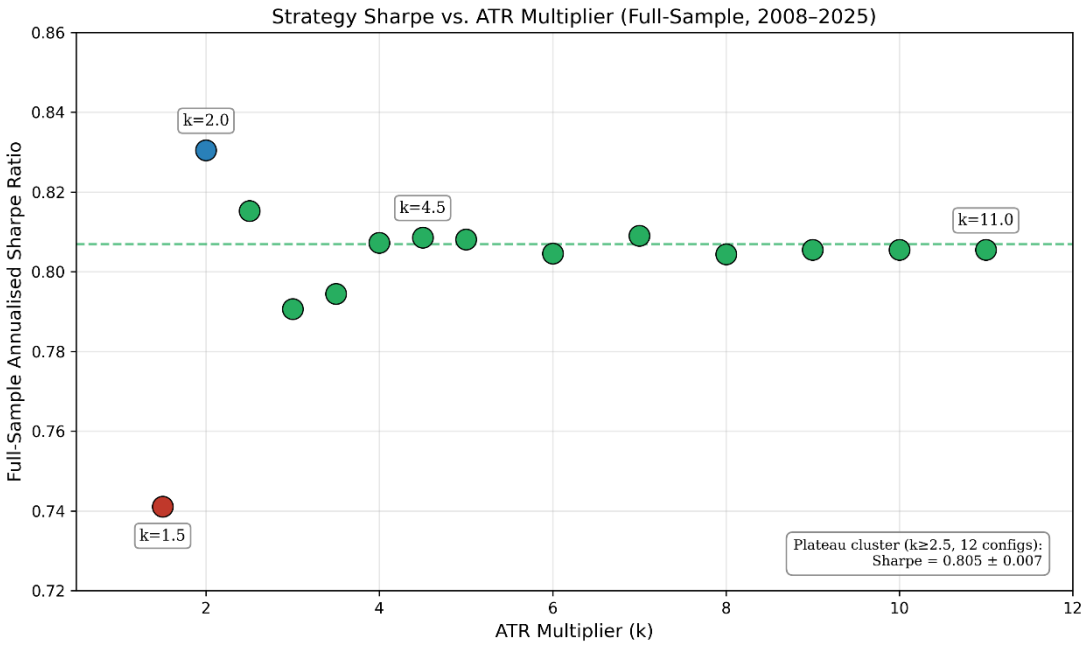

Figure 2. Full-sample annualised Sharpe ratio vs. ATR multiplier k, showing the plateau cluster, the $k = 2.0$ above-plateau point, and the $k = 1.5$ outlier.

Figure 3 overlays the equity curves of all 14 configurations on a single log-scale axis. Visually, through the entire 2008–2025 period, the curves are nearly indistinguishable from each other. Only the $k = 1.5$ configuration (lowest curve from 2017 onwards) and to a lesser extent the $k = 2.0$ configuration diverge from the dense cluster of the remaining 12 plateau configurations.

 doi:10.20944/preprints202606.0436.v1

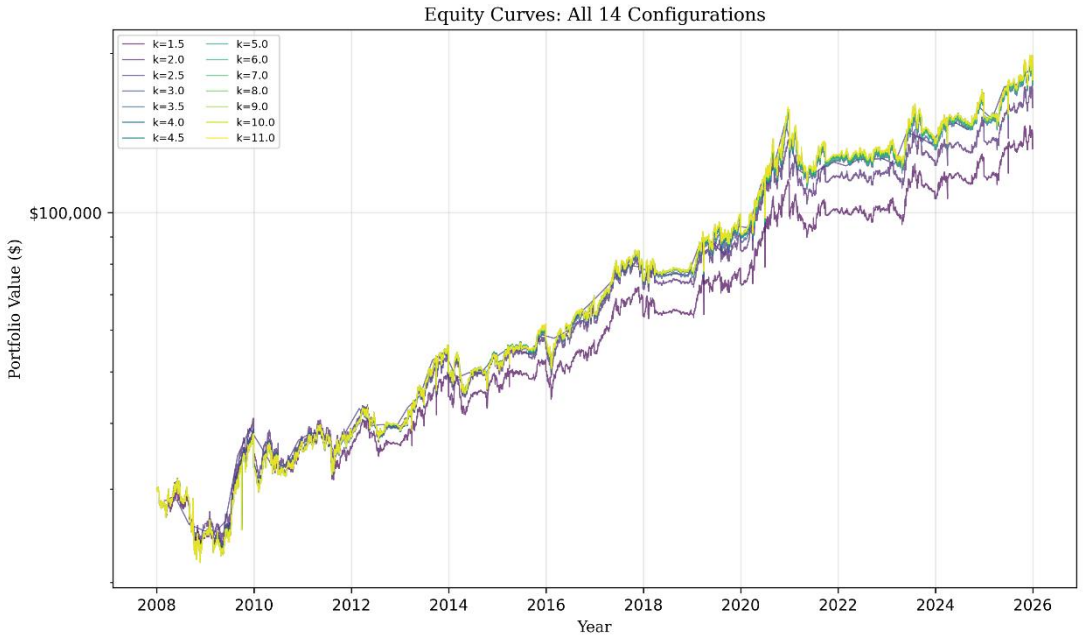

Figure 3. All 14 equity curves overlaid on log scale; the plateau region is visually indistinguishable.

7.2. Investor-Experience Vector

Table 1 reports the full investor-experience vector (IEV) for each of the 14 configurations: CAGR, annualised Sharpe, MaxAnnualDrawdown, NegYears (count of years with net negative return), MCNYR, TUW, and the minimum rolling 5-year Calmar ratio.

