



Master's Thesis

Volatility-Gated Momentum: An EGARCH-Filtered Intraday Trading Framework for Equity Index and Energy Futures

June 2026 thesis; editorial revision 14 October 2026

Portfolio edition—not a university resubmission or replacement of the original submitted thesis

Volatility-Gated Momentum

An EGARCH-Filtered Intraday Trading Framework for Equity Index and Energy Futures

Woosub Shin

Department of Economics, University of Copenhagen
M.Sc. in Economics

Advisor: Jacob Lundbeck Serup · June 2026

This thesis studies a long-only breakout framework for Nasdaq 100, S&P 500, and WTI Crude Oil futures using 5-minute data over 2019–2025. EGARCH-derived variables govern trade eligibility and position size. The reported development specification has a Sharpe ratio of 1.412, compared with 0.664 after removing the combined volatility layer. A retrospective exercise selecting trading parameters on 2010–2018 reports a later-period Sharpe of 1.195. These are historical research results: development exposure, future-informed volatility calibration, and unresolved accounting differences limit claims about prospective performance.

Abstract

This thesis investigates whether asymmetric conditional volatility can serve as an admissibility rule for whether an intraday breakout should be traded at all. The hypothesis is that EGARCH, used as a conditioning layer rather than a directional forecast, can identify volatility states in which intraday long breakout signals are more likely to carry directional information.

The empirical analysis uses 5-minute data for Nasdaq 100, S&P 500, and WTI Crude Oil futures over 2019–2025. The development specification reports a daily Sharpe ratio of 1.412 and terminal equity of 3.88 times the initial \$200,000. Removing the combined volatility gate, score, inverse-volatility sizing, and regime caps lowers the reported Sharpe to 0.664. A paired circular block bootstrap of this development comparison reports a nonpositive-tail fraction of 0.0038. Retrospective parameter selection on 2010–2018 produces a reported 2019–2025 Sharpe of 1.195, versus 0.382 for its ablation; the development bootstrap does not test that separate comparison. Timing shifts and frequency-preserving randomizations provide additional within-design sensitivity evidence, but do not identify the independent effect of EGARCH asymmetry.

The evidence supports further investigation of volatility-conditioned breakout trading, while leaving several questions unresolved. The 2019–2025 window was exposed to strategy development; full-sample volatility ranks use future observations; and separate output pipelines disagree on some PnL totals and observation counts. Historical parameter restrictions therefore do not establish an untouched holdout. This revision preserves reported estimates, distinguishes empirical associations from identified mechanisms, and specifies the reconciliation and prospective tests needed before the results can support implementation claims.

Revision status

Portfolio edition—not a university resubmission or replacement of the original submitted thesis

This edition revises the source corresponding to the publicly circulated 82-page thesis. It corrects prose, mathematical definitions, references, and the scope of statistical claims without rerunning models or changing unverified empirical outputs. Unreconciled numerical differences are identified beside the affected results and collected in Appendix I. The original thesis and research files remain unchanged.

Acknowledgments

I would like to thank my advisor, Jacob Lundbeck Serup, for his guidance throughout this project. I also acknowledge the use of Claude (Anthropic) and ChatGPT (OpenAI) for coding support, debugging, and research brainstorming during the development of the trading framework and robustness testing pipeline. Responsibility for the original thesis rests with the author. The October 2026 editorial revision was prepared with OpenAI Codex assistance; the revision does not imply that the author or institution has approved its final wording.

Table of Contents

Abstract	iii
Acknowledgments	iv
1 Introduction	1
2 Literature Review	3
2.1 Volatility Modeling: From Estimation to Application	3
2.2 Momentum and Breakout Strategies: Unconditional Strength, Conditional Fragility . . .	3
2.3 Volatility Scaling versus Volatility Conditioning	4
2.4 Positioning and Hypothesis	5
3 Data	6
3.1 Sample and Asset Selection	6
3.2 Intraday Frequency and Session Restriction	6
3.3 Data Preparation and Construction of Variables	7
3.4 Data Integrity and Cleaning Procedure	8
3.5 Descriptive Statistics	9
3.6 Price Dynamics and Cross-Asset Structure	9
3.7 Implications for the Empirical Design	11
4 Methodology	12
4.1 Data Construction and Sample Design	12
4.2 EGARCH Volatility Modeling	12
4.3 Volatility Expansion Gate and Continuous Volatility Score	16
4.4 Volatility Regime Classification	18
4.5 Intraday Signal Construction	18
4.6 Portfolio Construction and Position Management	19
4.7 Performance Measurement	21
4.8 Model Comparison and Ablation Design	22

4.9	Statistical Inference: Circular Block Bootstrap	23
4.10	Complete Parameter Specification	24
5	Empirical Results	26
5.1	Development-Enhanced Specification	26
5.2	The EGARCH Conditioning Effect: Ablation Evidence	27
5.3	Asset-Level Contribution and Trade Activity	28
5.4	Trade Quality by Volatility Regime	30
5.5	Year-by-Year Analysis and Subperiod Performance	31
5.6	Transaction Cost Sensitivity	33
5.7	Factor Exposure Analysis	34
5.8	Annual Summary and Calibration Provenance	36
5.9	EGARCH versus Alternative Volatility Filters	36
5.10	Placebo and Falsification Tests	37
5.11	Return Distribution and the Sharpe–Sortino Gap	40
5.12	Parameter Stability	42
5.13	Deflated Sharpe Ratio	45
5.14	Retrospective Temporal Split	47
5.15	Retrospective Pre-2019 Parameter Selection	47
5.16	Rolling Parameter Re-Selection	48
5.17	Interpretation of the Main Findings	50
6	Discussion	51
6.1	The Economic Mechanism: Why Expanding Volatility Improves Breakout Quality . . .	51
6.2	Why Asymmetry Matters for a Long-Only Framework	52
6.3	The Conditional Nature of the Edge	53
6.4	The Role of Crude Oil	54
6.5	The Equity Tailwind	54
6.6	Limitations	55
6.7	Future Research	56

7	Conclusion	57
A	Data Integrity and Cleaning Checks	59
B	Additional Strategy Variants	60
C	Cost Sensitivity by Asset	61
D	Supplementary Figures	62
E	CL Direction Restriction Analysis	63
F	Portfolio Composition: NQ+ES Only vs. Three-Asset	64
G	Contract-Count Scaling Sensitivity	65
H	Trading Rule Summary	67
I	Reconciliation and Reproducibility Register	69
	References	71

List of Figures

1	Normalized price series for NQ, ES, and CL (2019–2025)	10
2	Average intraday activity pattern based on normalized volume	11
3	Equity curves of the final model and No-EGARCH specification	27
4	Drawdown comparison of the final model and No-EGARCH specification	28
5	Cumulative net PnL contribution by asset	29
6	EGARCH volatility gate and trade entries for NQ during 2021	31
7	Year-by-year returns of the final three-asset model (2019–2025)	33
8	Gate-timing placebo test	39
9	Daily return distribution and normal Q-Q plot	40
10	Distribution of daily strategy returns: gain and loss days	41
11	Sharpe ratio under single-parameter perturbation	43
12	Joint sensitivity: breakout window $\times$ ATR stop multiplier	44
13	Joint sensitivity: EMA fast $\times$ EMA slow	44
14	Joint sensitivity: RSI entry $\times$ RSI exit threshold	45
15	Equity comparison between the final three-asset model and the No-EGARCH specification.	62
16	Sharpe ratio across the selected subperiods of the final three-asset model.	62

List of Tables

1	Descriptive statistics for 5-minute log returns (2019–2025)	9
2	EGARCH(1,1) parameter estimates by futures contract	14
3	EGARCH(1,1) residual diagnostics by futures contract	16
4	Complete parameter specification of the final three-asset model	25
5	Main performance results: final three-asset model vs. no-EGARCH ablation (2019–2025)	26
6	Paired circular block bootstrap of the development Sharpe difference	28
7	Trade counts and net PnL by asset in the final three-asset model (2019–2025)	29
8	Trade-level statistics of the final three-asset model (2019–2025)	30
9	Trade quality conditional on volatility regime at entry (2019–2025)	30
10	Year-by-year performance of the final three-asset model	32
11	Cost sensitivity of the final three-asset strategy	33
12	Factor exposure regressions of strategy daily returns	35
13	Saved annual summary (chronological recalibration provenance unresolved)	36
14	Performance comparison across volatility filter specifications	37
15	Gate-timing placebo test	38
16	Randomized frequency-preserving gate placebo	39
17	Sharpe and Sortino ratios at 5-minute and daily frequencies (champion specification, 2019–2025)	41
18	Parameter stability analysis: Sharpe ratio under single-parameter perturbation	42
19	DSR formula sensitivity conditional on archived inputs	46
20	Development period vs. later-period split performance	47
21	Retrospective parameter selection: training on 2010–2018, evaluation on 2019–2025 . .	48
22	Walk-forward analysis: annual parameter re-selection on rolling 24-month windows . . .	49
23	Residual-sign diagnostic of the vol-score distribution	53
24	Intraday futures data audit results	59
25	Selected strategy variants tested during model development	60
26	Per-asset cost sensitivity of the final three-asset strategy	61
27	CL direction restriction analysis: long-only vs. short-only vs. long-and-short (2019–2025)	63
28	Portfolio composition test: NQ+ES only vs. three-asset champion (2019–2025)	64

29	Scale L sensitivity: EGARCH-filtered vs. No-EGARCH specification (2019–2025) . . .	65
30	Asset-specific parameter summary	67

1 Introduction

Intraday breakout signals do not mean the same thing in every market environment. A close above a recent high may reflect directional buying, transient liquidity conditions, or noise. The hypothesis here is that daily volatility information can help distinguish among those possibilities. This trading hypothesis is distinct from the variance predictability established by the conditional-volatility framework of Engle (1982), Bollerslev (1986), and Nelson (1991). What is less developed is the use of this observation as an explicit rule for deciding whether an individual intraday trade should be taken at all.

This thesis looks at one such rule. The basic idea is that EGARCH volatility conditioning may help an intraday breakout framework not by forecasting direction directly, but by screening for the volatility states in which breakout signals are more likely to contain directional information. It is useful to distinguish a few ways conditional volatility can enter a trading setup. In the standard econometric use, volatility is itself the object being forecast. In another common use, it enters through portfolio-level position scaling, as in Moreira and Muir (2017). Here I use it differently. The question is not how much to hold once a trade is on, but whether a given signal should be traded at all. EGARCH is well suited to that role because it responds asymmetrically to shocks (Engle and Ng, 1993; Glosten et al., 1993; Nelson, 1991), which means that not all volatility expansions are treated as the same kind of state. For a long-only strategy, that matters: a breakout that appears during stress after a negative shock is not obviously the same object as a breakout that appears during a more constructive expansion.

The empirical analysis uses 5-minute intraday data for three liquid U.S. futures contracts (Nasdaq 100, S&P 500, and WTI Crude Oil) over the 2019–2025 evaluation period. The multi-asset design tests whether the conditioning mechanism operates across both equity index and energy futures, two market segments with distinct volatility dynamics and microstructure. Entries are restricted to asset-specific core sessions (09:30–14:00 for equity indices, 09:30–13:00 for Crude Oil) when liquidity and price discovery are most concentrated (Andersen and Bollerslev, 1997; Andersen et al., 2003).

The framework combines three layers: (i) a daily EGARCH(1,1) model that defines a volatility expansion gate and a continuous volatility score governing position sizing through inverse volatility targeting; (ii) long-only breakout, RSI, and EMA entry signals with asset-specific parameterization; and (iii) ATR-based trailing stops, a cooldown mechanism, and regime-conditional portfolio caps for Crude Oil. The design is modular. Components can be varied in the simulation, but the complete-layer ablation changes both selection and exposure; their individual effects require matched component comparisons.

The empirical findings can be summarized as three claims. I list them below in what I take to be descending order of how well they are identified.

First, the reported development results associate the combined EGARCH-derived filtering and sizing layer with higher risk-adjusted performance: Sharpe falls from 1.412 to 0.664 when the layer is removed. The development paired-bootstrap tail fraction is 0.0038. Gate-timing shifts and randomized

gate assignments support sensitivity to the observed gate, conditional on the selected architecture and historical calibration. They do not establish a causal effect or correct for the research process.

Second, retrospective parameter selection on 2010–2018 reports a 2019–2025 Sharpe of 1.195. This restricts the trading-parameter search window, but the architecture and validation exercises were developed after observing later data. It is a useful sensitivity exercise, not an independently established estimate of future performance. The development Sharpe of 1.412 is a selected historical estimate, not a statistical upper bound.

Third, performance varies across market states and periods. The saved results motivate a conditional trading hypothesis but do not identify stable boundaries within which the mechanism will remain profitable. The earlier near-zero training Sharpe and the implementation limitations are central to that interpretation.

The remainder of the thesis is organized as follows. Section 2 reviews the relevant literature on volatility modeling, momentum strategies, and volatility-conditioned trading. Section 3 describes the data and sample construction. Section 4 presents the methodology. Section 5 reports the empirical results. Section 6 discusses interpretation and limitations, and Section 7 concludes.

2 Literature Review

The thesis draws on three separate literatures that are relevant here, even though they are not usually brought together in quite this way: conditional volatility modelling, momentum and breakout strategies in futures markets, and the more recent work that uses volatility information to condition directional strategies. The review follows that sequence mainly because it matches the role volatility gradually comes to play in the thesis. It starts as something to estimate, then becomes an input into portfolio-level scaling, and finally becomes part of the rule for whether individual signals are acted on at all.

2.1 Volatility Modeling: From Estimation to Application

The ARCH framework of Engle (1982) and its GARCH generalization (Bollerslev, 1986) established that return variance is predictable, persistent, and clustered. Nelson's (1991) EGARCH specification introduced two properties central to the present thesis: logarithmic variance modeling, which guarantees positivity without imposing nonnegativity constraints on the variance coefficients, and asymmetric shock response, which allows negative innovations to generate larger volatility increases than positive ones. Glosten et al. (1993) and Engle and Ng (1993) confirmed this asymmetry empirically across equity markets, while Kang et al. (2009) documents persistent crude-oil volatility through comparisons of GARCH-family models. These references do not by themselves identify the economic source of CL asymmetry.

An asymmetric variance recursion distinguishes innovations by sign as well as magnitude. Whether that statistical distinction predicts the success of a subsequent long breakout is a separate hypothesis. The thesis does not observe risk premia, information arrival, or forced order flow directly.

Most applications of GARCH-family models continue to focus on variance forecasting, risk management, and option pricing, where the object of interest is the volatility estimate itself. Rather less attention has been given to a different question: whether conditional volatility can be used to decide not how much to trade but whether a particular trade should be taken at all. This thesis takes that question as its starting point.

2.2 Momentum and Breakout Strategies: Unconditional Strength, Conditional Fragility

Directional momentum is among the best-documented empirical regularities in finance. Jegadeesh and Titman (1993) established intermediate-horizon return continuation in equities; Moskowitz et al. (2012) extended the finding to futures across asset classes, with Szakmary et al. (2010) providing complementary evidence on trend-following profitability in commodity futures specifically; and Asness et al. (2013) showed that momentum profits reflect a common factor structure. Breakout rules operationalize momentum through price thresholds and are naturally suited to intraday execution, where

the signal can be observed and acted upon within a single session. The methodological tradition for evaluating such rules dates back to Brock et al. (1992), with subsequent surveys by Lo et al. (2000) and Park and Irwin (2007) systematizing the evidence on which rule families generalize across markets and which do not.

The fragility of these strategies is also well documented. Daniel and Moskowitz (2016) show that momentum crashes are concentrated in high-volatility states and particularly on the short side, which implies that the quality of a momentum signal depends heavily on the volatility regime in which it is observed. A plausible reading is that, in calm markets, a breakout above a recent high often reflects little more than a transient liquidity gap, whereas in expanding volatility the same breakout is more likely to be accompanied by elevated information arrival and supportive order flow. The unconditional case for momentum is strong, but this literature also gives a direct reason to implement it conditionally on the volatility state.

The long-only restriction is a design choice informed by equity drift and by the observed behavior of the tested strategy. The crash mechanism studied by Daniel and Moskowitz (2016) concerns cross-sectional momentum portfolios; it does not establish that an intraday long-only futures rule avoids crash risk. The choice of direction is therefore part of the research search space and requires separate validation.

2.3 Volatility Scaling versus Volatility Conditioning

A growing body of work uses volatility information to improve directional strategies. Existing applications, however, operate at two distinct levels, and the distinction between them is central to positioning the present thesis.

At the *portfolio level*, Moreira and Muir (2017) show that scaling portfolio exposure inversely with recent variance can improve risk-adjusted returns across their tested portfolios; the paper also considers related volatility-management rules. This approach treats volatility as a uniform position-sizing input: all trades receive the same scaling regardless of whether the current volatility state favors or disfavors the underlying signal. The contribution is to reduce exposure during high-volatility periods and increase it during low-volatility periods. This is a risk-management device that does not discriminate between informative and uninformative signals.

Related work connects volatility to momentum returns through different channels. Zhuang (2018) studies volatility-informed allocation between an exchange-rate momentum portfolio and Treasury bills; this remains a position-allocation method rather than evidence for trade-level admission. Misirli (2024) examines aggregate volatility risk and cross-sectional momentum returns. These papers motivate attention to volatility, but their results do not directly validate the intraday futures mechanism proposed here.

The present design applies daily volatility information to individual intraday breakout entries and to their simulated position quantities. The admission and sizing roles are distinct in the specification but interact

in performance. The empirical contribution is this particular documented combination and its historical comparisons; no claim of unique novelty or independently identified gate effect is required.

2.4 Positioning and Hypothesis

The thesis combines conditional-volatility modeling, intraday breakout signals, and volatility-dependent position sizing. EGARCH enters both trade admissibility and exposure, so the complete-layer ablation assesses their joint contribution. The empirical exercises comprise layer and component comparisons, gate-timing and randomized-gate placebos, and retrospective training-window restrictions. These comparisons motivate the hypothesis that volatility information can improve breakout performance; they do not isolate an asymmetry mechanism or constitute independent prospective validation.

3 Data

This section describes the futures contracts, sample construction, data preparation, and descriptive features of the dataset used in the empirical analysis.