Table 1. Investor-experience metrics across the 14-configuration ATR-multiplier grid (NASDAQ-100 PIT, 2008–2025, net of Dutch progressive tax). Sharpe is computed at annual frequency (18 observations, population standard deviation).

k	CAGR	Sharpe	MaxAnnDD	NegYears	MCNYR	TUW	Calmar5Y_min
1.5	8.56%	0.7411	−18.64%	1	1	0.000	0.1983
2.0	9.65%	0.8305	−15.06%	1	1	0.000	0.3456
2.5	10.19%	0.8153	−16.08%	2	1	0.111	0.3270
3.0	10.24%	0.7907	−18.25%	2	1	0.111	0.2687
3.5	10.35%	0.7945	−18.28%	2	1	0.111	0.2707
4.0	10.35%	0.8073	−18.28%	2	1	0.111	0.2682
4.5	10.39%	0.8086	−18.40%	2	1	0.111	0.2725
5.0	10.36%	0.8081	−18.40%	2	1	0.111	0.2749
6.0	10.28%	0.8046	−18.40%	2	1	0.111	0.2735
7.0	10.33%	0.8090	−18.40%	2	1	0.111	0.2734
8.0	10.38%	0.8044	−18.40%	2	1	0.111	0.2722
9.0	10.39%	0.8055	−18.40%	2	1	0.111	0.2736
10.0	10.39%	0.8055	−18.40%	2	1	0.111	0.2736

 doi:10.20944/preprints202606.0436.v1

11.0	10.39%	0.8055	−18.40%	2	1	0.111	0.2736

Two structural features of the grid are immediately visible. First, the CAGR range across the plateau $k \in [3.5,7.0]$ is only 0.11 percentage points (10.28% to 10.39%) — a band narrower than the noise on a single configuration's annual return. Second, MCNYR is uniformly equal to 1 across the entire grid, while Sharpe ranges from 0.7411 to 0.8305 (a 0.090 spread). This is exactly the property predicted by Theorem 2 (Sharpe and MCNYR are independent): the grid realises 14 distinct Sharpe values while MCNYR remains invariant.

Two configurations within the grid merit specific mention beyond the plateau-region framing. The configuration with the highest annual-frequency Sharpe ($k = 2.0$, Sharpe 0.8305) also exhibits the lowest maximum annual drawdown (−15.06%), the highest minimum rolling 5-year Calmar ratio (0.3456), and zero time-under-water — at the cost of approximately 0.74 percentage points of CAGR relative to the plateau mean. The headline plateau configuration ($k = 4.5$) achieves the maximum CAGR of the grid (10.39%) with a Sharpe of 0.8086, MaxAnnualDD of −18.40%, and Calmar5Y_min of 0.2725. The trade-off between tighter and wider stops is therefore real and multi-criteria; we report the plateau as the principal empirical finding because the PBO analysis in Section 7.4 shows that finer point-identification within or at the edge of the plateau is not supported by the available sample.

7.3. Structural Consistency of Loss Periods

Across the 12 plateau configurations ($k = 2.5$ through $k = 11.0$), the negative-return years are uniformly 2008 (GFC, with losses ranging from −16.08% to −18.40%) and 2021 (with losses ranging from −1.09% to −3.43%). The two tightest configurations ($k = 1.5$ and $k = 2.0$) avoid 2021 entirely. No other year in the 18-year sample exhibits a net negative return for any configuration.

The time-under-water years are similarly structurally consistent. Across the 12 plateau configurations, year-end portfolio values are below the running maximum in exactly two years (2021 and 2022), corresponding to TUW = 2/18 = 0.111. In 2021 the portfolio drops below the 2020 peak due to the negative return; in 2022 the portfolio earns a small positive return (1.37% to 1.49%) that is insufficient to recover the 2020 peak. The configurations with TUW = 0 ($k = 1.5$, $k = 2.0$) achieve full peak-respect throughout the sample.

The interpretation is that the strategy's underperformance is regime-driven rather than parameter-tuned: the same two stress periods affect all plateau configurations, with magnitudes that vary monotonically with the stop multiplier but with timing and identity that are invariant. This is the empirical signature of a strategy whose edge is genuine rather than a result of parameter selection.

7.4. Grid Statistical Structure, Probability of Backtest Overfitting, and Bonferroni Significance

We now examine the statistical structure of the 14-configuration grid as an object, the Probability of Backtest Overfitting that follows from that structure, and the Bonferroni-corrected significance of the plateau-level performance against the null of zero excess return.

7.4.1. The Grid Carries One Effective Degree of Freedom

The 14 configurations were chosen to span a 7.3× range of the ATR multiplier k, from a tight-stop regime through an operative-stop regime into a regime where the stop effectively never fires within a quarter. A priori, these configurations could have produced 14 statistically distinct return series. They do not.

Let $\mathbf{R}$ denote the 72×14 matrix of quarterly net-of-tax returns. The 14×14 correlation matrix of $\mathbf{R}$ has the following eigenvalue spectrum (descending):

$$\boldsymbol{\lambda_1} = 13.9097, \quad \boldsymbol{\lambda_2} = 0.0735, \quad \boldsymbol{\lambda_3} = 0.0083, \quad \boldsymbol{\lambda_{4..14}} < 0.005$$

The dominant eigenvalue absorbs 99.36% of the total variance ($\lambda_1 / \sum_i \lambda_i = 13.91/14 = 0.9936$); the second eigenvalue carries 0.52%; the remaining twelve eigenvalues collectively carry 0.12%. The full spectrum is reported in Appendix C. Applying the effective-M method of Li and Ji [38], which

 doi:10.20944/preprints202606.0436.v1

deducts from M the excess of correlation-matrix eigenvalues above unity, the 14-configuration grid yields $M_{eff} = 1.09$. Statistically, the grid behaves as essentially one configuration evaluated at 14 nearly identical perturbations. This is itself a finding worth recording: variation in k across the operative range — a range chosen ex ante to expose the strategy to materially different stop behaviour — produces return series that span essentially one statistical dimension within the available 72-quarter sample. The trailing-stop multiplier is a near-redundant parameter for the strategy as implemented over this sample.