3.1 Sample and Asset Selection

The empirical analysis uses three liquid U.S. futures contracts: Nasdaq 100 futures (NQ), S&P 500 futures (ES), and WTI Crude Oil futures (CL). These contracts are selected because they represent distinct but economically important market segments. NQ and ES capture equity index dynamics and broad risk sentiment, while CL provides energy commodity exposure driven by global supply-demand dynamics, geopolitical risk, and macroeconomic growth expectations. Taken together, they offer a focused cross-section of futures markets that combines correlated equity index behavior with genuine cross-asset diversification from the energy sector.

The choice of these three contracts serves two purposes: it tests whether an EGARCH intraday breakout strategy can operate across both equity index and commodity futures, and it examines whether an energy futures contract with different volatility dynamics and EGARCH asymmetry properties adds meaningful diversification to a portfolio otherwise dominated by equity index exposure.

Five-minute open-high-low-close-volume (OHLCV) bars are obtained from Kibot (www.kibot.com), a commercial vendor of historical intraday futures data, as continuous front-month series for each contract. The raw intraday data extend from late 2009 to early 2026, but the main empirical analysis is restricted to the January 2019 to December 2025 evaluation period. This interval was chosen because it spans seven years containing multiple distinct market environments, including a relatively calm pre-crisis year, the COVID crash and rebound, the post-pandemic inflation shock, the 2022 monetary tightening and commodity supercycle, and the subsequent recovery through 2025, providing a demanding and economically relevant window for evaluating whether the framework generalizes across materially different market states.

3.2 Intraday Frequency and Session Restriction

The data are sampled at the 5-minute frequency, reflecting a compromise between informational richness and noise reduction. Relative to daily data, 5-minute observations allow the strategy to react to intraday breakouts and volatility changes in a timely manner; relative to tick-by-tick or 1-minute data, they are less dominated by microstructure noise. This frequency provides sufficient granularity to capture short-run market dynamics while preserving tractability in both econometric filtering and backtesting.

Although the raw futures files contain a much broader trading window, the empirical analysis uses asset-specific session restrictions. Equity index futures (NQ and ES) are restricted to the 09:30 to 14:00 U.S. session, while Crude Oil futures (CL) use the 09:30 to 13:00 session. These restrictions

are imposed for both economic and practical reasons. Economically, the opening and midday portions of the regular U.S. trading day are the most informative intraday periods for these contract types, combining high liquidity, elevated volatility, and concentrated price discovery. The shorter session for CL reflects the observation that crude oil momentum signals tend to degrade in the afternoon as NYMEX pit-session patterns attenuate. Practically, restricting the strategy to these windows reduces exposure to less informative trading periods and makes the final framework easier to interpret as a realistic intraday execution design.

These are entry-session restrictions. They do not imply that the archived backtest liquidates positions at the session boundary; the position loop and elapsed holding time must be interpreted separately (Section 4.6).

3.3 Data Preparation and Construction of Variables

Each futures file contains standard intraday fields: date, time, open, high, low, close, and volume. The first step in data preparation is the construction of a unified datetime variable from the separate date and time fields. After conversion, the data are sorted chronologically and checked for consistency in order and formatting.

The main return series used in the econometric and descriptive analysis is the 5-minute log return,

$$r_t = \ln \left(\frac{P_t}{P_{t-1}} \right),$$

where P_t denotes the futures closing price at time t . Log returns are preferred because they are additive over time and standard in the volatility modeling literature.

Several derived variables are then constructed for the trading framework. Each is defined formally below, as these indicators serve as core components of the signal and risk management logic.

Average True Range (ATR). The ATR is computed over a 14-bar rolling window. Let H_t , L_t , and C_{t-1} denote the high, low, and previous close at bar t . The true range is

$$\text{TR}_t = \max(H_t - L_t, |H_t - C_{t-1}|, |L_t - C_{t-1}|),$$

and the ATR is the 14-bar simple moving-average of TR_t :

$$\text{ATR}_t = \frac{1}{14} \sum_{k=0}^{13} \text{TR}_{t-k}.$$

Exponential Moving Average (EMA). The EMA with span s is defined recursively as

$$\text{EMA}_{s,t} = \alpha_s \cdot C_t + (1 - \alpha_s) \cdot \text{EMA}_{s,t-1}, \quad \alpha_s = \frac{2}{s + 1},$$

where C_t is the 5-minute closing price. The strategy uses asset-specific EMA pairs: spans 15 and 40 for NQ and ES, and spans 20 and 50 for CL.

Relative Strength Index (RSI). The RSI is computed over a 14-bar lookback. Let $\Delta_t = C_t - C_{t-1}$ denote the bar-to-bar price change. Average gains and average losses are computed as 14-bar simple moving-averages:

$$AG_t = \frac{1}{14} \sum_{k=0}^{13} \max(\Delta_{t-k}, 0), \quad AL_t = \frac{1}{14} \sum_{k=0}^{13} \max(-\Delta_{t-k}, 0).$$

The RSI is then

$$RSI_t = 100 - \frac{100}{1 + AG_t / (AL_t + \epsilon)},$$

where $\epsilon = 10^{-8}$ prevents division by zero. Values above 50 indicate net upward momentum; the strategy requires $RSI > 58$ for entry and exits when $RSI < 52$.

Rolling breakout high. The upper breakout boundary over an asset-specific lookback window L_i is

$$U_{i,t} = \max\{H_{i,t-1}, H_{i,t-2}, \dots, H_{i,t-L_i}\},$$

where $L_{NQ} = L_{ES} = 60$ bars and $L_{CL} = 30$ bars. The one-bar shift ensures the breakout level is computed from bars strictly prior to the current observation.

EGARCH-derived variables. Daily conditional volatility from the EGARCH(1,1) model is mapped onto each 5-minute bar by date. Three derived quantities are computed from the daily series: (i) the volatility expansion gate, (ii) the continuous volatility score, and (iii) the volatility regime classification. These are defined formally in Sections 4.3–4.4.

The final strategy therefore operates on a dataset that combines raw intraday price information with both technical and econometric state variables. This structure matters because the model is not built from prices alone. Instead, it relies on a layered decision process in which breakouts are confirmed by momentum and trend filters and activated only under suitable volatility conditions.

3.4 Data Integrity and Cleaning Procedure

Before running the strategy, the futures datasets were subjected to a structured audit. This step is particularly important in intraday futures research because apparent strategy profitability can be distorted by data problems such as duplicate timestamps, broken bars, missing intervals, incorrect ordering, or discontinuities around contract transitions.

The audit addressed several potential issues. First, timestamp integrity was verified by checking for duplicates and ensuring monotonic chronological ordering. Second, OHLC consistency was checked to ensure that high prices were not below the open, close, or low, and that low prices were not above

the open, close, or high. Third, the data were screened for nonpositive prices and implausible intraday jumps. Fourth, the asset-specific trading sessions were examined for missing bars.

The audit results indicate that the datasets are structurally sound. Across NQ, ES, and CL, no duplicate timestamps were detected, all timestamps were correctly ordered, and no OHLC consistency violations were found. Similarly, no nonpositive prices were present in the cleaned files. While some missing session bars were identified, the counts were limited relative to the full sample and were not large enough to suggest systemic data corruption. Overall, the audit provides confidence that the intraday files are suitable for empirical backtesting.

This cleaning and verification stage is important for interpreting the later results. In particular, it reduces the likelihood that the final performance measures are driven by mechanical data problems rather than by the actual strategy logic.

3.5 Descriptive Statistics

Table 1 reports descriptive statistics for 5-minute log returns across the three futures contracts. The table includes the mean, standard deviation, minimum, maximum, skewness, kurtosis, and total number of observations.

Table 1: Descriptive statistics for 5-minute log returns (2019–2025)

Asset	Mean	Std Dev	Min	Max	Skewness	Kurtosis	Obs
NQ	0.000010	0.001659	-0.080721	0.062704	-1.328	169.954	140375
ES	0.000007	0.001374	-0.075536	0.067505	-2.171	313.628	140375
CL	0.000002	0.003884	-0.414123	0.300641	-13.367	2616.722	141173

Note: Statistics computed on 5-minute log returns within the asset-specific trading sessions.

The descriptive table reports the highest five-minute return dispersion for CL (0.003884), followed by NQ (0.001659) and ES (0.001374). All three distributions have large reported kurtosis. Skewness is materially negative for NQ (−1.328) and ES (−2.171), and especially extreme for CL (−13.367). The CL minimum log return of −0.414123 warrants direct inspection of the underlying dates, contract adjustments, and return construction. These observations motivate heavy-tail modeling and data checks; they do not support the original characterization of skewness as small.

These properties reinforce the case for conditional volatility modeling. Heavy tails, time-varying variance, and occasional asymmetry all point toward the limitations of constant variance approaches, motivating the EGARCH framework developed in the methodology section.

3.6 Price Dynamics and Cross-Asset Structure

To visualize the cross-asset structure of the sample, Figure 1 plots normalized price series for the three futures contracts over the evaluation period. The figure highlights that the assets are far from identical.

Equity index futures display broad directional swings associated with risk appetite and macro conditions, while Crude Oil reflects a different pattern driven by energy supply-demand dynamics, OPEC policy, and commodity cycle behavior. CL experienced a sharp collapse and recovery during 2020 and a sustained rally during the 2021–2022 commodity supercycle, providing genuine diversification relative to the equity index contracts.

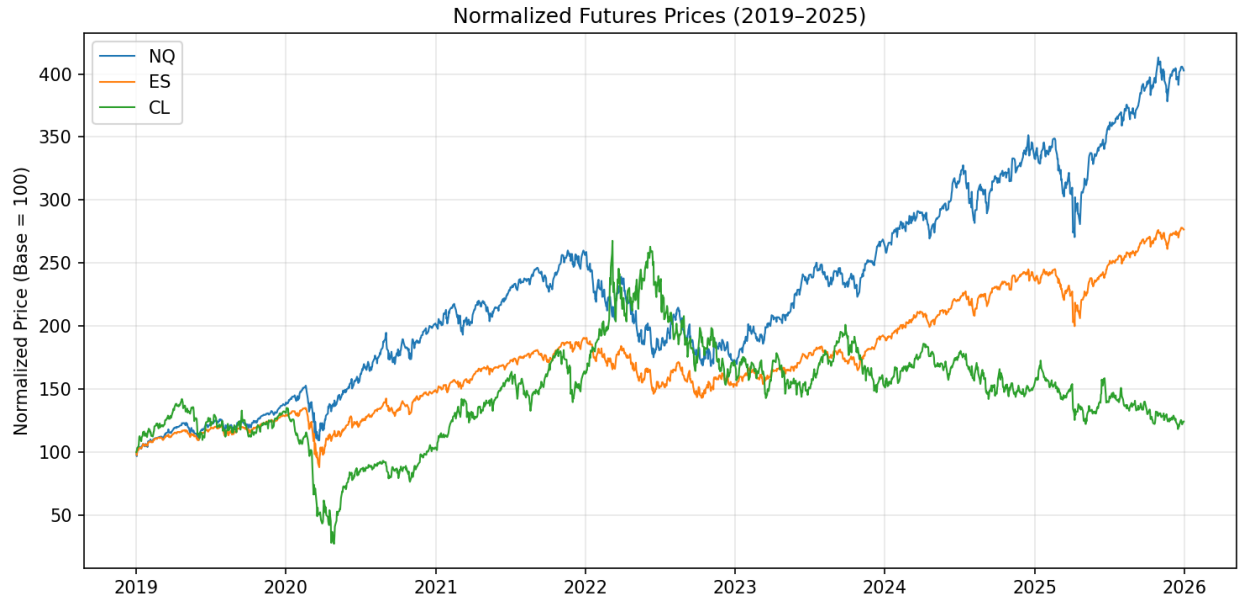


Figure 1: Normalized price series for NQ, ES, and CL over the 2019–2025 evaluation period. The figure illustrates the distinct price dynamics of the three futures markets used in the analysis.

Figure 2 reports the average intraday activity pattern based on normalized volume. The figure supports the session restrictions used in the thesis by showing that market activity is concentrated in the opening and midday portions of the trading day. This provides further justification for focusing the strategy on the 09:30 to 14:00 window for equity index futures and the 09:30 to 13:00 window for Crude Oil rather than applying the same logic uniformly across all intraday periods.

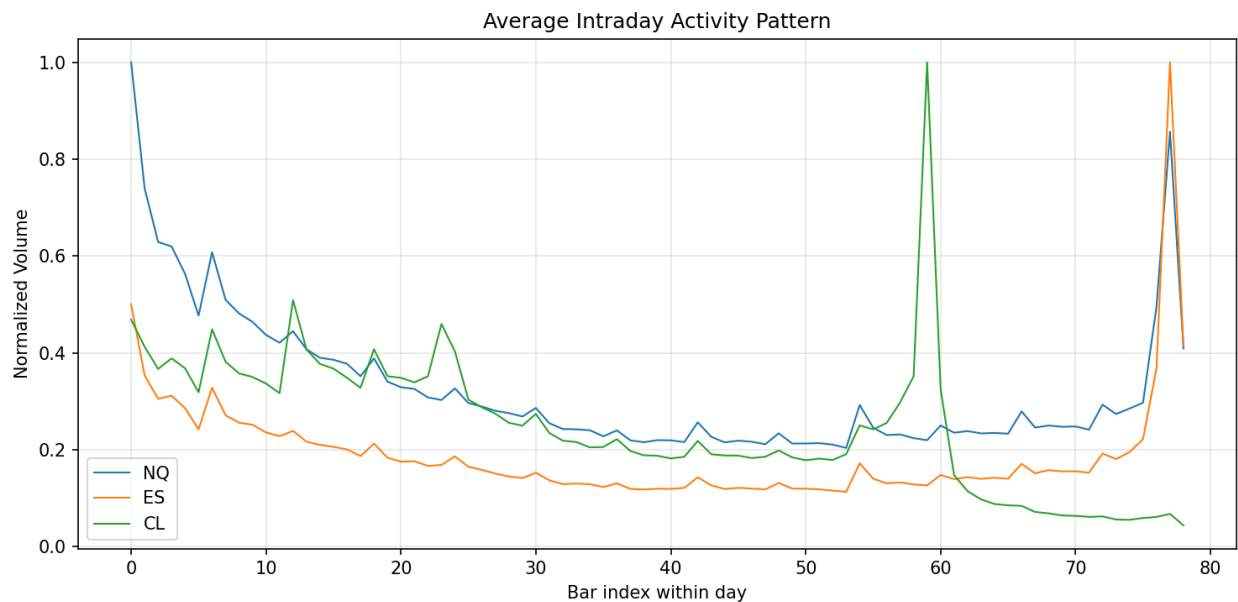


Figure 2: Average intraday activity pattern across the sample based on normalized volume. The opening and midday portions of the U.S. regular session show the strongest concentration of activity, supporting the asset-specific trading windows used in the thesis.

Together, the descriptive evidence confirms that the dataset combines meaningful variation across both time and asset class, a prerequisite for evaluating whether the framework generalizes beyond a single contract or market regime.

3.7 Implications for the Empirical Design

These data characteristics have direct implications for the empirical design. The heavy-tailed and heteroskedastic nature of 5-minute returns supports the use of EGARCH as a volatility filter. The clear differences in volatility dynamics across NQ, ES, and CL justify asset-specific parameterization. The concentration of market activity during the selected session windows supports restricting trading to the most liquid intraday intervals. The coexistence of calm and highly stressed subperiods within 2019–2025 makes the sample well suited for evaluating robustness across different market environments.

4 Methodology

This section describes the empirical design used to evaluate the proposed volatility-conditioned futures trading strategy. The framework combines daily volatility state identification with intraday trade execution. More specifically, EGARCH models are estimated on daily returns for each contract to classify the prevailing volatility environment, while signal generation and portfolio implementation are conducted using 5-minute data during regular trading hours. This separation is intentional. It allows volatility to be modeled at a frequency where parameter estimation is stable and economically interpretable, while preserving the higher frequency execution needed for breakout and momentum signals.

4.1 Data Construction and Sample Design

The data and sample construction are described in Section 3. For the methodology, the key definitions are as follows. Let $P_{i,t}$ denote the price of contract i at time t . Intraday log returns are

$$r_{i,t}^{(5m)} = 100 \cdot \log \left(\frac{P_{i,t}}{P_{i,t-1}} \right), \quad (1)$$

where the scaling by 100 expresses returns in percentage terms. Daily returns are constructed from daily closing prices derived from the 5-minute series:

$$r_{i,\tau}^{(d)} = 100 \cdot \log \left(\frac{P_{i,\tau}^{(d)}}{P_{i,\tau-1}^{(d)}} \right), \quad (2)$$

where $P_{i,\tau}^{(d)}$ denotes the daily closing price for contract i on day τ . The input files are continuous-contract series supplied by the vendor. The original description asserted a last-trading-day roll with back-adjustment, but the archived inputs do not by themselves verify that rule. Contract identifiers, roll dates, adjustment factors, and a reconstruction against individual contracts are needed before roll treatment can be treated as established. In particular, the absence of nonpositive prices in the cleaned CL series does not validate its treatment of the April 2020 contract dislocation.

4.2 EGARCH Volatility Modeling

Volatility is modeled using a contract-specific EGARCH(1,1) specification. The purpose of this step is not to predict returns directly, but to identify the volatility environment in which intraday signals are either allowed or suppressed. In this sense, EGARCH acts as a conditioning layer within the strategy rather than as an autonomous trading model.

For each contract, conditional variance is specified as

$$\log(\sigma_{i,\tau}^2) = \omega_i + \beta_i \log(\sigma_{i,\tau-1}^2) + \alpha_i \left(|z_{i,\tau-1}| - \sqrt{2/\pi} \right) + \gamma_i z_{i,\tau-1}, \quad (3)$$

where $\sigma_{i,\tau}^2$ denotes the conditional variance of daily returns, ω_i is a constant, β_i captures volatility persistence, α_i captures the effect of shock magnitude, and γ_i captures asymmetry in the volatility response to shocks. The standardized innovation is defined as

$$z_{i,\tau-1} = \frac{\varepsilon_{i,\tau-1}}{\sigma_{i,\tau-1}}, \quad (4)$$

where $\varepsilon_{i,\tau-1}$ is the innovation from the estimated daily mean equation. The diagnostic estimation script uses a constant mean and Student- t innovations; the provenance linking each cached trading-volatility series to those diagnostic estimates remains unresolved. The standardized Student- t specification has unit innovation variance. The displayed recursion follows the arch software convention (arch Developers, 2026), which centers the magnitude term at $\sqrt{2/\pi}$ even under Student- t likelihood. That constant is not generally $E|z|$ for a standardized Student- t random variable; changing the centering convention shifts the intercept. The degrees of freedom and precise estimation sample must be recorded with each fitted model for reproducibility.

Information available at the trading date. A one-day shift of a fitted volatility series prevents use of the current realized return, but does not remove information embedded in full-sample fitted coefficients or percentile thresholds. The archived daily fitting pipeline estimates EGARCH before shifting the resulting series. Unless coefficients and calibration distributions are fit only on observations available at each decision date, the resulting backtest is retrospective. The table originally labeled an expanding-rank exercise has unresolved generation provenance, as discussed below; it cannot certify the availability of the volatility inputs.

The logarithmic form of EGARCH ensures positive conditional variance without imposing nonnegativity constraints directly on the variance parameters. More importantly, it allows volatility to respond asymmetrically to shocks. This is especially relevant in equity futures, where negative shocks often trigger sharper increases in future volatility than positive shocks of similar size. Crude Oil futures also exhibit statistically significant negative asymmetry, though the effect is smaller in magnitude than in equity index contracts.

EGARCH models are estimated on daily log returns rather than directly on 5-minute returns. This choice is made for both econometric and economic reasons. From an econometric perspective, daily returns yield more stable parameter estimates and avoid the severe microstructure noise, excess zero returns, and intraday seasonal patterns that complicate volatility estimation at the 5-minute level. From an economic perspective, the goal is to identify persistent market states such as calm, normal, or stressed environments, rather than to capture bar-to-bar microstructure noise. Daily EGARCH therefore serves

as a slower moving state filter that is intentionally decoupled from the intraday execution frequency: it answers the question of whether today’s market environment is suitable for breakout trading, while the intraday signals determine when and how to trade within that environment. This frequency separation is a design choice, not a limitation. The EGARCH model identifies a persistent daily state (expanding, contracting, or extreme volatility) that changes slowly relative to the 5-minute execution horizon, ensuring that the conditioning signal is stable within each trading day and that no intraday feedback loop contaminates the volatility estimate. Once estimated, the daily conditional volatility series is mapped onto the intraday framework by date and used to govern signal activation.

Table 2 reports the estimated EGARCH(1,1) coefficients for each contract.

Table 2: EGARCH(1,1) parameter estimates by futures contract

Contract	ω	α	β	γ	SE(ω)	SE(α)	SE(β)	SE(γ)
NQ	0.0044	0.1361***	0.9774***	−0.1446***	0.0035	0.0146	0.0040	0.0122
ES	−0.0087**	0.1435***	0.9755***	−0.1738***	0.0038	0.0171	0.0044	0.0140
CL	0.0218***	0.1290***	0.9868***	−0.0497***	0.0041	0.0140	0.0026	0.0084

Note: EGARCH(1,1) models are estimated separately for each contract using daily log returns. Reported coefficients correspond to the specification $\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \alpha(|z_{t-1}| - \sqrt{2/\pi}) + \gamma z_{t-1}$. Standard errors are reported in adjacent columns. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The estimated coefficients indicate substantial volatility persistence across all three contracts, as shown by the consistently high values of β . This confirms that volatility clustering is a central feature of the sample. The asymmetry coefficient γ is negative and statistically significant for all three contracts, indicating that negative shocks increase future volatility more strongly than positive shocks of similar magnitude. The effect is strongest in ES ($\gamma = -0.1738$), followed by NQ ($\gamma = -0.1446$) and then CL ($\gamma = -0.0497$). The weaker but still significant asymmetry in Crude Oil is consistent with the observation that energy futures exhibit leverage-type volatility effects driven by supply disruptions and demand shocks, though the mechanism differs from the equity-market leverage-effect. These results support the use of a contract-specific EGARCH filter and show that asymmetric volatility dynamics are present across all three markets, with the strongest effects concentrated in the equity index contracts.

To verify the adequacy of the EGARCH(1,1) specification, standard residual diagnostics are reported in Table 3. For NQ, the Ljung–Box test on standardized residuals is marginally significant ($p = 0.036$), and the test on squared residuals rejects at the 1% level ($p = 0.008$), though the ARCH-LM test does not reject ($p = 0.086$). For ES, a similar pattern holds: both Ljung–Box statistics are significant at the 5% level ($p = 0.024$ and $p = 0.035$), but the ARCH-LM test is comfortably insignificant ($p = 0.305$). These mixed diagnostics do not establish correct specification: failure to reject an ARCH-LM test is not evidence that all conditional heteroskedasticity has been removed, and the Ljung–Box rejections warrant further investigation. For CL, the Ljung–Box test on squared residuals ($p < 0.001$) and the ARCH-LM test ($p < 0.001$) indicate meaningful residual volatility clustering not captured

by the EGARCH(1,1) specification. Higher-order specifications were tested as robustness checks: EGARCH(2,1), EGARCH(1,2), and EGARCH(2,2) with Student- t errors, as well as EGARCH(1,1) with a skewed Student- t distribution. The skewed- t specification achieves the best information criteria (the skewness parameter is significant at $p = 0.0001$, confirming that CL daily returns are left-skewed), but no specification eliminates the residual ARCH effects at the 5% level. The higher-order models add parameters that are generally insignificant while providing only marginal improvement in diagnostic p -values.

The persistence of residual volatility clustering reflects structural features of crude oil price formation that distinguish CL from the equity index contracts. Crude oil returns are shaped by a wider variety of exogenous shock types, including supply disruptions (OPEC decisions, geopolitical conflict, infrastructure outages), demand shocks tied to global growth, storage and contango dynamics, and regulatory events. Each of these can produce volatility jumps with distinct persistence profiles. The April 2020 negative-price episode, during which WTI front-month futures briefly traded below zero, represents an extreme instance of this complexity: CL exhibits 12 daily returns exceeding 15% in absolute value and excess kurtosis of 126.8 over the sample, compared with kurtosis below 20 for NQ and ES.

Persistent crude-oil volatility motivates comparison of parsimonious and more flexible GARCH-family parameterizations (Kang et al., 2009). The core difficulty is that GARCH-family models assume a single volatility process with smooth mean reversion, whereas CL's volatility is driven by discrete, heterogeneous shocks whose arrival rates and magnitudes vary across regimes. No standard specification, including the higher-order and skewed- t variants tested above, fully resolves this tension without introducing parameters that are poorly identified on a seven-year sample.

The remaining ARCH effects for CL show that the fitted model leaves predictable variation in squared residuals. Persistence and significant asymmetry do not establish that its gate classifications are accurate or that directional errors are harmless. Regime caps and smoothed sizing may limit exposure, but their effectiveness must be evaluated empirically rather than inferred from the model coefficients.

CL's regime caps and 40-day volatility smoothing limit simulated quantities, but cannot establish the accuracy of its gate. Its saved daily PnL correlation with the NQ+ES basket is 0.001, and the historical three-asset comparison reports Sharpe 1.412 versus 1.350 without CL. These are portfolio results within the selected sample. The standalone CL summaries disagree: the direction appendix reports Sharpe 0.479 and drawdown 6.8%, while the cost appendix reports 0.446 and 10.5% at baseline cost. They cannot be combined without resolving their return frequency and accounting conventions.

Table 3: EGARCH(1,1) residual diagnostics by futures contract

Contract	LB(10) z_t	LB(10) z_t^2	ARCH-LM(5)	p -value
NQ	19.35 [0.036]	23.84 [0.008]	9.64	[0.086]
ES	20.64 [0.024]	19.43 [0.035]	6.01	[0.305]
CL	10.90 [0.365]	41.61 [0.000]	25.01	[0.000]

Note: Ljung–Box statistics on standardized and squared standardized residuals with 10 lags (test statistic [p -value]). ARCH-LM test with 5 lags. Ljung–Box tests reject for NQ and ES despite non-rejection of their ARCH-LM tests; this does not establish adequate fit. CL exhibits significant residual ARCH effects. Displayed 0.000 values are rounded, not exact zero probabilities.

4.3 Volatility Expansion Gate and Continuous Volatility Score

The estimated daily conditional volatility series is used in two distinct ways within the trading framework. First, a volatility expansion gate determines whether trading is permitted on a given day. Second, a continuous volatility score determines the intensity of position sizing within eligible days.

The volatility expansion gate is the binary admission component of the volatility layer. It requires that the daily conditional volatility exceeds its own 10-day moving-average, lagged by one day:

$$\text{vol_expand}_{i,\tau} = \begin{cases} 1, & \hat{\sigma}_{i,\tau-1} > \text{MA}_{10}(\hat{\sigma}_{i,\cdot})_{\tau-1} \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where $\hat{\sigma}_{i,\tau-1}$ denotes the lagged daily conditional volatility for contract i . The one-day lag delays the observed volatility input; historical availability also requires the coefficients and calibration quantities to have been estimated without later observations. This condition captures whether volatility is locally expanding, which empirically coincides with periods in which breakout signals are more informative.

The continuous volatility score is a flat-middle function of the conditional-volatility percentile $p_{i,\tau}$. In the development pipeline, ranks are calculated across the available full sample. This uses future observations when ranking an earlier day and is a source of look-ahead in calibration, even when the underlying volatility value is lagged and the shape parameters are fixed. The archived trade-analysis implementation ranks a volatility value repeated on intraday rows, which can also weight days by their number of bars. A daily expanding empirical distribution, constructed strictly before the decision date, is the appropriate implementable alternative. The original claim of an expanding-rank check is not verified by the examined table-generation code (Table 13); it is not used here to dismiss this bias. The four score parameters are $(a, b, c, d) = (0.20, 0.40, 0.80, 0.90)$.

$$\text{vol_score}(p) = \text{clip} \left(\begin{cases} 0.30 + \frac{p}{a} \times 0.50, & p \leq a \\ 0.80 + \frac{p-a}{b-a} \times 0.20, & a < p \leq b \\ 1.00, & b < p \leq c \\ 1.00 - \frac{p-c}{d-c}, & c < p \leq d \\ 0, & p > d, \end{cases} \quad 0, 1 \right) \quad (6)$$

where $\text{clip}(x, 0, 1)$ bounds the output to $[0, 1]$. With $(a, b, c, d) = (0.20, 0.40, 0.80, 0.90)$, the score equals 0.30 at the 0th percentile, rises to 0.80 at the 20th, ramps linearly from 0.80 to 1.00 between the 20th and 40th percentiles, holds flat at 1.00 between the 40th and 80th percentiles, decays linearly from 1.00 to 0.00 between the 80th and 90th percentiles, and is zero above the 90th percentile. The name “flat-middle” refers to the plateau at full weight across the 40th–80th percentile range.

This shape is motivated by the empirical finding that the strategy performs well across a broad range of moderate-to-elevated volatility conditions. The flat plateau ensures that position sizing remains stable throughout the core of the volatility distribution, avoiding unnecessary variation in exposure. The ramp below the 20th percentile assigns reduced but nonzero weight to low-volatility periods, acknowledging that breakout signals in very calm markets are weaker but not entirely uninformative. The sharp decay above the 80th percentile and hard cutoff at the 90th prevent trading during extreme-volatility episodes where breakout signals become unreliable. A position is blocked entirely when $\text{vol_score} < 0.05$.

The volatility expansion gate and the continuous score serve different roles. The gate determines whether a trade is *eligible* (binary on/off), while the continuous score determines the *intensity* of position sizing within eligible days. The model therefore does not assume that volatility itself predicts return direction. It assumes that volatility conditions the probability that an observed breakout or momentum signal will lead to meaningful continuation.

Two placebo exercises are used to test whether this admissibility rule adds information beyond generic trade suppression. The first shifts the gate forward and backward by $\{1, 3, 5, 10\}$ trading days while holding all other components fixed. The second randomly reassigns gate-active days within each asset and calendar year while preserving the exact number of active days in that asset and year. The timing-shift placebo asks whether contemporaneous alignment matters; the randomized placebo asks whether the result could be replicated by an economically meaningless gate with the same annual frequency. In both cases, only gate timing or gate content is altered.

4.4 Volatility Regime Classification

In addition to the expansion gate and continuous score, the estimated daily conditional volatility series is used to classify each trading day into one of three volatility regimes: low, medium, or high. The regime classification uses the 42nd and 58th percentile thresholds:

$$\mathcal{R}_{i,\tau} = \begin{cases} \text{Low,} & \hat{\sigma}_{i,\tau} < q_{0.42,i} \\ \text{Medium,} & q_{0.42,i} \leq \hat{\sigma}_{i,\tau} < q_{0.58,i} \\ \text{High,} & \hat{\sigma}_{i,\tau} \geq q_{0.58,i}, \end{cases} \quad (7)$$

where $q_{0.42,i}$ and $q_{0.58,i}$ denote the 42nd and 58th percentiles of the contract-specific conditional volatility distribution. Full-sample thresholds carry the same future-calibration issue as the score ranks. Lagging the assigned label does not by itself make a full-sample threshold available in real time.

The regime classification does *not* serve as a binary entry filter. Unlike the volatility expansion gate, which directly gates trade eligibility, the regime labels are used only for Crude Oil position caps. This reflects the empirical finding that applying a hard regime entry filter (blocking trades during high-volatility regimes) actually degrades performance, while using regimes to modulate CL exposure improves risk-adjusted returns. The regime caps for Crude Oil are:

$$\text{CL: } w_{\text{low}} = 0.00, \quad w_{\text{medium}} = 0.40, \quad w_{\text{high}} = 0.80.$$

NQ and ES do not use regime-based caps; their full exposure is governed by the volatility expansion gate and the continuous vol-score. Each daily regime label is mapped onto all 5-minute observations within the corresponding trading day.

4.5 Intraday Signal Construction

Once the daily volatility gate has been evaluated, signals are generated from 5-minute data. The strategy combines momentum and breakout logic rather than relying on volatility alone. In this design, EGARCH determines the market state, while the intraday indicators determine timing and direction.

A long trade in contract i is entered at bar t if and only if **all six** of the following conditions hold simultaneously:

1. **Volatility expansion gate:** $\text{vol_expand}_{i,\tau} = 1$ (daily conditional volatility exceeded its 10-day MA on the previous day; see Equation 5).
2. **Volatility score positive:** $\text{vol_score}(p_{i,\tau}) > 0.05$ (the continuous volatility score, based on today's conditional volatility percentile rank, is not in the extreme-vol exclusion zone; see Equation 6).

3. **Breakout:** $C_{i,t} > U_{i,t}$ (the current 5-minute close exceeds the rolling high of the previous L_i bars, where $L_{NQ} = L_{ES} = 60$ and $L_{CL} = 30$; see Section 3.3).
4. **RSI momentum confirmation:** $RSI_{i,t} > 58$ (the 14-bar RSI exceeds the entry threshold, ensuring that directional pressure is already established before a breakout is acted upon).
5. **EMA trend alignment:** $EMA_{s_1,t} > EMA_{s_2,t}$ (the short-period EMA exceeds the medium-period EMA, where $(s_1, s_2) = (15, 40)$ for NQ and ES and $(s_1, s_2) = (20, 50)$ for CL). The use of faster EMA periods for equity index contracts reflects their more responsive trend dynamics; CL benefits from a slower trend filter that better captures the longer-horizon momentum structure of energy futures.
6. **Cooldown and session:** At least 5 bars (25 minutes) have elapsed since the last trade exit, and the current bar falls within the asset-specific session window (09:30–14:00 ET for NQ and ES; 09:30–13:00 ET for CL).

The strategy is restricted to long-only positions; short breakouts are not traded. This restriction reflects both the theoretical arguments discussed in Section 2.2 (momentum crash risk and positive equity drift) and the empirical finding that the breakout signal produces more robust results on the long side over the seven-year evaluation period. For CL specifically, where the positive drift argument does not apply, Appendix E confirms that the short-only specification achieves a Sharpe ratio of just 0.122 versus 0.479 for long-only. These conditions govern new entries when the strategy is flat. Failure of an entry condition does not close an existing position; existing positions follow the exit rules.

4.6 Portfolio Construction and Position Management

Signals are generated separately for each contract and then aggregated into a portfolio. Positions are taken only when the contract-specific volatility gate and intraday entry conditions are both satisfied. When flat, the strategy opens no new position without a valid signal. An existing position persists until an exit condition is met.

Portfolio construction follows a diversified multi-asset approach in which contracts are combined into a single strategy portfolio. Position sizing uses an EGARCH inverse volatility targeting (IVT) framework inspired by Moreira and Muir (2017). The sizing and PnL accounting proceed in four steps:

Step 1: Base contract count (IVT). The base contract count is the ratio of the pre-sample median conditional volatility to the 40-day moving-average of lagged daily conditional volatility, multiplied by a contract-count scaling parameter:

$$n_{i,t}^{\text{base}} = \text{clip} \left(\frac{\tilde{\sigma}_{\text{median},i}}{\text{MA}_{40}(\hat{\sigma}_{i,\tau-1})} \times L, \quad 0.25, \quad 3L \right),$$

where $\tilde{\sigma}_{\text{median},i}$ is the median of daily EGARCH conditional volatility computed over the pre-OOS period (before January 2019), $\text{MA}_{40}(\hat{\sigma}_{i,\tau-1})$ is the 40-day moving-average of lagged conditional volatility, $L = 5$ is a contract-count scaling parameter, and the result is clipped to $[0.25, 15.0]$. When current conditional volatility is below its historical median, IVT allocates more than L contracts; when it is above, fewer. This is a model-based inverse-volatility sizing rule; both the fitted model and realized-volatility alternatives use historical returns. It is not an explicit portfolio volatility target, since it does not solve for dollar exposure or cross-asset covariance.

Step 2: Vol-score scaling. The base count is multiplied by the continuous volatility score:

$$n_{i,t}^{\text{scaled}} = n_{i,t}^{\text{base}} \times \text{vol_score}(p_{i,\tau}).$$

Step 3: Regime cap (CL only). For Crude Oil, effective contracts are capped by the regime-conditional limit:

$$n_{\text{CL},t}^{\text{eff}} = \min(n_{\text{CL},t}^{\text{scaled}}, w_{\mathcal{R}} \times L),$$

where $w_{\mathcal{R}} \in \{w_{\text{low}} = 0.00, w_{\text{med}} = 0.40, w_{\text{high}} = 0.80\}$ is the regime-conditional cap weight. For NQ and ES, the cap is $1.00 \times L = 5.0$ contracts in all regimes:

$$n_{i,t}^{\text{eff}} = \min(n_{i,t}^{\text{scaled}}, 5.0), \quad i \in \{\text{NQ}, \text{ES}\}.$$

Step 4: Profit and loss computation. Dollar PnL per bar is

$$\text{PnL}_{i,t} = \text{pos}_{i,t-1} \times (C_{i,t} - C_{i,t-1}) \times V_i \times n_{i,t-1}^{\text{eff}} - \mathbf{1}_{\text{entry},t} \times \kappa_i \times n_{i,t}^{\text{eff}},$$

where $\text{pos}_{i,t-1} \in \{0, +1\}$ is the position direction from the prior bar (long-only specification), V_i is the contract point value (\$20 for NQ, \$50 for ES, \$1,000 for CL), and κ_i is the per contract round-turn transaction cost (\$14.51 for NQ, \$29.51 for ES, \$12.80 for CL). The original inputs are described as fees plus an estimated half-spread; a separately timestamped fee, spread, and slippage calibration is not documented here. Transaction costs are charged at entry on the effective number of contracts. Portfolio PnL is the sum across all three contracts.