Robustness across model variants: The same eigenvalue analysis applied to a parallel grid in which the macro regime classifier is disabled (constant maximum exposure across all dates — Appendix B for details) yields $\lambda_1 = 13.9091$, $M_{\text{eff}} = 1.091$— essentially identical to the conditional grid. The eigenvalue collapse is therefore structural to the ATR-multiplier strategy class on this universe, not an artefact of the regime-conditioning module. We return to this point in Section 7.4.4 and Appendix B.

7.4.2. PBO Reflects In-Sample Noise Selection on a Statistically Near-Degenerate Grid

The CSCV procedure of Section 6.6 selects, in each of the C(12,6) = 924 partitions, the configuration with the highest Sharpe ratio on a randomly chosen half of the sub-periods and computes its rank on the complementary half. The PBO is the fraction of partitions in which the in-sample winner's out-of-sample rank falls below the median.

Under a pure-noise null — configurations statistically identical, in-sample winner chosen uniformly at random — the expected PBO is 0.5. Under the maximum-Sharpe selection rule used by CSCV, the in-sample winner is the configuration that benefitted most from noise on the chosen sub-periods; this benefit reverses on average out-of-sample, producing PBO substantially above 0.5 when the configurations are near-degenerate and the signal-to-noise ratio is high. Bailey et al. [8] characterise this as an inherent feature of CSCV applied to grids with low effective dimensionality. We compute:

$$\text{PBO} = 0.9329.$$

This value is consistent with the eigenvalue collapse of Section 7.4.1. When $M_{\text{eff}} = 1.09$, grid-internal selection operates on the 0.65% of variance orthogonal to the dominant common mode; this orthogonal variance systematically reverses out-of-sample, pushing PBO toward its high-side asymptote. The mechanism is the standard regression-to-the-mean of maximum selection on noisy estimates, operating here on a grid whose configurations are statistically near-identical. PBO is therefore an expected consequence of the eigenvalue collapse, not an indication of overfitting in any meaningful sense: the grid is statistically too narrow to support point identification, and selecting a configuration from within it produces systematic out-of-sample underperformance as it must. Appendix B demonstrates that under reduced signal strength (the unconditional-classifier variant) PBO declines to 0.786 while M_{eff}remains at 1.09; the two quantities therefore measure related but distinct properties of the grid, both of which support the region-based interpretation that follows.

7.4.3. The Plateau-Level Performance Is Highly Significant

Performance significance is established separately from grid-internal selection. The full-sample quarterly-frequency annualised Sharpe ratio of the highest-Sharpe configuration ($k = 2.5$, quarterly Sharpe 0.7081) yields a t-statistic of 5.97 over the 72-quarter sample, with uncorrected two-sided p-value 2.4×10^{-9}. Multiple-comparisons correction by Bonferroni:

- At the conservative count $M = 14$: corrected p-value 3.4×10^{-8}.
- At the eigenvalue-spectrum-derived count $M_{\text{eff}} = 1.09$: corrected p-value 2.6×10^{-9}.

Both p-values lie far below $\alpha = 0.05$. The significance conclusion is therefore robust to the choice of effective M. The naïve $M = 14$ value serves as a conservative upper bound; the M_{eff} value reflects the empirical correlation structure of the grid. We retain the strict $M = 14$ figure in headline

 doi:10.20944/preprints202606.0436.v1

reporting throughout the paper. The plateau-mean quarterly Sharpe (≈ 0.700 across the 12 clustered configurations) yields a similarly significant p-value under the same correction.

Note on sampling frequency: The Sharpe values reported in this subsection are computed at quarterly frequency on the 72-quarter sample, to match the rebalancing frequency of the strategy. These differ numerically from the annual-frequency Sharpe values reported in Table 1, which are computed on 18 annual observations to align with the annual-based investor-experience metrics (NegYears, MCNYR, TUW). The qualitative finding of plateau homogeneity is invariant to the sampling-frequency choice. Under quarterly Sharpe, the in-sample winner is $k = 2.5$; under annual Sharpe, the configuration with the highest single-point Sharpe is $k = 2.0$. Both lie within the plateau region defined by the IEV vector.

7.4.4. Joint Interpretation

The four findings of this subsection are mutually consistent and jointly support the region-based interpretation of strategy parameters developed in Section 8:

1. The 14-configuration grid carries $M_{\text{eff}} = 1.09$ effective independent configurations.
2. PBO = 0.9329 follows from (1) under the maximum-selection / regression-to-the-mean mechanism described above.
3. The plateau-level performance is significant at $p \leq 3.4 \times 10^{-8}$ under the strictest Bonferroni correction, and at $p \leq 2.6 \times 10^{-9}$ under the eigenvalue-spectrum-aware correction.
4. The eigenvalue collapse $M_{\text{eff}} = 1.09$ persists in an unconditional-classifier variant (Appendix B), establishing it as a structural property of the ATR-multiplier strategy class on the NASDAQ-100 universe rather than an artefact of regime conditioning.

The defensible empirical claim is therefore not "$k = 2.5$ is the optimal configuration" — that claim is contradicted by (1) and (2) — but rather: the strategy as a whole, instantiated by any configuration on the plateau, produces a Sharpe ratio statistically distinguishable from zero by margins surviving multiple-comparisons correction at the most conservative effective-M assumption. The eigenvalue spectrum **and the persistence of that spectrum across the regime-classifier ablation** establish that the ATR multiplier is a near-redundant parameter for this strategy class within the available sample. The grid identifies a **region** of equivalent performance; within that region, the choice of k is below the resolution of the available data.