The strategy starts with initial capital of \$200,000. The rule permits fractional contract quantities and does not measure account leverage. Actual gross notional exposure would be $\sum_i |q_{i,t}| P_{i,t} V_i / E_t$, where E_t is current equity; margin requirements are a separate constraint. Integer contract rounding, margin, market impact, and financing or collateral interest are not validated by this simulation.

The archived bar-level accounting uses the preceding bar's effective quantity, which may change during a position, whereas the trade-statistics script fixes quantity at entry. The cost equation charges a round turn at entry and does not charge separate costs for those quantity changes. These accounting

conventions explain why the two summaries need reconciliation; they must not be silently treated as a single executable trade ledger.

Trade management. Positions are opened when the signal criteria in Section 4.5 are satisfied and closed when any of the following conditions is met:

1. **Trailing ATR stop-loss.** The stop level S_t is initialized at entry bar t_0 as $S_{t_0} = C_{t_0} - m_i \times \text{ATR}_{t_0}$ for a long position, where m_i is the asset-specific ATR multiplier ($m_{\text{NQ}} = m_{\text{ES}} = 2.2$, $m_{\text{CL}} = 2.5$). On each subsequent bar, the stop ratchets upward:

$$S_t = \max(S_{t-1}, C_t - m_i \times \text{ATR}_t).$$

The position is closed when $C_t \leq S_t$. The wider multiplier for CL (2.5 versus 2.2) accommodates the higher intraday volatility of energy futures and prevents premature stopouts.

2. **RSI exit.** The position is closed if RSI falls below 52.
3. **Maximum holding period.** The position is closed after 72 observed bars. This equals six elapsed hours only if those bars are consecutive at five-minute spacing. The archived simulation uses the session condition for entry and contains no explicit session-end liquidation in the examined position loop; positions can therefore remain open outside the entry window. A strictly intraday implementation requires a separately specified and tested liquidation rule.

No fixed profit target is used; the trailing ATR stop allows winning trades to run without a predetermined cap on gains. A five-bar cooldown period (25 minutes) follows each exit before a new entry in the same contract is permitted. The saved comparisons aim to share execution logic across variants. Nonetheless, the differing bar-level and trade-level accounting conventions require reconciliation before comparisons are interpreted as executable results. Signals and close-based exits are evaluated at the same bar close used for accounting, which assumes fills at that price; a next-observable-price execution check is needed.

4.7 Performance Measurement

Strategy performance is evaluated using cumulative return, Sharpe ratio, Sortino ratio, and maximum drawdown. Let R_t denote the daily portfolio return at trading day t , formed by aggregating intraday PnL into a single daily PnL observation and dividing by the initial capital base. The Sharpe ratio is defined as

$$\text{Sharpe} = \frac{\bar{R}}{\sigma_R} \sqrt{252}, \quad (8)$$

where $\bar{R}$ is the mean daily return and σ_R is the standard deviation of daily returns, annualized by the square root of the conventional 252 trading days per year. The historical ratio labeled Sortino is computed as

$$\text{Historical negative-return ratio} = \frac{\bar{R}}{\sigma_R} \sqrt{252}, \quad (9)$$

where the archived diagnostic computes σ_R^- as the sample standard deviation of the negative daily returns about their conditional mean. Thus the historical quantities labeled “Sortino” in the preserved tables are negative-return-standard-deviation ratios, not the conventional target-downside-deviation Sortino ratio. For a zero target, the conventional denominator would instead be $\sqrt{T^{-1} \sum_{t=1}^T \min(R_t, 0)^2}$. These definitions can differ materially. Conventional Sortino values require recomputation from a reconciled daily return series.

Reported Sharpe ratios use raw PnL divided by initial capital, without subtracting a risk-free rate, and the conventional $\sqrt{252}$ annualization factor. This scaling does not adjust for autocorrelation. Dollar equity is $E_t = E_0 + \sum_{u \leq t} \text{PnL}_u$ and maximum drawdown is $\min_t (E_t / \max_{u \leq t} E_u - 1)$. A terminal equity multiple of 3.88 corresponds to a cumulative net return of 288%, not 388%. Annual results computed after resetting equity to E_0 must not be compounded or compared with full-path drawdowns as if they shared the same capital denominator.

The passive benchmark holds one futures contract against the stated capital base. One futures contract is not inherently unlevered: notional exposure depends on price and point value relative to capital. The strategy changes eligibility and quantity, so raw returns and drawdowns are not exposure-matched comparisons. Sharpe comparisons describe the saved return streams, but do not identify the source of improvement. A useful follow-up would hold the candidate-trade set and sizing rule constant while changing only the gate, and separately vary sizing while fixing the gate.

In addition to full sample performance, the analysis evaluates subperiod behavior to assess how the strategy performs across different market environments. This matters because a volatility-conditioned framework should not be expected to perform uniformly in calm and stressed regimes. The subperiod analysis therefore serves as a complement to the full sample results by showing whether the strategy’s economic role is concentrated in particular market states.

4.8 Model Comparison and Ablation Design

To determine whether the econometric layer contributes meaningfully to performance, the thesis implements a structured model comparison exercise. The preferred specification is compared with ablated versions in which specific EGARCH components are removed. This design makes it possible to distinguish between the contribution of the signal logic itself and the additional contribution of volatility conditioning.

The complete-layer ablation removes the expansion gate, volatility score, inverse-volatility sizing, and regime caps while retaining breakout, RSI, EMA, and trade-management rules. It therefore measures a bundled change in selection and exposure. Component variations provide additional sensitivity evidence

but do not convert that bundled contrast into a pure gate effect. Isolating EGARCH asymmetry requires a matched volatility-model comparison with $\gamma = 0$, using the same historical estimation windows and decision-time calibration. No such isolated causal estimate is claimed here.

At the same time, the comparison of many variants raises a multiple testing concern. Although the strategy variants are evaluated within a structured empirical framework, repeated comparison across the same evaluation period weakens the interpretation of the test window as a perfectly untouched out-of-sample exercise. The reported results should therefore be interpreted as comparative evidence within an iterative research design rather than as a single-shot validation exercise immune to model selection effects.

Two falsification exercises complement the ablations. First, the EGARCH expansion gate is shifted forward and backward by $\pm\{1, 3, 5, 10\}$ trading days, misaligning the volatility signal with the intraday execution window while holding all other parameters fixed. Negative shifts (using future volatility information) are included only as an anticipatory sensitivity check. Second, the gate-active days are randomly reassigned within each asset and calendar year while preserving the exact number of active days in that asset and year. The first test asks whether contemporaneous timing matters; the second asks whether the observed result could be replicated by an economically meaningless gate with the same annual frequency. Both are reported in Section 5.10.

4.9 Statistical Inference: Circular Block Bootstrap

A paired circular block bootstrap (Politis and Romano, 1991) describes uncertainty in the difference between the saved development-specification and No-EGARCH return streams. Fixed-length blocks preserve some local dependence and are sampled jointly for both specifications. This resampling is conditional on the chosen strategies, dataset, and calibration; it does not repeat the research search or establish an independent validation sample.

The procedure operates on paired daily portfolio returns $\{r_t\}_{t=1}^T$. The original bootstrap description reports $T = 1,762$, while the saved daily-frequency comparison reports 1,805 observations. The observation calendar and exact bootstrap input therefore require reconciliation. Let r_t^{final} and $r_t^{\text{no-EGARCH}}$ denote the two aligned streams. Define the annualized Sharpe difference as

$$\Delta S = \sqrt{252} \left(\frac{\bar{r}^{\text{final}}}{\hat{\sigma}^{\text{final}}} - \frac{\bar{r}^{\text{no-EGARCH}}}{\hat{\sigma}^{\text{no-EGARCH}}} \right), \quad (10)$$

where both ratios are annualized by the standard $\sqrt{252}$ factor.

The bootstrap proceeds as follows. The return series is treated as circular: observation $T + 1$ wraps to observation 1. In each replication $b = 1, \dots, B$, a bootstrap sample of length T is constructed by drawing blocks of consecutive observations of fixed length ℓ , starting from uniformly random positions in the circular series. The same block positions are used for both return arrays. This preserves

contemporaneous pairing only if the arrays contain identical dates in identical order. The examined generator drops date labels and truncates arrays to a common length; explicit date alignment must be verified before treating the pairing as established. The Sharpe difference $\Delta S^{*(b)}$ is computed from each bootstrap sample.

The reported nonpositive-tail fraction is the fraction of B bootstrap replications in which the resampled Sharpe difference is zero or negative:

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B \mathbf{1}[\Delta S^{*(b)} \leq 0]. \quad (11)$$

The archived result uses block length $\ell = 20$ days and $B = 10,000$ replications. The reported fraction, 0.0038, is the share of resampled differences at or below zero. The resampling distribution is not recentered under a null hypothesis, so this is a bootstrap tail summary rather than a fully calibrated null-imposed test. The original text claimed similar conclusions at block lengths 10 and 40; those results must be linked to saved outputs before that sensitivity claim can be relied upon.

Dependence-preserving resampling is useful for these non-normal return streams, but its adequacy depends on the observation calendar, block length, and stationarity assumptions. The one-month block choice is not justified by a five-bar cooldown alone. A reproducible inference update should align the exact calendar, report confidence intervals and block-length sensitivity, and distinguish within-design sampling uncertainty from selection uncertainty.

4.10 Complete Parameter Specification

Table 4 consolidates all strategy parameters for reference. Parameters are divided into three groups: EGARCH/volatility layer parameters (shared across contracts), intraday signal parameters (asset-specific), and trade management parameters.

Table 4: Complete parameter specification of the final three-asset model

Parameter	NQ	ES	CL	Description
EGARCH / Volatility Layer (shared except where noted)				
EGARCH specification	EGARCH(1,1), Student- t			Estimated on daily log returns
Vol expansion gate MA	10 days			MA window for expansion condition
Vol expansion multiplier	1.00			$\hat{\sigma}_{\tau-1} > 1.00 \times \text{MA}_{10}$
Vol expansion lag	1 day			Lags the observed series
Vol-score shape (a, b, c, d)	(0.20, 0.40, 0.80, 0.90)			Flat-middle tent function
IVT smoothing window	40 days			MA of lagged cond. vol
Contract-count scale L	5			Applied in IVT formula
IVT clip bounds	[0.25, 15.0]			$3 \times L$ upper bound
Regime thresholds	42nd / 58th pct.			For CL caps only
Regime caps $(w_{\text{low}}, w_{\text{med}}, w_{\text{high}})$	—	—	0.0, 0.4, 0.8	NQ/ES fixed cap L
Intraday Signal Parameters (asset-specific)				
Session window (ET)	09:30–14:00	09:30–14:00	09:30–13:00	Trading hours
Breakout lookback L_i	60 bars	60 bars	30 bars	Rolling high window
EMA fast / slow spans	15 / 40	15 / 40	20 / 50	Trend alignment
RSI lookback	14 bars			Standard RSI period
RSI entry threshold	58			Long entry > 58
Trade Management				
ATR lookback	14 bars			True Range smoothing
ATR stop multiplier m_i	2.2	2.2	2.5	Trailing stop distance
RSI exit threshold	52			Long exit < 52
Max holding period	72 observed bars			Bar-count exit
Cooldown	5 bars (25 min)			Post-exit waiting period
Direction	Long only			No short trades
Cost and Capital				
Initial capital	\$200,000			Starting equity
Point value V_i	\$20	\$50	\$1,000	Per-point contract value
Round-turn cost κ_i	\$14.51	\$29.51	\$12.80	Fixed cost assumption

5 Empirical Results

This section presents the saved historical results. The development specification was refined using 2019–2025; the subsequent training-window, subperiod, and rolling-search exercises are retrospective checks within that research history. None is described as a pristine prospective holdout. Numerical outputs are retained unless an internally supported correction is available, and outstanding differences are disclosed at the relevant table.

5.1 Development-Enhanced Specification

The development specification implements the parameter set in Table 4, refined using the 2019–2025 evaluation period. It is the baseline for historical comparisons. Its reported performance is exposed to selection; the later checks restrict particular aspects of that selection but do not quantify or remove all resulting bias.

Table 5 summarizes performance for both the EGARCH-filtered and unfiltered specifications.

Table 5: Main performance results: final three-asset model vs. no-EGARCH ablation (2019–2025)

Model	Sharpe	Sortino [†]	Max DD	Return Mult.	Trades	Neg. Years
Final 3-asset	1.412	1.579	-10.6%	3.88×	1215	0
No EGARCH	0.664	1.261	-29.3%	4.65×	3842	0

Note: Reported development results over 2019–2025 with \$200,000 initial capital. Transaction-cost inputs per round turn: NQ \$14.51, ES \$29.51, CL \$12.80. The ablation removes the combined gate, score, sizing, and cap layer. Return Mult. is terminal equity divided by initial capital; it is not cumulative net return.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The development model reports a Sharpe ratio of 1.412, maximum drawdown of 10.6%, and terminal equity of 3.88 times initial capital over seven years, with 1,215 completed trades. These selected historical estimates are not an upper confidence bound. The retrospective training-window exercise reports Sharpe 1.195 (Section 5.15). The saved annual summaries contain no negative calendar years, but their exact capital and trade-accounting conventions require reconciliation with the full-period series.

The asset-specific parameterization is economically interpretable rather than ad hoc. NQ and ES use longer breakout windows and faster trend filters to match faster equity-index continuation, while CL uses a shorter breakout window, slower EMA filter, wider ATR stop, and shorter session to reflect noisier energy-futures microstructure. These differences help keep the empirical comparison focused on the conditioning mechanism rather than on forcing one uniform specification across distinct contracts.

Figure 3 illustrates the cumulative equity paths of the final three-asset model and the No-EGARCH specification.

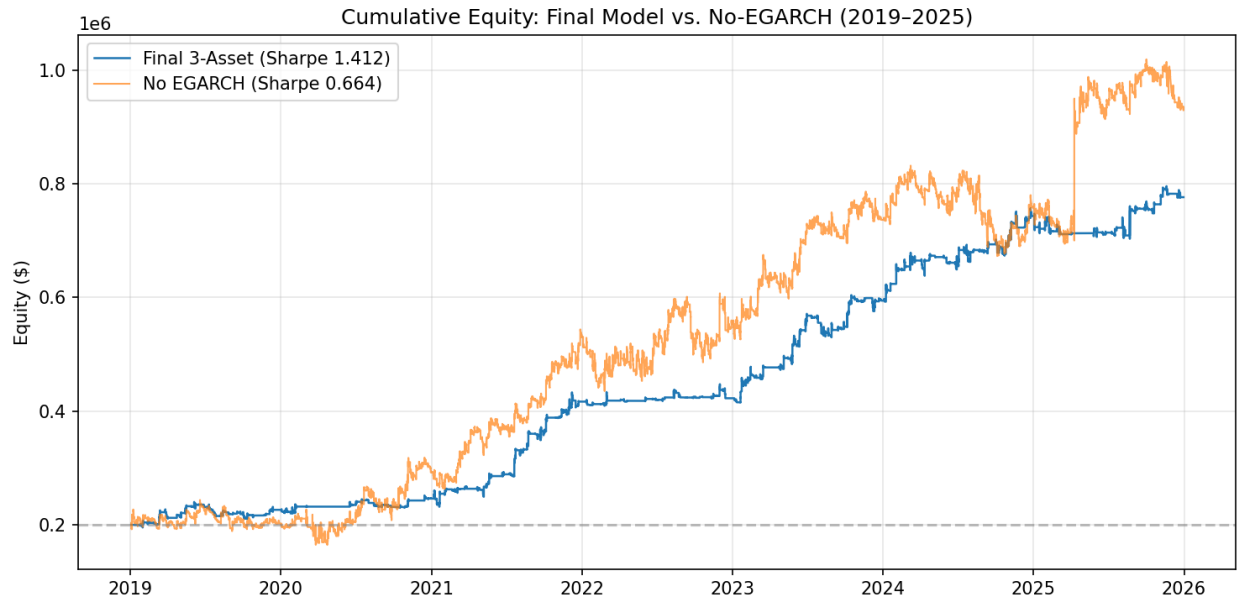


Figure 3: Equity curves of the final three-asset model and the No-EGARCH specification over the 2019–2025 evaluation period.

The equity curve illustrates the EGARCH gate’s interaction with market regimes: during stable volatility (2019), the gate permits few entries; during extreme spikes (March 2020, H1 2022), the score cutoff can suppress entry and IVT can reduce position sizes even if the expansion gate is open; during moderately expanding volatility (2021, 2023–2024), the gate opens and the vol-score assigns full conviction, producing the strongest returns. The No-EGARCH version achieves a higher terminal return ($4.65\times$ versus $3.88\times$) but at the cost of a 29.3% maximum drawdown.

5.2 The EGARCH Conditioning Effect: Ablation Evidence

The full-layer ablation removes the expansion gate, continuous volatility score, inverse-volatility sizing, and regime caps. It holds the stated directional signals and exit rules fixed, but changes both trade selection and risk exposure. The resulting comparison measures the joint effect of the volatility layer in the selected historical design.

Removing the layer changes the reported Sharpe from 1.412 to 0.664 and maximum drawdown from 10.6% to 29.3%. The unfiltered simulation records 3,842 trades versus 1,215. This is consistent with greater selectivity and altered exposure, but neither the trade count nor the Sharpe difference identifies their separate contributions. The difference of the rounded Sharpe values is 0.748; the bootstrap table reports 0.749 from its own stored calculation.

Component comparisons suggest that gate timing matters within the chosen architecture. Their interpretation remains conditional on the jointly selected score, sizing, session, and exit rules; they do not independently validate the proposed information-arrival or liquidity mechanisms.

Figure 4 shows the drawdown profiles of the final model and the No-EGARCH version.

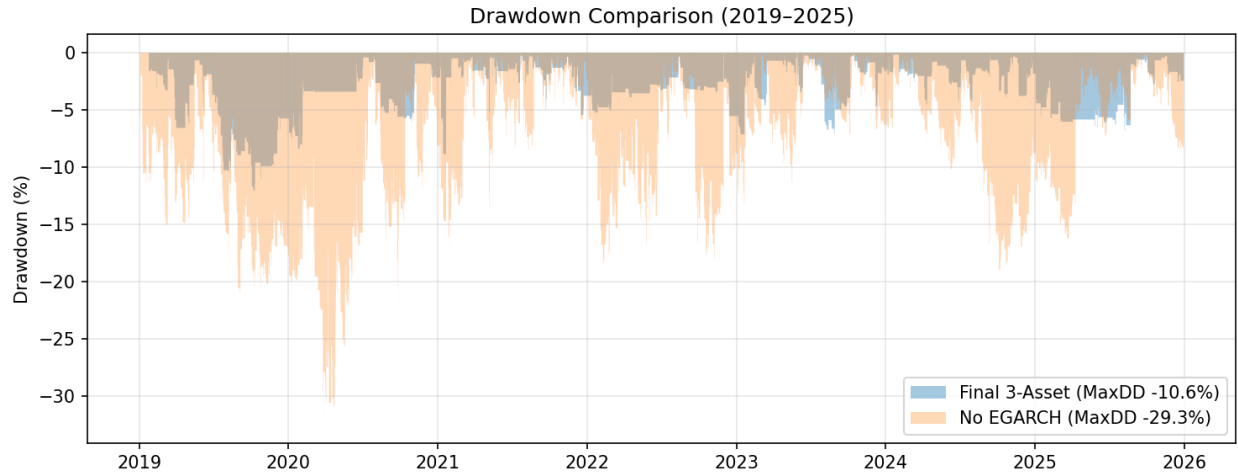


Figure 4: Drawdown comparison of the final three-asset model and the No-EGARCH specification over the 2019–2025 sample.

Without EGARCH conditioning, the framework follows a substantially more unstable return path, with deeper losses concentrated during the periods (2020, 2022) when the combined volatility layer would have changed entries and exposure.

Table 6 reports the paired circular block bootstrap of the development comparison, with 20-day blocks and 10,000 replications. The nonpositive-tail fraction is 0.0038. It must not be transferred to the retrospective pre-2019-selected comparison, for which the two reported Sharpes are 1.195 and 0.382.

Table 6: Paired circular block bootstrap of the development Sharpe difference

Comparison	Sharpe Diff (ann.)	Tail fraction
Final vs. No-EGARCH	0.749	0.0038

Note: Block length = 20 days, 10,000 replications. Sharpe Diff is the annualized Sharpe ratio of the final model minus the no-EGARCH specification. Tail fraction is the one-sided fraction of bootstrap replications in which the Sharpe difference is zero or negative.

5.3 Asset-Level Contribution and Trade Activity

The aggregate portfolio results conceal meaningful variation across assets. Figure 5 plots the cumulative contribution of each contract to total profit and loss over the evaluation period.

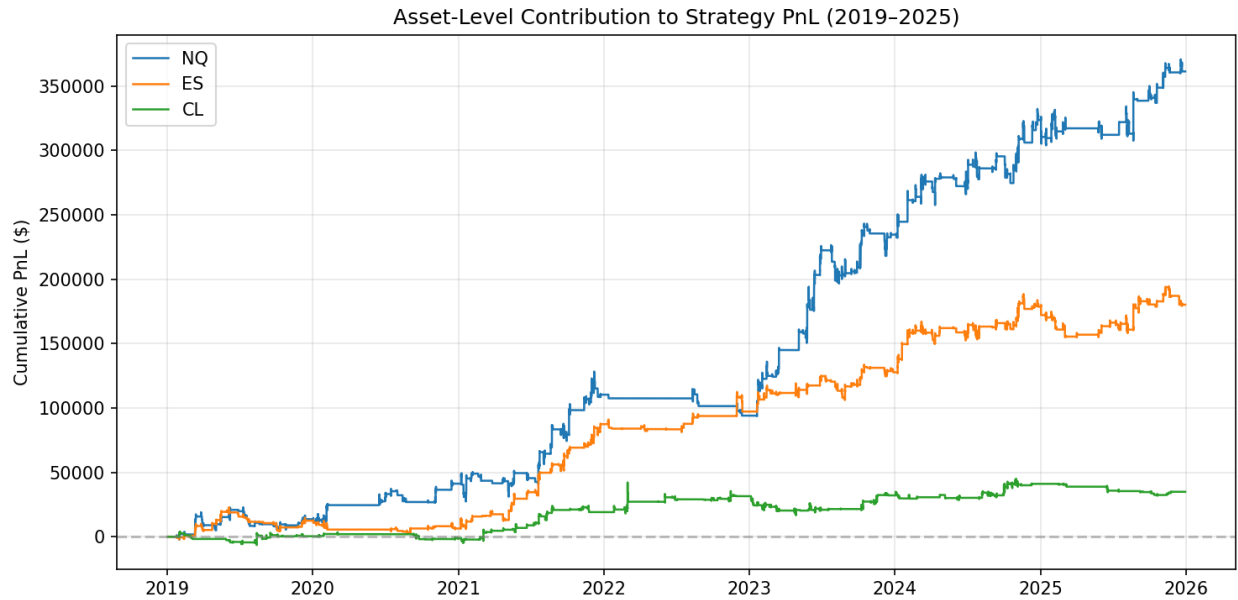


Figure 5: Cumulative net profit and loss contribution by asset in the final three-asset model.

NQ and ES are the primary directional contributors. CL provides diversification: its daily PnL correlation with the equity index basket is 0.001, and removing CL introduces a negative calendar year (2022) that the three-asset version avoids (Appendix F). Gold and Treasury contracts were tested during development and found to detract from performance.

Table 7 reports trade counts by asset for the final three-asset model.

Table 7: Trade counts and net PnL by asset in the final three-asset model (2019–2025)

Asset	Trades	Net PnL
NQ	405	\$361,429
ES	429	\$180,235
CL	381	\$34,888

The bar-level asset PnLs in Table 7 sum to \$576,552. The trade-level totals in Table 8 sum to \$577,460, despite identical asset trade counts. The \$908 difference is retained as an unresolved accounting discrepancy. The examined scripts use different quantity conventions (bar-varying versus entry-fixed); a reconciliation against one canonical position and cash-flow ledger is needed before either total is adopted as definitive.

The trade count distribution confirms that the strategy remains actively engaged across all three markets, though not equally. Differences reflect the distinct behavior of each contract, the interaction of breakout signals with volatility and momentum filters, and the different session lengths. This variation indicates that the model adapts to each asset class rather than mechanically forcing uniform activity.

Table 8 provides trade-level statistics that characterize the strategy’s microstructure.

Table 8: Trade-level statistics of the final three-asset model (2019–2025)

	Trades	Win Rate	Avg Win	Avg Loss	Profit Factor	Avg Hold	Total PnL
NQ	405	47.9%	\$4,617	\$-2,536	1.67	17 bars	\$360,504
ES	429	46.2%	\$2,553	\$-1,406	1.56	18 bars	\$180,699
CL	381	34.4%	\$1,373	\$-575	1.25	15 bars	\$36,257
Portfolio	1215	43.0%	\$3,023	\$-1,450	1.58	17 bars	\$577,460

Note: Win Rate is the percentage of trades with positive net PnL. Avg Win and Avg Loss are mean net PnL of winning and losing trades respectively. Profit Factor is total winning PnL divided by total losing PnL (absolute value). Avg Hold is the mean trade duration in 5-minute bars. All figures are net of transaction costs.

The portfolio win rate is 43.0%, meaning that fewer than half of all trades are profitable. Nevertheless, the strategy is profitable because the average winning trade (\$3,023) is approximately twice the average losing trade (\$1,450 in absolute terms), producing a profit factor of 1.58. This asymmetry is consistent with the trailing-stop design: losing trades are cut quickly by the ATR stop, while winning trades are allowed to run until the trailing stop or RSI exit triggers. The saved average holding count is 17 observed bars. It does not establish elapsed minutes or the share of trades closed within the entry session when timestamps contain gaps or entries occur near the session boundary. NQ generates the largest per-trade profits (average win \$4,617), consistent with its role as the strongest directional contributor. CL has the lowest win rate (34.4%) but remains net profitable through its regime-conditional sizing, which limits losses when signals are weakest.

5.4 Trade Quality by Volatility Regime

Table 9 describes executed trades by volatility regime. These are selected observations with different position sizes and asset composition, particularly because CL is excluded in its low regime. The table does not directly test whether conditioning improves a common set of candidate breakouts.

Table 9: Trade quality conditional on volatility regime at entry (2019–2025)

Regime	Trades	Win Rate	Avg PnL	Total PnL	Profit Factor	Avg Hold
Low	282	34.0%	\$427	\$120,546	1.62	16 bars
Medium	250	50.0%	\$632	\$157,992	1.70	18 bars
High	683	44.2%	\$438	\$298,922	1.51	16 bars

Note: Regime is classified using the lagged EGARCH conditional volatility percentile with 42nd/58th thresholds. Trades are assigned the regime prevailing at entry. CL regime caps restrict exposure to zero in low-volatility regimes, which accounts for the lower trade count in that category. All figures are net of transaction costs.

Medium-volatility trades have the highest win rate (50.0%) and profit factor (1.70), followed by high-volatility trades (44.2%, PF 1.51). Trade volume concentrates in the high-volatility regime (683 of 1,215

trades), consistent with the gate’s design. The medium-regime’s superior win rate suggests that breakout signals are most informative when volatility is moderate and expanding rather than already extreme, consistent with the vol-score tent function’s plateau across the 40th–80th percentile range.

Figure 6 illustrates the EGARCH gating mechanism in action for NQ during 2021, the strategy’s strongest year.

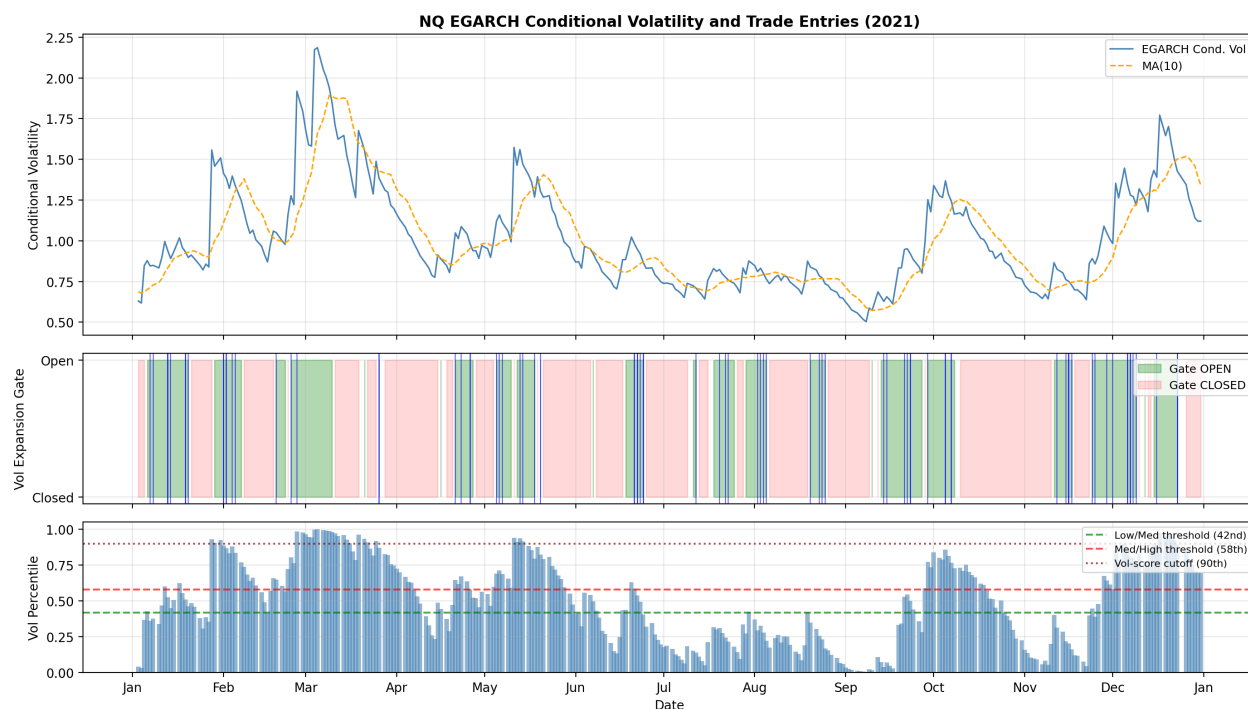


Figure 6: EGARCH volatility gate and trade entries for NQ during 2021. Top panel: daily EGARCH conditional volatility and its 10-day moving average. Middle panel: gate status (green = open, red = closed) with trade entry dates marked in blue. Bottom panel: volatility percentile rank with regime thresholds. The saved graphic reports 138 active daily observations out of 311 and 80 trades; the calendar and counting convention require verification.

The figure’s reported denominator of 311 daily observations in 2021 exceeds the conventional exchange trading-day count and may include additional dates from the input calendar. It is preserved as an artifact count, not asserted to be 311 exchange sessions. Its active-day fraction must be reconciled with the daily calendar before being interpreted economically.

5.5 Year-by-Year Analysis and Subperiod Performance

A further objective is to assess how the strategy behaves across different market environments. Table 10 reports year-by-year results.

Table 10: Year-by-year performance of the final three-asset model

Period	Trades	Sharpe	Sortino [†]	Max DD (%)	Return
Full evaluation (2019–2025)	1,215	1.412	1.579	-10.6	+288%
2019	195	0.916	1.241	-10.6	+13.2%
2020	111	0.887	0.805	-6.0	+9.9%
2021	229	2.622	3.713	-9.4	+85.3%
2022	82	0.201	0.162	-8.9	+2.9%
2023	210	2.270	3.672	-10.8	+85.9%
2024	223	1.870	2.107	-8.9	+74.5%
2025	164	0.458	0.450	-16.7	+16.6%

Note: The full row denotes the historical evaluation period, not an untouched holdout. Trades reports total completed trades. Max DD denotes maximum drawdown. Return is the annual strategy return (full evaluation row reports cumulative net return). Year-level ratios use the reported annualization convention. The annual trade counts sum to 1,214, versus 1,215 in the full row. Annual drawdowns can differ from full-path drawdowns when each year resets its capital base; the saved outputs require a common-calendar, common-equity reconciliation.

The saved annual table contains no negative calendar-year return, with the largest reported gains in 2021, 2023, and 2024. The smaller gains in 2020 and 2022 show substantial variation across years. Because later years were observed during the research process, their positive results do not independently demonstrate generalization beyond development.

Figure 7 summarizes the year-by-year returns of the final strategy.

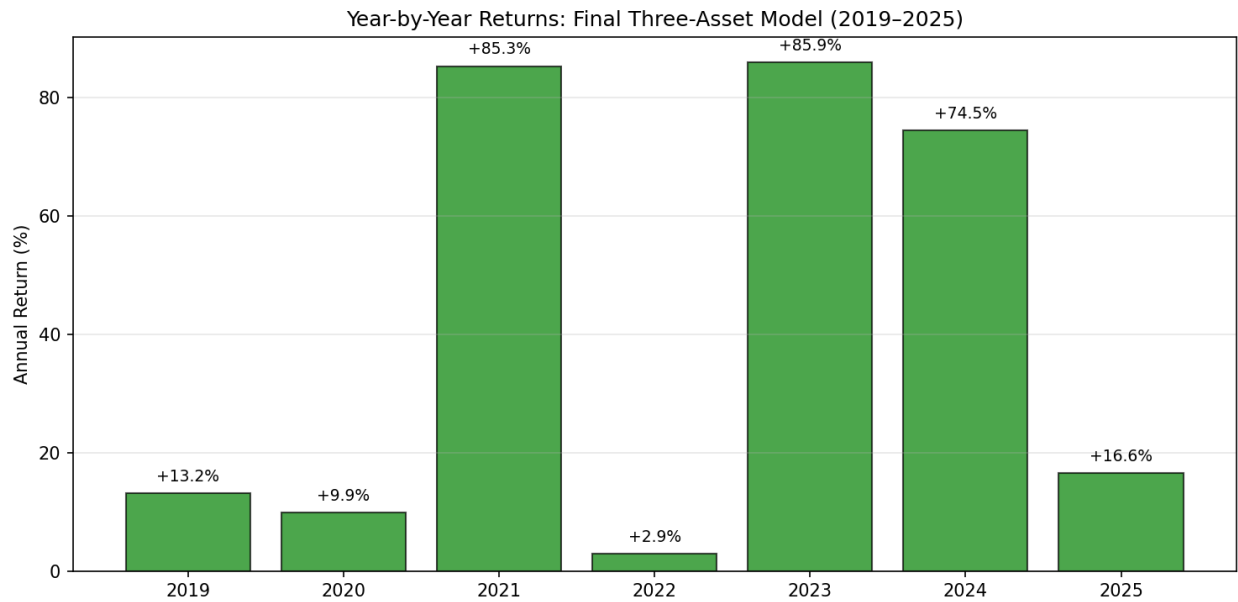


Figure 7: Year-by-year returns of the final three-asset model across the 2019–2025 evaluation period.

The pattern is consistent with the conditioning mechanism’s design: activity concentrates in states where the expansion gate identifies informative volatility conditions. The generality beyond the sample period remains untested.

5.6 Transaction Cost Sensitivity

Table 11 reports performance under cost multipliers ranging from $0.5\times$ to $8\times$ the baseline assumptions (\$14.51 round-turn for NQ, \$29.51 for ES, \$12.80 for CL, each treated as the stated fixed round-turn cost input).

Table 11: Cost sensitivity of the final three-asset strategy

Cost Mult.	Sharpe	Sortino [†]	Max DD	Return Mult.
0.5×	1.510	1.701	-9.4%	4.09×
1.0×	1.412	1.579	-10.6%	3.88×
2.0×	1.215	1.339	-13.1%	3.46×
4.0×	0.810	0.866	-21.9%	2.63×
8.0×	-0.022	-0.022	-64.1%	0.96×

Note: Base round-turn costs: NQ \$14.51, ES \$29.51, CL \$12.80. Cost multiplier scales all transaction costs proportionally.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

At twice baseline costs, Table 11 reports Sharpe 1.215, replacing the inconsistent 1.329 in the original prose. At eight times baseline, it reports -0.022 and terminal equity of 0.96 times initial capital. This

is sensitivity to a fixed per-contract cost model, not an execution-capacity study. It omits empirically calibrated impact and separate charges for quantity changes during a trade; the cost assumptions are therefore not characterized as conservatively validated market costs.