7.5. Per-Year Annual Returns

Figure 4 displays the per-year YoY net-of-tax returns as a heatmap, with configurations on the vertical axis and calendar years on the horizontal axis. The plateau structure is visible as horizontal banding (the 12 plateau configurations produce nearly identical year-by-year returns), and the two structural stress periods are visible as vertical dark-red stripes (2008 and 2021). The annotated cell values confirm the magnitudes reported in Sections 7.2 and 7.3.

 doi:10.20944/preprints202606.0436.v1

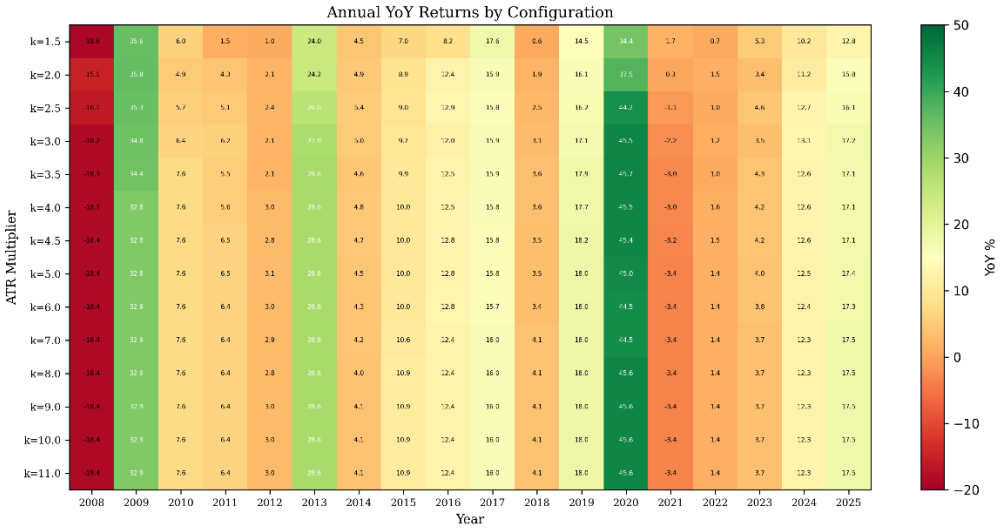

Figure 4. Per-year YoY net-of-tax returns (Dutch progressive tax applied) across the 14 ATR-multiplier configurations, 2008–2025. Diverging colour scale: deep red < −15%, white ≈ 0%, deep green > +30%. Horizontal banding indicates the plateau; vertical red columns (2008 and 2021) indicate the structurally consistent stress periods.

8. Discussion

8.1. Implications of the Bias Taxonomy

The three-class taxonomy in Section 3 provides a structured vocabulary for identifying specific filtration violations in backtest pipelines. Each class is defined by reference to a distinct locus of leakage — universe membership (Class I), input price series (Class II), within-day stop computation (Class III) — and each can be independently verified or violated. We anticipate two practical uses of the taxonomy: (a) as a checklist for practitioners reviewing their own backtesting infrastructure, and (b) as a vocabulary for reviewers requesting that authors of empirical studies report compliance with each class explicitly.

8.2. Implications of the ATR Ratcheting Lemma

Lemma 1 is structurally simple but has practical consequences. The monotonicity property means that the ATR trailing stop provides monotonically non-decreasing downside protection during any open trade, in contrast to adaptive volatility-based stops (e.g., GARCH-derived stops) which can move in either direction. The ratcheting property is therefore not just a convenience but a structural guarantee that the per-trade loss is bounded by $P_{\text{entry}} - S_{\text{final}}$. We have made no claim about which stop type produces better terminal performance; the choice involves a trade-off between the ratcheting guarantee of ATR stops and the volatility-adaptiveness of GARCH stops, and the empirical answer is likely to depend on the dominant volatility regime in the sample.

8.3. Implications of Sharpe-MCNYR Independence

Theorem 2 establishes that Sharpe and MCNYR are mathematically independent. The empirical results in Section 7 demonstrate that this independence is realised in a practical strategy setting: the 14-configuration grid produces Sharpe variation from 0.7411 to 0.8305 while MCNYR is uniformly 1. This finding has direct implications for performance reporting. The literature on investor capital withdrawal [31–33] suggests that consecutive losses can trigger irreversible behavioural responses (forced withdrawals, strategy abandonment) that are not captured by Sharpe-based performance measures. Reporting MCNYR, TUW, and rolling Calmar alongside the standard set therefore provides information genuinely orthogonal to Sharpe, not merely a redundant summary.

 doi:10.20944/preprints202606.0436.v1

8.4. Interpretation of PBO = 0.9329

The PBO value is striking and warrants explicit interpretation. PBO = 0.9329 does not indicate that the strategy is overfit in the sense of being economically vacuous. It indicates that the identity of the optimal configuration within a clustered plateau cannot be reliably identified from a 72-quarter sample. This is consistent with the Bonferroni-corrected significance of the plateau Sharpe ($p \ll 0.05$): the plateau exists and is statistically distinguishable from zero, but its internal structure is below the noise threshold of the available sample.

The methodologically appropriate response is not to discard the plateau but to acknowledge that within-plateau selection is meaningless and to defend the plateau as a region. The question "which is the best configuration?" is one the data cannot answer; and the "is the strategy real?" question is one for which the joint evidence of plateau homogeneity, structural-loss consistency, and Bonferroni-significant Sharpe provides a defensible affirmative answer.

8.5. Limitations

Several limitations apply. First, our empirical study covers only the NASDAQ-100 universe over the 2008–2025 sample. The structural findings — the eigenvalue collapse $M_{\text{eff}} = 1.09$, the plateau structure across $k \in [2.5,11.0]$, the uniform $\text{MCNYR} = 1$ — are reported for this combination of strategy, universe, and period and may not generalise to small-cap (e.g., Russell 2000), international, or single-stock universes, where liquidity, commission, slippage, and within-universe correlation structure differ materially. Generalisation requires separate empirical validation.

Second, the PBO analysis is performed on a grid of 14 configurations with 72 quarterly observations, providing approximately 5.1 observations per configuration. This is below the heuristic threshold of 20 observations per configuration commonly used for stable Sharpe estimation across grid searches. A larger sample (e.g., daily returns producing 4,500 observations) would reduce PBO substantially; we adopt quarterly returns for consistency with the rebalancing frequency of the strategy. The reported PBO should be interpreted considering this sample-adequacy constraint.