5.7 Factor Exposure Analysis

If the strategy's returns can be fully explained by common risk factors, the EGARCH mechanism adds no independent value. To investigate, the daily portfolio return series is regressed on two sets of risk factors.

The primary specification uses the Fama–French three-factor model (Fama and French, 1993), the standard benchmark for evaluating risk-adjusted returns in academic finance:

$$r_t^{\text{strategy}} - r_t^f = \alpha + \beta_1(\text{Mkt-RF})_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \varepsilon_t, \quad (12)$$

where factors are obtained from Kenneth French's Data Library at daily frequency. The strategy's daily return is computed as daily portfolio PnL divided by initial capital (\$200,000), and excess returns are formed by subtracting the risk-free rate. A secondary specification uses three self-constructed factors from the same futures data: an equal-weighted market return, a time-series momentum proxy (Moskowitz et al., 2012), and the daily change in 20-day realized volatility of ES (Moreira and Muir, 2017). Both specifications are estimated by OLS with heteroskedasticity-robust (HC1) standard errors.

Table 12: Factor exposure regressions of strategy daily returns

	(1) FF3	(2) Self-Constructed
Alpha	0.0014*** (3.34)	0.0014*** (3.56)
Mkt–RF	0.2095*** (6.09)	
SMB	0.0495 (0.73)	
HML	-0.1154*** (-3.14)	
Market		0.1105*** (5.86)
Momentum		-0.0145 (-0.96)
Volatility		0.8557*** (2.84)
R^2	0.0300	0.0574
Adj. R^2	0.0284	0.0558
N	1759	1739

Note: OLS regressions with heteroskedasticity-robust (HC1) standard errors. t -statistics in parentheses. Annualized alpha (Spec 1): 35.13%; (Spec 2): 36.48%. Fama–French factors from Kenneth French Data Library (daily). Self-constructed factors follow the existing pipeline: Market = equal-weighted futures return, Momentum = sign of trailing 5-day return $\times$ daily return, Volatility = $\Delta RV(20)$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 12 reports the results. The intercept (alpha) is highly significant in both specifications: 0.0014 per day ($t = 3.34$, $p < 0.001$) under Fama–French and 0.0014 per day ($t = 3.56$, $p < 0.001$) under self-constructed factors, corresponding to annualized alphas of approximately 35–36%. Both specifications explain only a small share of daily return variance ($R^2 = 3.0\%$ and 5.7% respectively), so most variation remains unexplained by those included linear factors. The HC1 significance is reported under that covariance assumption and is not a validation of time-series alpha.

The estimated market coefficient is 0.210 ($t = 6.09$), conditional on the stated factor model and PnL normalization. It is not a measure of the fraction of time invested or the percentage of account notional exposed. The SMB estimate is insignificant ($t = 0.73$); HML is negative (-0.115 , $t = -3.14$). These are in-sample linear associations. Low R^2 does not establish a causal volatility-gating mechanism, and HC1 standard errors do not adjust for serial correlation.

Under the self-constructed specification, the momentum loading is insignificant ($t = -0.96$), while the volatility loading is positive (0.856, $t = 2.84$). The result depends on these particular factor proxies and

does not show that the strategy is independent of momentum or equity risk. Interpretation of the intercept also depends on collateral interest: subtracting the risk-free rate from futures PnL that excludes interest is not the same as evaluating the excess return of a fully specified collateralized account.

5.8 Annual Summary and Calibration Provenance

Table 13 was originally labeled an expanding-window rank-recalibration exercise. However, the examined output-generation block computes annual metrics by slicing the saved strategy series; it does not itself refit percentile ranks using prior-year observations. Until an independent recalibration run and its inputs are linked to this table, it is treated as an annual historical summary. It cannot be cited as evidence that full-sample rank look-ahead has been removed.

Table 13: Saved annual summary (chronological recalibration provenance unresolved)

Year	Trades	Return
2019	195	+13.2%
2020	110	+9.9%
2021	229	+85.3%
2022	83	+2.9%
2023	210	+85.9%
2024	223	+74.5%
2025	164	+16.6%

Note: The saved output contains annual slices of the historical champion series. Prior-data rank recalibration is not verified. Counts differ from the main annual table in 2020 (110 versus 111) and 2022 (83 versus 82), while both annual tables sum to 1,214.

Revision note: the examined generation block slices saved annual returns and does not verify the claimed prior-year rank recalibration.

The saved annual summary reports positive results across its seven years. This describes the existing series and does not validate a new chronological calibration procedure.

5.9 EGARCH versus Alternative Volatility Filters

The ablation in Section 5.2 shows that removing volatility conditioning entirely causes severe deterioration. A sharper question is whether EGARCH *specifically* matters, or whether any volatility filter would produce similar improvements. Table 14 addresses this by replacing EGARCH with two simpler alternatives while holding all other components fixed.

Table 14: Performance comparison across volatility filter specifications

Volatility Filter	Sharpe	Max DD	Return Mult.	Trades	Neg. Yrs
EGARCH (final)	1.412	-10.6%	3.88×	1215	0
RV(20)	1.114	-10.1%	3.52×	1512	2
RV(60)	0.866	-16.6%	3.06×	1421	2
No filter	0.664	-29.3%	4.65×	3842	0

Note: All specifications use identical signal logic (breakout, RSI, EMA) and position sizing. RV(20) and RV(60) replace EGARCH conditional volatility with 20- and 60-day rolling standard deviation of daily log returns. No filter removes all volatility conditioning.

Two alternatives replace EGARCH with rolling standard deviations of daily returns: RV(20) and RV(60). These differ from intraday realized variance constructed by summing squared high-frequency returns. The reported Sharpes are 1.114 for RV(20), 0.866 for RV(60), 0.664 unfiltered, and 1.412 for EGARCH. The differences describe whole volatility inputs, including their persistence, smoothing, and calibration. They do not isolate an asymmetry coefficient or prove that EGARCH is the best admissibility model.

5.10 Placebo and Falsification Tests

The complete-layer comparison in Section 5.2 changes gate and sizing together. The following timing and randomization exercises vary the expansion gate within the selected remaining architecture. They ask whether its observed timing performs differently from particular alternatives, conditional on the same historical calibration.

The first is a gate-timing placebo that shifts the expansion gate forward and backward by $\{1, 3, 5, 10\}$ trading days. If the mechanism argument is correct, the correctly timed gate should outperform all mistimed versions.

Table 15: Gate-timing placebo test. The EGARCH expansion gate is shifted forward (positive) or backward (negative) by the indicated number of trading days. Shift = 0 is the correctly timed gate used in the thesis. Negative shifts use future information and are included as a anticipatory sensitivity check. All other parameters are held at champion values. Evaluation period: 2019–2025.

Gate Shift (days)	Sharpe	Sortino [†]	Max DD (%)	Return Mult.	Trades
-10	0.859	1.033	-15.2	2.57×	1472
-5	-0.155	-0.145	-48.3	0.77×	1091
-3	0.438	0.460	-17.9	1.77×	1028
-1	1.297	1.509	-11.9	3.52×	1120
+0 ←	1.412	1.579	-10.6	3.88×	1215
+1	0.884	0.938	-14.1	2.78×	1264
+3	0.502	0.477	-20.0	1.92×	1314
+5	0.205	0.181	-29.6	1.33×	1236
+10	0.309	0.324	-23.0	1.58×	1325

Note: Negative shifts allow the gate to use future volatility information (look-ahead); positive shifts delay the gate signal beyond its original 1-day lag. The correctly timed gate (shift = 0) is marked with an arrow.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The unshifted gate reports Sharpe 1.412; a one-day delay reports 0.884 and a five-day delay 0.205. A future-information shift of −1 day reports 1.297. Poor performance under a particular anticipatory shift does not prove that the unshifted pipeline is free of look-ahead in estimation or calibration. The result establishes sensitivity to gate alignment within the historical specification.

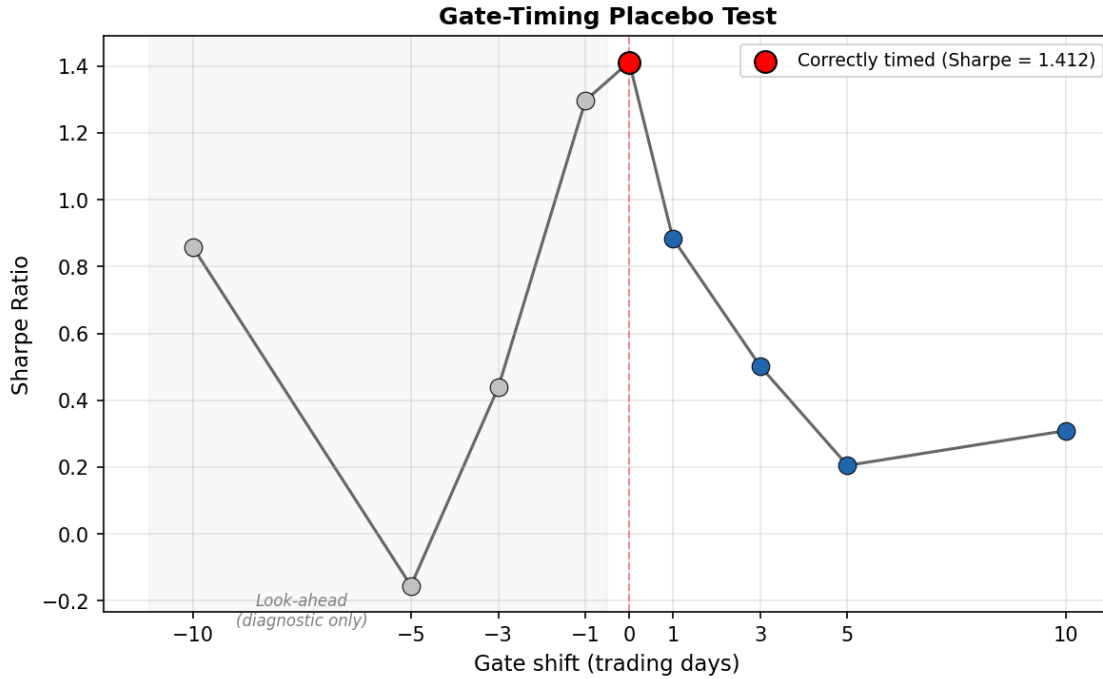


Figure 8: Gate-timing placebo test: Sharpe ratio under shifted EGARCH expansion gates. Shift = 0 is the correctly timed gate. Positive shifts delay the signal; negative shifts (shaded) use future information and are included as an anticipatory sensitivity check. The decay under positive shifts describes sensitivity to timing within the saved pipeline; it is not a test of all sources of information leakage.

The second placebo is stronger against the trade-suppression objection. Within each asset and calendar year, gate-active days are randomly reassigned while preserving the exact number of active days. This holds gate frequency fixed while destroying the economic content of the signal, though it does not preserve finer serial dependence such as within-month clustering.

Table 16: Randomized frequency-preserving gate placebo. For each asset and calendar year, the EGARCH gate is randomly reassigned across trading days while preserving the exact number of gate-active days. This keeps asset-year gate frequency fixed while removing the economic content of the gate. All other parameters remain at champion values. Evaluation period: 2019–2025.

Specification	Sharpe	Max DD (%)	Return Mult.	Trades
Actual EGARCH gate	1.412	-10.6	3.88×	1215
Randomized placebo mean	0.561	-20.2	1.93×	1328
Randomized placebo median	0.573	-18.8	1.94×	1327
Randomized placebo 95th pct.	0.953	-11.3	2.62×	1372
Best randomized placebo	1.191	-7.7	3.16×	1314

Note: Based on 500 placebo simulations with random seed 42. One-sided empirical placebo p-value, computed as $(k+1)/(N+1)$ with k placebo Sharpe ratios at least as large as the actual Sharpe, is 0.002. Here $k = 0$ and $N = 500$. The actual Sharpe lies at the 100.0th percentile of the placebo distribution.

Table 16 shows that the randomized placebo distribution lies far below the actual specification. Across 500 simulations, the mean placebo Sharpe is 0.561 and the 95th-percentile placebo Sharpe is 0.953, versus 1.412 for the actual gate. None of the 500 randomizations match or exceed the actual Sharpe, implying a one-sided empirical placebo p -value of 0.002. The smoothed empirical tail fraction is $(0 + 1)/(500 + 1) \approx 0.002$. This comparison rejects only the particular randomization scheme, conditional on the selected model. It changes serial clustering and may change asset synchrony and downstream trade opportunities; it does not rule out other structured gates or correct for selecting the design after observing outcomes.

5.11 Return Distribution and the Sharpe–Sortino Gap

At daily frequency the historical negative-return-standard-deviation ratio is 1.579, compared with Sharpe 1.412. At five-minute frequency the corresponding ratio is 0.605 and Sharpe 1.555. These values use different aggregation and annualization conventions, and the former ratio is not the conventional Sortino statistic (Section 4.7). A negative bar return is a price change over that bar, not necessarily a dip below the trade’s entry price. Figure 9 shows the saved daily distribution.

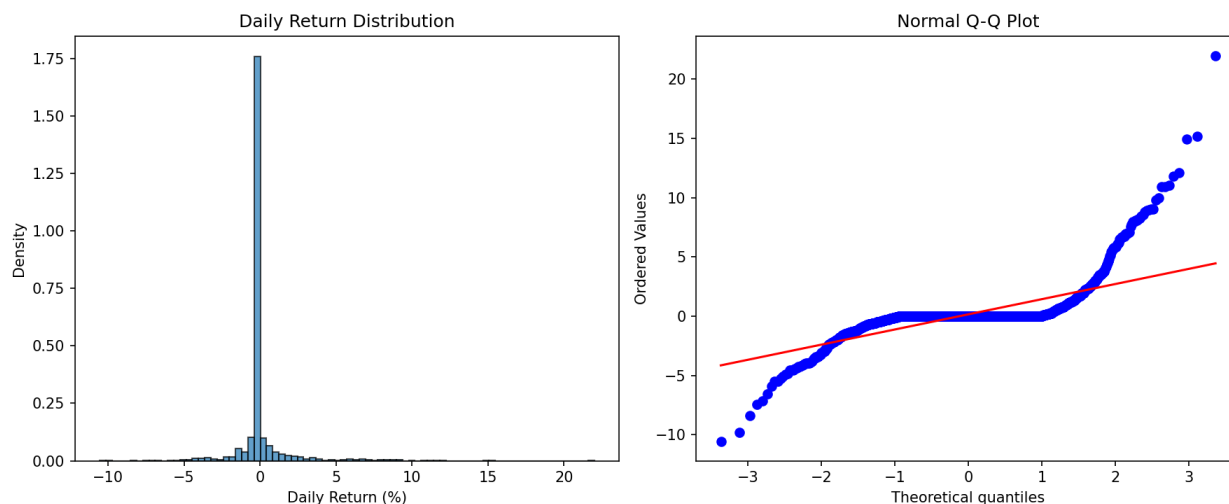


Figure 9: Daily return distribution of the final model (left) and normal Q-Q plot (right). The distribution exhibits positive skewness and substantial excess kurtosis, indicating fat tails and occasional large gains. Evaluation period: 2019–2025.

The saved daily distribution has positive skewness, heavy tails, and a substantial mass at zero. The gain/loss graphic reports 295 positive days and 321 negative days, so loss days are more frequent among its 616 active observations. The distribution and the negative-return ratio are descriptive diagnostics under the saved accounting convention.

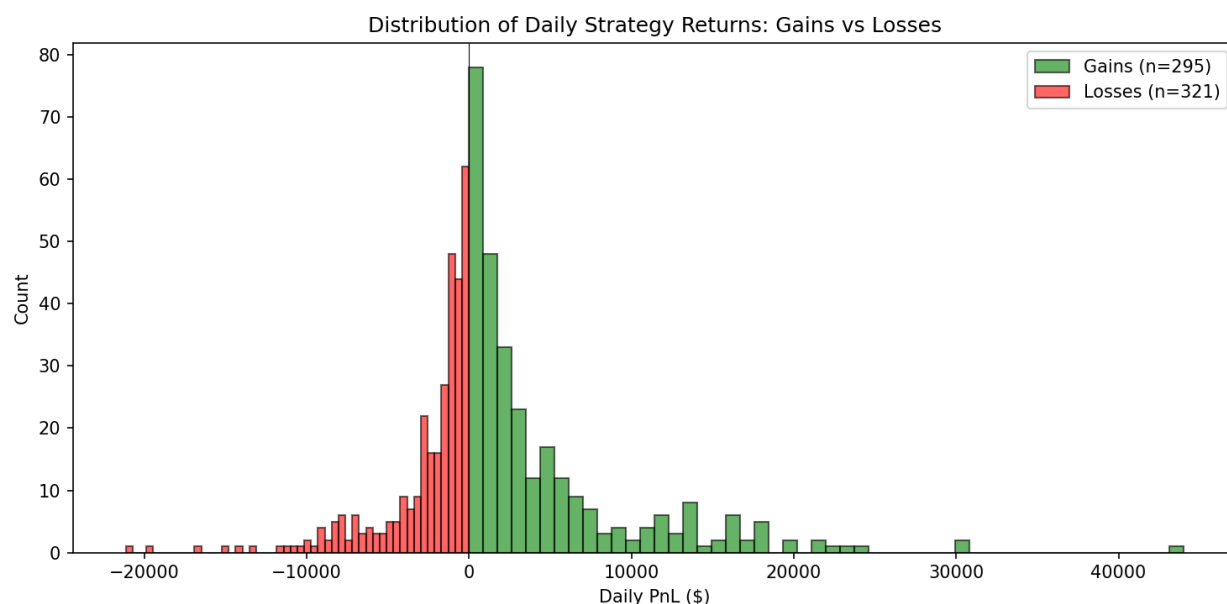


Figure 10: Distribution of daily strategy returns separated into gain and loss days. The positive mean, positive skewness, and fat-tailed loss distribution are consistent with the daily Sortino ratio of 1.579 exceeding the daily Sharpe ratio of 1.412. Evaluation period: 2019–2025.