Third, the Dutch tax treatment is jurisdiction specific. Investors in other jurisdictions, or those holding instruments via tax-deferred vehicles, will obtain different net returns. We report net-of-tax figures because they correspond to the actual cash flows experienced by the investor and therefore directly populate the investor-experience metrics in Section 5. Pre-tax CAGR figures are typically higher than the reported values by approximately 4–6 percentage points depending on the configuration.

Fourth, the strategy as described in Section 6 omits specific component details — threshold values, exact entry-rule specification, and the macro regime classifier construction — for commercial reasons. The category-level description is sufficient to characterise the type of strategy under study but not to reproduce its specific behaviour from scratch. We acknowledge the trade-off: rigorous reproducibility would require full disclosure. The supplementary CSV outputs and figures contain sufficient detail for independent verification of the results even where the engine producing them is not released.

Fifth, the Class II (back-adjusted price series) concern is acknowledged but not fully mitigated. A future revision incorporating raw unadjusted prices with explicit corporate-action handling would strengthen the bias-compliance claim.

Sixth, the Class III empirical magnitude on this specific strategy was investigated separately and found to be configuration-dependent and negligible (below 0.1 percentage points of CAGR) on the period-by-period engine at quarterly rebalancing. The bias remains a theoretical concern (Definition 4) and is documented to motivate vigilance in vectorised stop-computation implementations; we do not claim that all systematic equity strategies are equally robust to Class III violations.

9. Conclusions

 doi:10.20944/preprints202606.0436.v1

We have presented a formal taxonomy of look-ahead bias for systematic equity backtesting, organised around the natural-filtration condition that defines point-in-time correctness. Three bias classes — universe-membership contamination, price-data forward leakage, and stop-exit sequencing violations — are characterised as distinct types of filtration violation and discussed with mitigation strategies. We have formalised the ATR trailing stop as a stochastic recurrence and recorded its monotonic non-decreasing ratcheting property (Lemma 1), providing a structural guarantee on per-trade loss. We have exhibited a closed-form construction (Theorem 2) demonstrating that Sharpe ratio and maximum consecutive negative-year run are mathematically independent, providing theoretical grounding for the treatment of investor-experience metrics as first-class objectives.

An 18-year empirical study on the NASDAQ-100 with point-in-time index constituency reconstruction and an acknowledged residual Class II exposure (Section 6.3.1), evaluating 14 ATR-multiplier configurations under combinatorially symmetric cross-validation, yields three principal findings. First, the strategy exhibits a stop-multiplier-insensitive Sharpe plateau $k \in [2.5, 11.0]$, with 12 configurations producing annualised Sharpe ratios within a 0.025 band (mean 0.805) and CAGR within a 0.20 percentage-point band (mean 10.34% net of Dutch progressive tax). Second, MCNYR is uniformly equal to 1 across the entire grid, demonstrating empirically the Sharpe–MCNYR independence formalised in Theorem 2. Third, the structurally consistent loss periods (2008 GFC and 2021) and time-under-water years (2021–2022) across the plateau configurations argue that the strategy's underperformance is regime-driven rather than parameter-tuned.

The Probability of Backtest Overfitting (PBO = 0.9329) confirms that point-identification within the plateau is statistically unstable, while the Bonferroni-corrected significance of the plateau Sharpe ($p \ll 0.05$) supports the plateau's existence as a genuine economic phenomenon. The combined evidence argues for region-based parameter defence in preference to single-point optimisation, and for the joint reporting of Sharpe-based and investor-experience metrics in systematic strategy evaluation.

Scope of the empirical findings: The empirical results are reported for a single strategy class on the NASDAQ-100 universe over the 2008–2025 sample, under Dutch progressive taxation. The eigenvalue collapse and the structural plateau are properties of this specific strategy-universe-period combination. Whether the near-redundancy of the ATR multiplier generalises to other equity universes (Russell 2000, international developed, emerging markets), to fundamentally different strategy classes (mean-reversion, breakout, multi-asset), or to longer sample periods that span additional macroeconomic regimes is an empirical question this paper does not resolve. The methodological framework — the bias taxonomy, the ATR ratcheting Lemma, the Sharpe-MCNYR independence theorem, and the Bonferroni-Li-Ji significance procedure — applies without modification to any such extension.

Author Contributions: All stages of the paper have been prepared by the author.

Funding: This research received no external funding.

Data Availability Statement: Reproducibility is stratified. The methodological framework of the paper (Sections 3, 4, 5, 6.6, 6.7) is fully specified and independently reimplementable. The empirical headline numbers (Sections 7, Appendix B) are not independently reproducible without three components withheld for commercial reasons: the trend-confirmation filter and entry-rule specifics, the macro regime classifier, and the strategy engine source code. What can be verified from supplementary materials is the computation of the headline statistics from the consolidated return series. The following are provided as supplementary materials: consolidated CSV outputs underlying Tables 1, 2, and 3 for both conditional and unconditional grids; verification scripts that reproduce the headline statistical analyses (Li-Ji effective M, PBO, Bonferroni significance, Theorem 2 numerical verification); and figure-generation scripts producing Figures 1–4. Underlying tick-level intraday data is commercially licensed and cannot be redistributed. Reconstructed point-in-time NASDAQ-100 constituency datasets and per-trade backtest logs are available from the author on reasonable request.

 doi:10.20944/preprints202606.0436.v1

Conflicts of Interest: The author declares no conflicts of interest. No external entity has funded or directed this research.

Appendix A. Numerical Verification of Theorem 2

The closed-form construction of Theorem 2 (Section 5.2) produces, for any target (μ, σ, T) and any target MCNYR $k \in \{1,2,\dots,T-1\}$ satisfying $d > 0$, a deterministic return sequence r^B with empirical mean μ, empirical standard deviation σ (population, ddof = 0), and $\mathrm{MCNYR}(r^B) = k$. The Sharpe ratio of r^B is therefore μ/σ for every k, independent of the prescribed maximum consecutive negative-year run.

We verify the construction numerically. Target parameters are fixed at $\mu = 0.10$, $\sigma = 0.20$, $T = 18$ (matching the empirical sample length used in Sections 6–7). For each $k \in \{1,2,\dots,9\}$, we apply the construction of Theorem 2 to produce $r^B(k)$, then compute four diagnostics directly from the resulting 18-period sequence: empirical mean, empirical standard deviation (population), Sharpe ratio (mean/std), and observed MCNYR. The condition $d > 0$(equivalently $\mu < \sigma\sqrt{q/p}$) is satisfied for all $k \in \{1,\dots,9\}$ at these target parameters; the upper limit $k_{\max}(T,\mu,\sigma) = 12$ is computed but not exercised, as values approaching $k_{\max}$ produce visually identical sequences with $g \to 0^+$ that are less instructive.

Table 2. Numerical verification of Theorem 2 at target parameters $(\mu, \sigma, T) = (0.10, 0.20, 18)$ across $k = 1,\dots,9$. The constructed sequence $r^B(k)$ has k negative entries of value $-d$ and $T-k$ positive entries of value g; both d and g vary materially across k, but the empirical mean, standard deviation, and Sharpe ratio are invariant by construction. Observed MCNYR matches the prescribed k exactly.

k	$-d$(neg-period value)	g(pos-period value)	$\mathbb{E}[r^B]$	Std(r^B)	Sharpe(r^B)	MCNYR(r^B)
1	−0.7246	0.1485	0.10000	0.20000	0.5000	1
2	−0.4657	0.1707	0.10000	0.20000	0.5000	2
3	−0.3472	0.1894	0.10000	0.20000	0.5000	3
4	−0.2742	0.2069	0.10000	0.20000	0.5000	4
5	−0.2225	0.2240	0.10000	0.20000	0.5000	5
6	−0.1828	0.2414	0.10000	0.20000	0.5000	6
7	−0.1507	0.2595	0.10000	0.20000	0.5000	7
8	−0.1236	0.2789	0.10000	0.20000	0.5000	8
9	−0.1000	0.3000	0.10000	0.20000	0.5000	9

Appendix B. Unconditional-Classifier Variant

The empirical results in the main body are produced with a macro regime classifier modulating position sizing across all 14 configurations (Section 6.2). Because the classifier specifics are commercially confidential, the question arises whether the principal findings — the eigenvalue collapse, the plateau structure, and the structurally consistent loss periods — are properties of the strategy class or artefacts of the regime conditioning. This appendix reports a parallel grid run in which the classifier output is held constant at maximum exposure across all decision dates (the "unconditional" variant), with all other parameters identical.

B.1. Engineering

 doi:10.20944/preprints202606.0436.v1

The classifier-disable mechanism is implemented as a runtime flag in the backtest engine that bypasses classifier evaluation; the classifier module itself is unmodified. The 14 grid configurations are otherwise identical to the main study (same point-in-time roster, same data, same costs, same tax treatment, same engine).

B.2. Investor-Experience Vector

The unconditional grid IEV vector is reported as **Table 3** alongside the conditional baseline. Annualised CAGR is uniformly reduced across the grid by 0.55–1.13 percentage points, with the plateau configurations ($k = 2.5$ through $k = 11.0$) clustering tightly around a 1.0–1.1 pp penalty. Annual-frequency Sharpe drops by 0.07–0.12 across the grid. MCNYR remains uniformly 1 in all 14 unconditional configurations, matching the conditional grid exactly. The CAGR penalty is approximately uniform across $k$in the plateau region (range -0.95 to -1.13 for $2.5 \leq k \leq 11.0$), suggesting the classifier acts as a near-constant exposure modulator rather than interacting with the stop-multiplier choice. The wide-stop saturation ($k = 9.0,10.0,11.0$) produces identical IEV values in both grids — the trailing stop effectively never fires within a quarter at these multipliers, removing k as an active parameter — and the penalty is the same (-1.13 pp CAGR, -0.0731 Sharpe) across all three rows.

Table 3. Conditional vs. unconditional 14-configuration ATR-multiplier grid (NASDAQ-100 point-in-time universe, 2008–2025, net of Dutch progressive tax). The conditional grid uses the macro regime classifier of Section 6.2; the unconditional grid disables the classifier (constant maximum exposure across all dates). Sharpe is reported at annual frequency (18 observations, population standard deviation), matching Table 1. Δ-CAGR is in percentage points; Δ-Sharpe in absolute units.

k	**CAGR cond**	**CAGR uncond**	∇-CAGR (pp)	Sharpe cond	Sharpe uncond	∇-Sharpe	MCNYR cond	MCNYR uncond
1.5	8.56%	8.01%	−0.55	0.7411	0.6445	−0.0966	1	1
2.0	9.65%	8.60%	−1.05	0.8305	0.7093	−0.1212	1	1
2.5	10.19%	9.24%	−0.95	0.8153	0.7236	−0.0917	1	1
3.0	10.24%	9.20%	−1.04	0.7907	0.7131	−0.0776	1	1
3.5	10.35%	9.30%	−1.05	0.7945	0.7217	−0.0728	1	1
4.0	10.35%	9.33%	−1.02	0.8073	0.7405	−0.0668	1	1
4.5	10.39%	9.32%	−1.07	0.8086	0.7380	−0.0706	1	1
5.0	10.36%	9.34%	−1.02	0.8081	0.7408	−0.0673	1	1
6.0	10.28%	9.29%	−0.99	0.8046	0.7351	−0.0695	1	1
7.0	10.33%	9.26%	−1.07	0.8090	0.7316	−0.0774	1	1
8.0	10.38%	9.26%	−1.12	0.8044	0.7324	−0.0720	1	1
9.0	10.39%	9.26%	−1.13	0.8055	0.7324	−0.0731	1	1
10.0	10.39%	9.26%	−1.13	0.8055	0.7324	−0.0731	1	1
11.0	10.39%	9.26%	−1.13	0.8055	0.7324	−0.0731	1	1

B.3. Eigenvalue Spectrum and Effective Tests

 doi:10.20944/preprints202606.0436.v1

The correlation-matrix eigenvalue spectrum of the unconditional grid is essentially identical to the conditional grid: $\lambda_1 = 13.9091$ (99.35% of variance, versus 13.9097 / 99.36% conditional). The Li–Ji effective number of independent tests is $M_{\text{eff}} = 1.091$(versus 1.090 conditional). The eigenvalue collapse is therefore structural to the ATR-multiplier strategy class in this universe, independent of the regime-conditioning module.