Daily aggregation nets positive and negative intraday PnL, changing the distribution of negative observations. This can change a conditional-negative-return standard deviation substantially; it does not validate the conventional Sortino interpretation or establish low tail risk. Table 17 preserves the original frequency diagnostics. The five-minute factor $\sqrt{78 \times 252}$ also requires justification for this simulation’s actual observation calendar and entry windows.

Table 17: Sharpe and Sortino ratios at 5-minute and daily frequencies (champion specification, 2019–2025)

Return frequency	Sharpe	Sortino [†]	Sharpe/Sortino	Observations
5-minute bar-level cross-check	1.555	0.605	2.57	140,376
Daily (all days)	1.412	1.579	0.89	1,805
Daily (active days only)	2.435	4.628	0.53	616

Note: Historical “Sortino” values use the sample standard deviation of negative returns, not target downside deviation. Five-minute annualization uses $\sqrt{78 \times 252}$ and daily annualization $\sqrt{252}$. Active-day ratios are conditional diagnostics, not calendar-time performance. The daily count is supported by the archived 1,805-date champion return file.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The daily Sharpe convention is retained for consistency with the main saved results. Active-day-only annualization discards zero-PnL days and is not a calendar-time performance measure. Neither a ratio above one nor the absence of a negative calendar year guarantees limited future downside. Drawdown, tail losses, overnight exposure, and margin capacity require separate assessment.

5.12 Parameter Stability

A strategy whose performance collapses under small parameter perturbations is likely overfit to the specific values chosen. To assess this risk, Table 18 reports the Sharpe ratio obtained when each key parameter is perturbed individually across a range of plausible values, while all other parameters are held at their champion settings.

Table 18: Parameter stability analysis: Sharpe ratio under single-parameter perturbation

Parameter	Champion	Champion Sharpe	Sharpe Range	Mean Sharpe
breakout_window	60	1.417	1.292–1.417	1.343
atr_stop	2.2	1.392	1.352–1.392	1.379
ema_fast	15	1.400	1.351–1.400	1.371
ema_slow	40	1.415	1.328–1.415	1.362
rsi_long	58	1.412	1.327–1.412	1.371
cooldown	5	1.412	1.373–1.415	1.401
ivt_ma	40	1.412	1.375–1.422	1.401
scale L	5.0	1.412	1.412–1.412	1.412

Note: Each row perturbs one parameter across its tested range while holding all others at champion values. Sharpe Range shows the minimum and maximum annualized Sharpe across the perturbation grid. A narrow range indicates the result is not sensitive to the exact parameter choice.

Revision note: the saved row-specific “Champion Sharpe” values are not all 1.412; their baseline or parameter-overwrite conventions are unresolved. These rows cannot be treated as a single exactly matched baseline until the generation provenance is reconciled.

The results show broad stability across all parameters. The narrowest range is the saved contract-count scaling exercise, which rescales its return stream proportionally; this does not establish invariance under live contract, cost, or margin constraints. Among the signal and trade-management parameters, the widest variation occurs for breakout window (1.292–1.417) and EMA slow period (1.328–1.415), both of which remain well above 1.0 at all tested values. No parameter exhibits a sharp spike at the champion value with collapse on either side: the performance surface is a plateau rather than a peak.

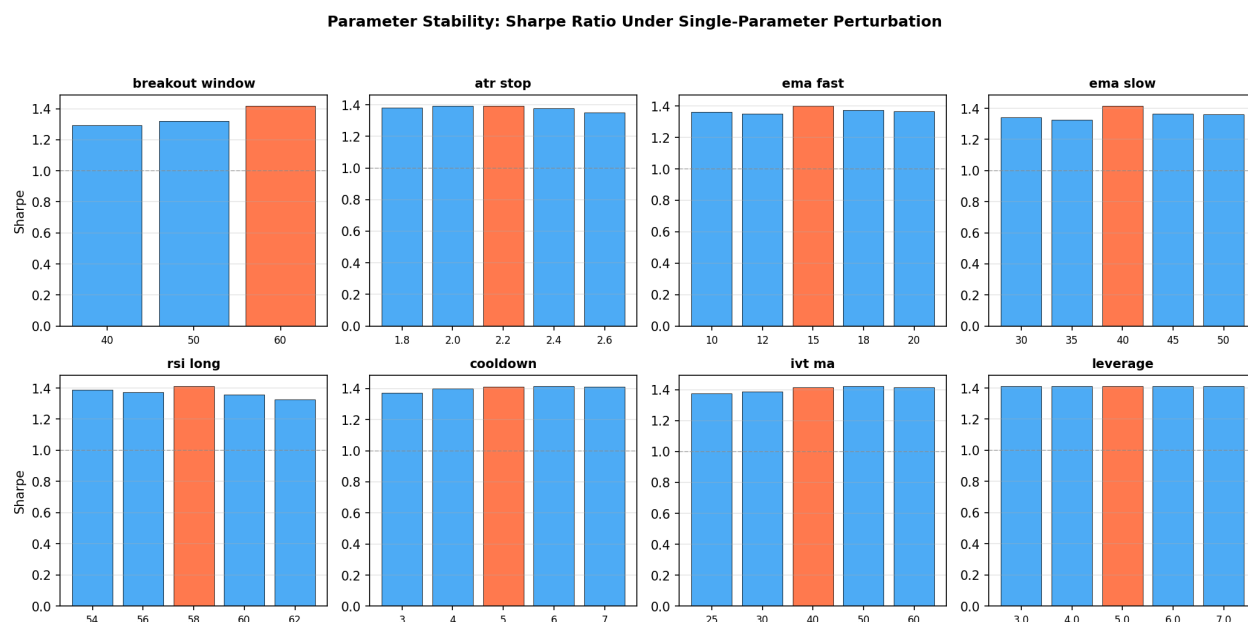


Figure 11: Sharpe ratio under single-parameter perturbation. Each panel varies one parameter while holding all others at champion values. The champion value is highlighted in red. The dashed line marks Sharpe = 1.0. Evaluation period: 2019–2025.

Figure 11 visualizes this stability. The absence of cliff edges suggests that the champion parameter set does not sit on a fragile optimum. However, one-dimensional perturbations cannot rule out joint sensitivity: two parameters might individually appear stable while interacting adversely when moved simultaneously.

Joint parameter stability. To address this concern, Figures 12–14 present two-dimensional heatmaps that vary pairs of economically related parameters simultaneously while holding all others at champion values. Each cell reports the portfolio Sharpe ratio over the full 2019–2025 evaluation period.

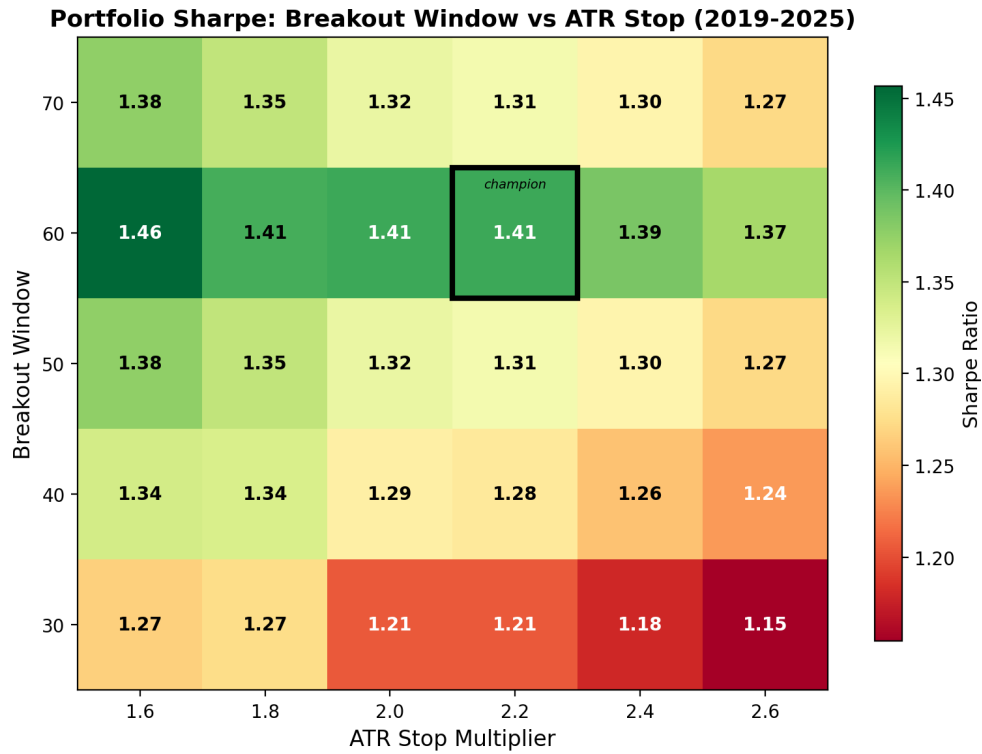


Figure 12: Joint sensitivity: breakout window $\times$ ATR stop multiplier (NQ/ES). Champion cell marked. All 30 combinations yield Sharpe > 1.15 . Range: 1.15–1.46.

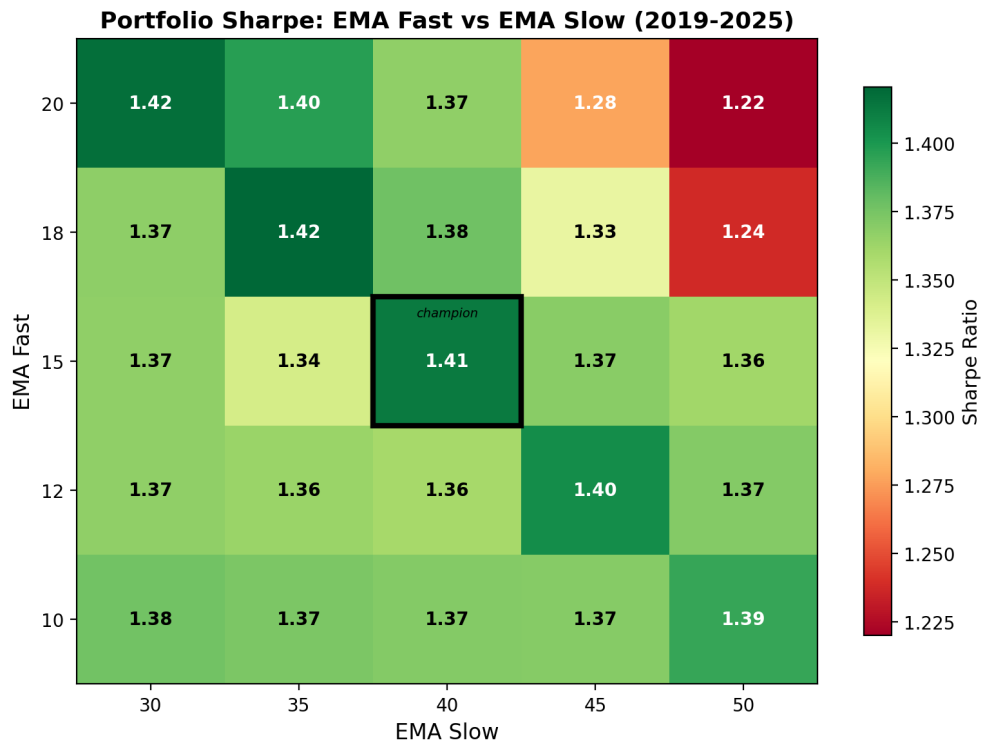


Figure 13: Joint sensitivity: EMA fast $\times$ EMA slow (NQ/ES). Champion cell marked. All 25 combinations yield Sharpe > 1.22 . Range: 1.22–1.42.

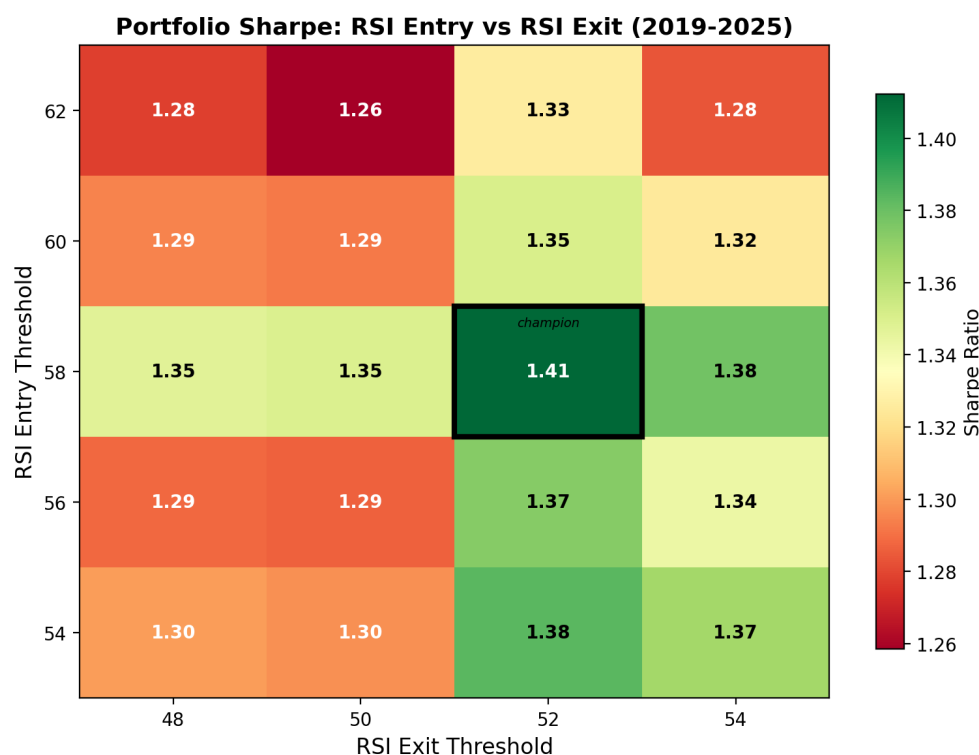


Figure 14: Joint sensitivity: RSI entry threshold $\times$ RSI exit threshold. Champion cell marked. All 20 combinations yield Sharpe > 1.26 . Range: 1.26–1.41.

All 75 reported combinations across the three heatmaps have Sharpe above 1.15, supporting local stability within the tested ranges. The selected cell is not the maximum in every surface, but this does not rule out selection of the architecture, asset universe, ranges, or reporting windows. Local smoothness and absence of a sharp optimum are useful sensitivity evidence, not evidence that the full strategy-development process was free of overfitting.

5.13 Deflated Sharpe Ratio

The parameter stability analysis addresses whether nearby parameter values yield similar results. A separate concern is selection bias: when many strategy variants are tested on the same evaluation period, the best-performing specification is expected to overstate true prospective Sharpe simply by chance. The Deflated Sharpe Ratio (DSR) of Bailey and López de Prado (2014) adjusts for this effect by comparing the observed Sharpe ratio against the expected maximum Sharpe under the null hypothesis that the tested strategies have zero true Sharpe. The Bailey–López de Prado formulation scales the expected maximum standard-normal draw by the cross-sectional dispersion of the Sharpe ratios across the tested family, which captures how far apart the candidate strategies actually are in performance space.

Table 19: DSR formula sensitivity conditional on archived inputs and a local trial family

Assumed trials N	Expected maximum m_N	Local dispersion	Unit dispersion
6	1.300	> 99.9%	63.369%
15	1.771	> 99.9%	13.742%
30	2.073	> 99.9%	2.199%
50	2.276	> 99.9%	0.424%

Note: This is a formula-only correction using the archived annualized Sharpe 1.4122771, annualized standard error 0.3282149, and annualized local-family dispersion 0.0322980. The saved summary and daily series contain 1,805 dates, replacing the original DSR note’s unsupported 1,762. The formula uses Equation (2) of Bailey and López de Prado (2014). The local dispersion comes from 38 nearby variants, not the complete research search. Trial count N and the standard-error construction are assumptions; these percentages are conditional sensitivities, not calibrated probabilities of prospective profitability or a verified full-search correction.

The cited expected-maximum approximation is

$$m_N = (1 - \gamma_E)\Phi^{-1}(1 - 1/N) + \gamma_E\Phi^{-1}(1 - 1/(Ne)),$$

where γ_E is the Euler–Mascheroni constant. The table computes $\Phi((SR - \sigma_{SR}m_N)/SE)$ using the archived unrounded inputs, and substitutes $\sigma_{SR} = 1$ only in the unit-dispersion column. All quantities in that quotient are in annualized Sharpe units. This replaces the original expected-maximum approximation and its corresponding numerical sensitivities; no empirical model or return series was rerun.

The small local dispersion, $\sigma_{SR} \approx 0.032$, comes from 38 nearby one-parameter perturbations sharing an already selected architecture. It omits the broader research search over assets, direction, sessions, volatility models, score shapes, and other iterations. High local correlation does not make the broader search equivalent to one independent trial. The table therefore cannot establish that the overall data-mining concern is small.

The unit-dispersion column is an illustrative numerical sensitivity. Unit dispersion and trial independence are different assumptions; neither column is a proven bound on the selection bias of the complete research history. The saved sample and annualized units are now identified, but serial-dependence treatment, trial-universe completeness, and the effective independent trial model still need justification before interpreting the percentages as calibrated probabilities.

Selection affects both performance levels and contrasts when variants or comparisons are chosen after observing their outcomes. Accordingly, the ablation and placebo results are not immune to the number and scope of research iterations. The retrospective training-window exercise supplies additional evidence but does not erase architecture selection. A full-search trial inventory and an untouched, preregistered evaluation are needed for stronger inference.

5.14 Retrospective Temporal Split

Table 20 divides the saved evaluation returns into 2019–2023 and 2024–2025. The original thesis described the later interval as a pure holdout, but elsewhere records refinement over 2019–2025. Without a dated specification freeze showing that no design or reporting decision used the later interval, this is a retrospective temporal split. The historical label “holdout” in the source output does not establish independence.

Table 20: Development period vs. later-period split performance

Period	Sharpe	Sortino [†]	Max DD	Return Mult.	Trades
Earlier (2019–2023)	1.594	2.053	–10.6%	2.97×	828
Later (2024–2025)	1.197	1.266	–13.4%	1.91×	387
Full period (2019–2025)	1.412	1.579	–10.6%	3.88×	1215
Later No-EGARCH	0.449	1.016	–51.7%	1.97×	1146

Note: Retrospective split of the saved development-exposed specification. A dated freeze proving the later period unused for design has not been established. All rows use the same reported parameter set and \$200,000 initial capital for the corresponding period. Period-specific equity resets affect drawdown comparability.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The later-period reported Sharpe is 1.197, compared with 1.594 in 2019–2023. The later No-EGARCH variant reports Sharpe 0.449 and drawdown 51.7%, versus terminal equity multiples of 1.97 and 1.91 for unfiltered and filtered versions. These figures describe a repeated historical contrast. They do not establish that the development effect survives on untouched data.