B.4. PBO and Significance

PBO under CSCV in the unconditional grid is 0.7857, compared with 0.9329 in the conditional grid. This decline is informative: PBO depends on both grid effective dimensionality (here unchanged) and signal-to-noise ratio (here reduced through removal of the conditioning module). Under weaker signal, the in-sample winner's noise advantage is less reliably reproduced across the in-sample / out-of-sample split, producing more partitions in which the in-sample winner happens to retain above-median rank out-of-sample by chance. PBO and M_{eff} therefore measure related but distinct properties: M_{eff} captures cross-sectional near-degeneracy among configurations, while PBO captures the temporal reproducibility of maximum-Sharpe ranking. Both support the region-based interpretation of the strategy parameter.

The in-sample winner under quarterly Sharpe in the unconditional grid is also $k = 2.5$ (quarterly Sharpe 0.5619, versus 0.7081 conditional), yielding a t-statistic of 4.74 and Bonferroni-corrected p-value of 3.1×10^{-5}at $M = 14$ or 2.4×10^{-6}at $M_{\text{eff}} = 1.09$. Both lie well below $\alpha = 0.05$; the unconditional plateau is statistically significant, but at a notably weaker margin than the conditional plateau. The classifier contributes approximately one order of magnitude of additional significance margin while leaving the plateau's existence and structural properties intact.

B.5. Interpretation

The unconditional variant establishes three things:

1. The plateau structure of the strategy is not an artefact of the regime classifier. The **identity** of the apparent best-Sharpe configuration shifts between grids ($k = 2.0$ conditional, $k = 5.0$ unconditional, both within the plateau), but the existence and extent of the plateau are unchanged. This is consistent with Section 7.4's interpretation that within-plateau point identification is below the resolution of the available sample.
2. The eigenvalue collapse $M_{\text{eff}} = 1.09$ is structural to the ATR-multiplier strategy class on the NASDAQ-100 universe.
3. The regime classifier provides additional value (≈1.0 pp CAGR, ≈0.07 annual Sharpe) without altering the qualitative grid structure.

The first two findings allow the methodological claims of the paper to be evaluated and reproduced without disclosure of the classifier specifics, even though numerical reproduction of the headline conditional results would additionally require the classifier. The third finding constrains the magnitude of the gap between the conditional and any independently reimplemented unconditional baseline.

Appendix C. Correlation-Matrix Eigenvalue Spectrum of the 14-Configuration Grid

The collapse onto a single dominant eigenvalue is the central empirical justification for the region-based interpretation of strategy parameters developed in Sections 7.4 and 8.4: variation in $k$across the operative range produces near-identical statistical objects from the perspective of the available 72-quarter sample.

Table 4. Eigenvalues of the 14×14 correlation matrix of quarterly net-of-tax returns across the 14 ATR-multiplier configurations on the NASDAQ-100 sample, 2008–2025 (72 quarterly observations per configuration). Eigenvalues are reported in descending order, with individual and cumulative variance shares relative to the

 doi:10.20944/preprints202606.0436.v1

trace ($\sum_i \lambda_i = 14$). The dominant eigenvalue absorbs 99.36% of total variance; the remaining 13 eigenvalues collectively carry 0.64%. The effective number of independent tests, computed via the Li–Ji method [38], is $M_{\text{eff}} = 1.09$.

i	$\boldsymbol{\lambda_i}$	$\boldsymbol{\lambda_i/14}$	**Cumulative**
1	13.9097	99.36%	99.36%
2	0.0735	0.525%	99.88%
3	0.0083	0.059%	99.94%
4	0.0046	0.033%	99.97%
5	0.0016	0.011%	99.98%
6	0.0012	0.009%	99.99%
7	0.0005	0.004%	100.00%
8	0.0004	0.003%	100.00%
9	0.0001	0.001%	100.00%
10-14	$< 10^{-4}$	$< 10^{-4}$	100.00%

Abbreviations

The following abbreviations are used in this manuscript:

ATR	Average True Range
CAGR	Compound Annual Growth Rate
CSCV	Combinatorially Symmetric Cross-Validation
DSR	Deflated Sharpe Ratio
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
IEV	Investor Experience Vector
MCNYR	Maximum Consecutive Negative Year Run
NDX	NASDAQ-100 Index
PBO	Probability of Backtest Overfitting
PIT	Point-in-Time
SMA	Simple Moving Average
TUW	Time-Under-Water
VIX	CBOE Volatility Index

References

1. N. Jegadeesh, and S. Titman, "Returns to buying winners and selling losers: Implications for Stock Market Efficiency.," *The Journal of Finance,* vol. 48, no. 1, pp. 65-91, 1993.
2. E. F. Fama, and K. R. French, "Multifactor explanations of asset pricing anomalies," *The journal of finance,* vol. 51, no. 1, pp. 55-84, 1996.
3. K. G. Rouwenhorst, "International momentum strategies," *The journal of finance,* vol. 53, no. 1, pp. 267-284, 1998.
4. C. S. Asness, T. J. Moskowitz, and L. H. Pedersen, "Value and momentum everywhere," *The journal of finance,* vol. 68, no. 3, pp. 929-985, 2013.
5. C. Geczy, and M. Samonov, "Two centuries of multi-asset momentum (equities, bonds, currencies, commodities, sectors and stocks)," *Available at SSRN 2607730*, 2017.
6. G. Antonacci, *Dual Momentum Investing*: McGraw-Hill Education, 2014.
7. C. Wilcox, and E. Crittenden, "Does trend following work on stocks?," *The Technical Analyst,* vol. 14, pp. 1-19, 2005.

 doi:10.20944/preprints202606.0436.v1

8. D. Bailey, J. Borwein, M. Lopez de Prado, and Q. J. Zhu, "The probability of backtest overfitting," *The Journal of Computational Finance,* vol. 20, no. 4, pp. 39-69, 2017.
9. D. H. Bailey, and M. López de Prado, "The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting and non-normality," *Journal of Portfolio Management,* vol. 40, no. 5, pp. 94-107, 2014.
10. D. H. Bailey, J. M. Borwein, M. L. De Prado, and Q. J. Zhu, "Pseudomathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance," *Notices of the AMS,* vol. 61, no. 5, pp. 458-471, 2014.
11. M. L. De Prado, *Advances in financial machine learning*: John Wiley & Sons, 2018.
12. J. W. Wilder, *New concepts in technical trading systems*: Greensboro, N.C.: Trend Research, 1978.
13. M. M. Carhart, "On persistence in mutual fund performance," *The Journal of finance,* vol. 52, no. 1, pp. 57-82, 1997.
14. T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen, "Time series momentum," *Journal of financial economics,* vol. 104, no. 2, pp. 228-250, 2012.
15. B. Hurst, Y. H. Ooi, and L. H. Pedersen, "A century of evidence on trend-following investing," *Available at SSRN 2993026*, 2017.
16. B. Hurst, Y. H. Ooi, and L. H. Pedersen, "Demystifying managed futures," *Journal of Investment Management,* vol. 11, no. 3, pp. 42-58, 2013.
17. C. R. Harvey, Y. Liu, and H. Zhu, "… and the cross-section of expected returns," *The Review of financial studies,* vol. 29, no. 1, pp. 5-68, 2016.
18. S. J. Brown, W. Goetzmann, R. G. Ibbotson, and S. A. Ross, "Survivorship bias in performance studies," *The Review of Financial Studies,* vol. 5, no. 4, pp. 553-580, 1992.
19. J. Henker, T. Henker, and T. D. Huynh, "Survivorship bias and alternative explanations of momentum effect." pp. 1-43.
20. J. Duarte, C. S. Jones, M. Khorram, H. Mo, and J. L. Wang, "Too good to be true: Look-ahead bias in empirical options research," *Review of Financial Studies*, 2023.
21. Z. Gao, W. Jiang, and Y. Yan, "A Test of Lookahead Bias in LLM Forecasts," *arXiv preprint arXiv:2512.23847*, 2025.
22. A. Sehgal, "R&D Alpha: Investment Intensity and Long-Term Stock Returns," *Available at SSRN 6002295*, 2026.
23. R. H. Battalio, C. W. Holden, M. Pierson, J. J. Shim, and J. Wu, "Latency and the look-ahead bias in trade and quote data," *Available at SSRN 5907665*, 2026.
24. M. Benhenda, "Look-Ahead-Bench: a Standardized Benchmark of Look-ahead Bias in Point-in-Time LLMs for Finance," *arXiv preprint arXiv:2601.13770*, 2026.
25. S. Vervoort, "Average True Range Trailing Stops," *Stocks and Commodities,* vol. 27, no. 6, pp. 34-40, 2009.
26. N. Li, A. Laryea, and Y. Ihlamur, "Optimal Stop-Loss and Take-Profit Parameterization for Autonomous Trading Agent Swarm," *arXiv preprint arXiv:2604.27150*, 2026.
27. A. Arratia, and A. Dorador, "On the Effectiveness of Stop-Loss Rules: An Analytical Framework Based on Modeling Overnight Gaps and the Stationary Bootstrap," *Available at SSRN 3087196*, 2017.
28. V. Zakamulin, "A comprehensive look at the empirical performance of moving average trading strategies," *Available at SSRN 2677212*, 2015.
29. W. F. Sharpe, "The sharpe ratio," *Streetwise–the Best of the Journal of Portfolio Management,* vol. 3, no. 3, pp. 169-85, 1998.
30. T. W. Young, "Calmar ratio: A smoother tool," *Futures,* vol. 20, no. 1, pp. 40, 1991.
31. J. B. Berk, and R. C. Green, "Mutual fund flows and performance in rational markets," *Journal of political economy,* vol. 112, no. 6, pp. 1269-1295, 2004.
32. V. Agarwal, and N. Y. Naik, "Risks and portfolio decisions involving hedge funds," *The Review of Financial Studies,* vol. 17, no. 1, pp. 63-98, 2004.
33. T. Sapp, and A. Tiwari, "Does stock return momentum explain the "smart money" effect?," *The Journal of Finance,* vol. 59, no. 6, pp. 2605-2622, 2004.
34. F. M. Kinniry Jr, C. M. Jaconetti, M. A. DiJoseph, Y. Zilbering, and D. G. Bennyhoff, "Putting a value on your value: Quantifying Vanguard advisor's alpha. Vanguard Research," 2014.

 doi:10.20944/preprints202606.0436.v1

35. J. H. Kim, A. Shamsuddin, and K.-P. Lim, "Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data," *Journal of empirical finance,* vol. 18, no. 5, pp. 868-879, 2011.
36. S. J. Larson, and J. Madura, "What drives stock price behavior following extreme one-day returns," *Journal of Financial Research,* vol. 26, no. 1, pp. 113-127, 2003.
37. D. Blitz, and P. Van Vliet, "The volatility effect: Lower risk without lower return," *Journal of portfolio management*, pp. 102-113, 2007.
38. J. Li, and L. Ji, "Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix," *Heredity,* vol. 95, no. 3, pp. 221-227, 2005.