The difference between these period Sharpes combines changes in market conditions, sample uncertainty, and possible selection effects. It cannot quantify the cost of parameter selection or justify a prospective Sharpe range of 1.0–1.2. The next exercise restricts the trading-parameter selection window, which addresses a different aspect of the research design.

5.15 Retrospective Pre-2019 Parameter Selection

Table 21 reports a retrospective exercise selecting trading parameters by sequential search over 2010–2018 and applying the selected values to 2019–2025. The search-window restriction is distinct from the timing of architecture development, fitted-volatility estimation, and percentile calibration. It therefore cannot be described as eliminating every form of later-period exposure.

Table 21: Retrospective parameter selection: training on 2010–2018, evaluation on 2019–2025

Model	Sharpe	Sortino [†]	Max DD	Return Mult.	Trades
Earlier-selected + EGARCH	1.195	1.314	−14.0%	3.36×	1302
Champion + EGARCH	1.412	1.579	−10.6%	3.88×	1215
Earlier-selected + No-EGARCH	0.382	0.673	−49.9%	2.88×	4079
Champion + No-EGARCH	0.664	1.261	−29.3%	4.65×	3842

Note: The retrospective trading-parameter search uses 2010–2018 observations. This restriction does not establish independence of the architecture, volatility coefficients, or calibration distributions from later data. The champion is development-exposed. Ratios use the reported daily convention and \$200,000 initial capital; transaction-cost inputs are shared.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The pre-OOS selected specification achieves a Sharpe ratio of 1.195 on the 2019–2025 evaluation period, with a terminal equity multiple of 3.36× on \$200,000 initial capital, approximately 15% below the champion’s 1.412. The parameter choices are structurally similar: the pre-OOS search selects the same breakout windows (60 bars for NQ/ES, 30 for CL), the same session hours, and comparable IVT smoothing, differing primarily in slightly tighter ATR stops and a marginally stricter RSI entry threshold. This describes similarity among selected parameter values, conditional on an already chosen architecture; it does not establish architectural generalization or separate refinement from market-period effects.

The reported Sharpe of 1.195 is a historical sensitivity result under a restricted trading-parameter search. It is 0.217 below the development value of 1.412, approximately 15.4% of that value. This arithmetic difference is not an estimate of prospective overstatement or the causal cost of refinement.

The training-period Sharpe on 2010–2018 is only 0.066. This near-zero result is material negative evidence about stability across periods and remains so even if the strategy is state-dependent. The reported gate-active rates are broadly similar across the earlier and later windows: 41.97%, 43.18%, and 42.09% for NQ, ES, and CL before 2019, versus 40.03%, 39.81%, and 38.13% later. Similar firing rates do not establish comparable signal quality. The original comparison of annual buy-and-hold returns (14.2%, 9.2%, and −2.4% earlier; 25.8%, 17.0%, and 6.9% later) describes unconditional drift, not drift estimated specifically on gate-active days. It cannot explain away the weak training Sharpe. The saved later-period ablation changes Sharpe from 1.195 to 0.382 and reports 49.9% drawdown, which is supportive historical comparative evidence but not an independent causal test.

5.16 Rolling Parameter Re-Selection

A rolling search complements the single training-window exercise by reselecting a subset of trading parameters annually. It uses past-window observations for that restricted search, while retaining the broader development-selected architecture. This is a partial rolling validation, not a fully independent evaluation of every design choice.

The implementation re-selects parameters annually on a 24-month rolling training window. For each evaluation year $y \in \{2012, \dots, 2025\}$, the training window is $[y - 2, y - 1]$. A small grid is searched per asset over the breakout window and ATR stop multiplier (the two parameters identified as most consequential by the parameter-stability analysis), with all other parameters held at their champion values. The combination that maximizes training-window Sharpe is selected and applied to year y . Year- y PnL is recorded, and the procedure rolls forward. Only the searched breakout and ATR-stop values are selected using the prior two years. The examined rolling-search script still uses full-data volatility ranks and regime quantiles, and a sizing anchor based on 2010–2018 even for tests within that earlier interval. The remaining champion settings and candidate grid also come from development. The saved 14-year result is therefore retrospective, not an end-to-end chronological backtest.

Table 22: Walk-forward analysis. Annual re-selection of breakout window and ATR stop on a rolling 24-month training window, with selected parameters applied to the next calendar year. All other parameters fixed at champion values. Comparison against the champion specification with fixed parameters over the same evaluation years.

Specification	2012–2018	2019–2025	2012–2025 (full)
Champion (fixed parameters)	−0.01	1.418	0.962
Walk-forward (annual re-fit)	0.376	1.283	0.939

Note: Annualized Sharpe ratios computed on daily-aggregated portfolio returns. The walk-forward design grids over breakout window and ATR stop multiplier per asset: $\{45, 60, 75\}$ bars and $\{1.8, 2.2, 2.6\}$ for NQ and ES, $\{20, 30, 40\}$ bars and $\{2.0, 2.5, 3.0\}$ for CL. The evaluation set comprises 14 calendar years (2012–2025); 2010–2011 is reserved as the initial training window. The walk-forward track records one negative calendar year (2025: −4.6%) in 14 years; the champion fixed records three (all in 2012–2018) over the same span.

Over 2012–2025 the saved rolling-search Sharpe is 0.939 versus 0.962 for the fixed-parameter comparator. This proximity shows limited sensitivity to these two parameter choices within the saved experiment. No paired uncertainty calculation is provided for the 0.023 difference, and the similarity does not establish that the complete strategy is independent of development choices.

Second, the comparison decomposes by subperiod in an informative way. Over 2012–2018, the walk-forward Sharpe (0.376) materially exceeds the champion fixed (−0.01) in this saved retrospective comparison. That difference does not isolate the effect of adaptation because calibration and other design choices are not held to an entirely historical information set. Over 2019–2025, the comparison reverses (1.283 for walk-forward versus 1.418 for champion), as expected since the champion has the advantage of having been refined on that window. Neither subperiod is implausible. The walk-forward records zero negative calendar years in 2012–2018 against the champion’s three, while in 2019–2025 the walk-forward has one negative year (2025) against the champion’s zero. The aggregate 14-year track contains one negative year out of fourteen.

The exercises answer different retrospective questions: one selects trading parameters on an earlier fixed window, and the other reselects two parameter dimensions annually. Neither freezes the complete research architecture before all evaluation data. Their results motivate an end-to-end chronological validation with all model fitting, calibration, search, and execution conventions fixed in advance.

5.17 Interpretation of the Main Findings

Pulling the empirical results together, the evidence supports three claims. The first is the one the thesis rests on most heavily, and the others are a step more interpretive. I separate them this way because not all parts of the argument are equally strong, and it seems better to be explicit about that than to present them as if they carried the same weight.

Claim 1: the saved volatility-layer comparisons favor the filtered strategy within this design. The development comparison reports Sharpe 1.412 versus 0.664; the retrospective earlier-selection comparison reports 1.195 versus 0.382. Only the former is associated with the reported 0.0038 bootstrap tail fraction. Timing and randomization checks provide complementary evidence conditional on the selected pipeline.

Claim 2: restrictions on parameter selection are informative but incomplete. The reported 1.195 Sharpe arises from a retrospective earlier-window search. Development history and future-informed volatility calibration prevent interpreting it as a clean prospective estimate. The 0.066 training Sharpe is evidence of substantial period dependence.

Claim 3: The mechanism is bounded rather than universal. Returns concentrate in years with expanding but not extreme volatility (2021, 2023, 2024), while the EGARCH gate limits participation in adverse years (2020, 2022). This domain specificity is consistent with the economic channels described in Section 6.1, but the seven-year sample does not contain enough distinct regimes to map the boundaries precisely.

6 Discussion

This section interprets the empirical findings in terms of the economic mechanism they imply, identifies the conditions under which the mechanism should and should not operate, and assesses the boundaries of what the evidence can support.

6.1 The Economic Mechanism: Why Expanding Volatility Improves Breakout Quality

The ablation results show that EGARCH conditioning matters for performance, but by themselves they do not explain the mechanism. This subsection sketches the reasons I think are most plausible. The explanation is not something the thesis can prove directly, but several ideas from the finance literature point in the same general direction. Taken together, they suggest a sensible reason why periods of expanding volatility may improve the quality of some intraday breakout signals, while other volatility states do not.

Information arrival and price discovery. The mixture-of-distributions hypothesis (Andersen and Bollerslev, 1997) links volatility to the rate of information arrival: periods of expanding volatility correspond to elevated rates of news incorporation into prices. A breakout that coincides with increasing information flow is more likely to reflect genuine price discovery, the sequential revelation of information through order flow, than a mechanical crossing of a stale price level. In a low-information environment, the same threshold crossing is more likely to reflect a transient liquidity gap or the execution of a single large order against thin depth. The expansion gate is intended as an indirect proxy for such conditions. Without observations of news or order flow, the thesis cannot verify that gate-active days have higher information arrival or more directional content.

Liquidity transition and mean-reversion weakening. As volatility rises above its recent average, market-makers respond by widening bid-ask spreads and reducing the depth they post at each price level. This response weakens the mean-reversion force that ordinarily kills breakouts in calm markets: in low-volatility states, the resting depth in the order book absorbs breakout momentum, causing price to revert; during the transition to expanding volatility, this depth thins, and the breakout faces less resistance. The implication is that a breakout during expanding volatility has a higher probability of leading to directional continuation because the microstructure conditions that normally suppress continuation are temporarily weakened. This hypothesized channel motivates comparing expansion and level measures. The saved comparisons do not directly measure depth or identify which microstructure process drives performance.

Participation and order flow. A further hypothesis is that changes in market participation amplify some breakouts when volatility expands. The thesis contains no order-level data with which to test this channel. The options-market evidence of Bollen and Whaley (2004) concerns implied volatility and order flow; it is not direct evidence that EGARCH gates identify momentum-amplifying futures flows.

The three channels work in the same direction rather than in tension with each other. Higher information arrival provides a reason for price to move; thinner resting depth removes part of the force that normally pushes it back; and the increase in discretionary and algorithmic participation during active days helps the move persist once it starts. All three effects are, however, tied to the current day's volatility environment, which is why both placebo exercises in Section 5.10 matter: mistiming the gate degrades performance sharply, and randomizing gate days while holding frequency fixed produces much weaker outcomes. These channels remain hypotheses. The next subsection distinguishes the model implication of asymmetry from the evidence required to test its incremental trading value.

6.2 Why Asymmetry Matters for a Long-Only Framework

The estimated EGARCH asymmetry coefficients are negative for all three contracts. Holding the previous variance and shock magnitude fixed, a negative standardized innovation produces a larger modeled variance response than an equally sized positive innovation. This property motivates a hypothesis about exposure after declines; it does not prove better long-only trading performance than a symmetric model.

For otherwise equal lagged variance and shock magnitude, $\gamma < 0$ raises the modeled variance response to a negative shock relative to a positive one. Under the score's upper-tail taper, a sufficiently high variance rank can reduce exposure. That is a mathematical channel in the specified model. Whether declines actually create worse breakout opportunities, whether the expansion gate counteracts this reduction, and whether a symmetric alternative performs differently are empirical questions not settled by this argument.

EGARCH reports Sharpe 1.412 versus 1.114 for RV(20), a difference of 0.298 in the selected historical design. These inputs differ in estimation, persistence, smoothing, and distributional treatment as well as asymmetry. Consequently, the comparison cannot attribute the gap specifically to γ . A matched $\gamma = 0$ EGARCH experiment under the same chronological protocol is needed.

Table 23 describes the score distribution on gate-active days split by the previous innovation's sign. It is a diagnostic of the fitted model's behavior, not a controlled test of an economic mechanism. For negative-minus-positive differences, the proposed model pattern is positive for volatility percentiles and upper-tail shares but negative for volatility scores.

Table 23: Residual-sign diagnostic: vol-score distribution on gate-active days, split by sign of the EGARCH standardized residual on day $t - 1$ (2019–2025 evaluation window)

Asset	γ	$\Delta \overline{\text{vol_pct}}$	$\Delta \overline{\text{vol_score}}$	$\Delta \%_{>0.80}$ (pp)	$\Delta \%_{>0.90}$ (pp)
ES	−0.174	+0.106	−0.129	+15.4	+12.8
NQ	−0.145	+0.091	−0.129	+15.5	+13.0
CL	−0.050	+0.031	−0.041	+3.3	+4.8

Note: All entries are differences (negative-residual bucket minus positive-residual bucket), so the proposed pattern is positive for percentile and tail-share differences but negative for score differences. The standardized residual on day $t - 1$ is approximated by $r_{t-1}/(\hat{\sigma}_{t-1}/100)$, where r_{t-1} here is a decimal daily log return (not the percentage return used in Section 4.1) and $\hat{\sigma}_{t-1}$ is in percentage units. This proxy omits the estimated conditional mean and is not exactly the fitted standardized innovation. The saved gate-active counts are 885 for ES (POS: 493, NEG: 390), 890 for NQ (POS: 517, NEG: 372), and 850 for CL (POS: 431, NEG: 416). The sign buckets leave 2, 1, and 3 observations respectively unclassified; their zero or missing-value treatment requires reconciliation. $\Delta \%_{>0.80}$ is the difference in the share of gate-active days falling into the vol-score suppression zone ($\text{vol_pct} > 0.80$, where exposure rolls off linearly); $\Delta \%_{>0.90}$ is the difference for the zero-exposure zone ($\text{vol_pct} > 0.90$).

The pattern lines up with the predicted direction across every metric and tracks $|\gamma|$ across assets. Gate-active days following negative residuals have higher mean vol_pct than gate-active days following positive residuals (+0.106 for ES, +0.091 for NQ, +0.031 for CL), correspondingly lower mean vol_score , and substantially more mass in the suppression zone ($\text{vol_pct} > 0.80$) and the zero-exposure zone ($\text{vol_pct} > 0.90$). For ES and NQ, the two contracts with the strongest asymmetry ($|\gamma| = 0.174$ and 0.145 respectively), the share of gate-active days falling into the zero-exposure zone is roughly 13 percentage points higher after negative residuals than after positive residuals. For CL, with the weakest asymmetry ($|\gamma| = 0.050$), the same gap is only 4.8 percentage points. The ordering of the empirical effect matches the ordering of the asymmetry parameter estimates from Section 4.2.

The sign-bucket pattern is consistent with how the fitted asymmetric recursion affects the score. It does not show what the same trades would earn under a matched symmetric recursion, nor establish that negative-shock entries are the worst opportunities. Conditioning on the model’s gate also selects observations using the same volatility process. Mechanism attribution therefore remains a hypothesis.

6.3 The Conditional Nature of the Edge

The framework’s performance is state-dependent rather than unconditional, and that is visible in the return pattern. Most of the returns come in years with moderately expanding volatility (2021, 2023, 2024), while the combined score and sizing layer can reduce exposure during more extreme episodes (2020, 2022). So the thesis is not claiming unconditional profitability. The narrower finding is a historical association between the selected combined layer and performance in these years. Stable state-space boundaries and reliable withdrawal from adverse states have not been identified.

This state dependence has specific and testable implications for the mechanism’s boundaries. Three conditions should weaken or break it:

Prolonged low volatility. A sustained regime with no volatility expansion episodes would produce few eligible trading days, reducing the framework’s opportunity set to near zero. The evaluation period includes low-volatility subperiods (early 2019, mid-2021) but not a multi-year low-volatility regime comparable to, for example, 2005–2006.

Negative-shock-dominated expansion. A market environment in which volatility expansions are consistently driven by negative shocks with no subsequent directional recovery could activate the expansion gate while the score taper suppresses some entries. Any surviving entries could still face adverse conditions; suppression is not guaranteed by the sign of the preceding shock. The evaluation period includes two such episodes (March 2020, H1 2022), both of which the gate handled well, but neither lasted more than several months.

Microstructure regime change. A structural change in the speed or nature of intraday price discovery, such as a shift in the role of algorithmic market-making or a change in the informational content of order flow, could alter the relationship between volatility expansion and breakout quality. The channels described in Section 6.1 depend on specific microstructure features that are stable within the evaluation period but not guaranteed to persist indefinitely.

The evidence supports a historical performance association for the combined layer within the selected 2019–2025 design. Whether its domain is stable across very different market environments is a separate question, and one that the available data cannot settle.

6.4 The Role of Crude Oil

The saved three-asset comparison reports a Sharpe increase from 1.350 to 1.412 and daily PnL correlation of 0.001 between CL and the equity-index basket. This is consistent with diversification in the observed sample, conditional on the selected allocation. It does not establish stable positive expectancy or attribute CL performance to model fit. The conflicting standalone CL ratios and drawdowns in the cost table (Table 26) and direction appendix (Appendix E) require reconciliation.

6.5 The Equity Tailwind

The 2019–2025 evaluation period was broadly favorable for long-only equity index exposure. The thesis addresses this concern through four pieces of evidence, each of which mitigates it partially without resolving it fully.

The reported market coefficient of 0.210 and low linear-factor R^2 do not rule out state-dependent equity exposure or omitted factors. The small positive saved return for 2022 is useful descriptive evidence, but a single year is not a stress guarantee. The earlier training Sharpe of 0.066 is a counterweight to the

stronger later performance, and similar gate firing rates do not resolve that gap. The repeated ablation contrast in 2024–2025 remains part of the same development-exposed research history.

What the thesis can claim is narrower: within the observed sample, the results are not well explained by passive equity exposure alone. What it cannot claim is that comparable performance would survive a prolonged multi-year bear market with no recovery, because the evaluation window does not contain such an episode. That boundary is an implication of disciplined inference rather than a weakness of the paper’s honesty.

6.6 Limitations

This section identifies the main unresolved threats in descending order of importance: first identification, then external scope, then implementation realism, and finally model-choice uncertainty. For each, the scope of what the thesis can and cannot claim is stated explicitly, along with the evidence that would resolve the concern.

Identification and temporal leakage. The architecture, asset choices, and reported comparisons were developed with knowledge of 2019–2025. Restricting a later parameter search to pre-2019 observations does not reverse that exposure. Full-sample fitted coefficients, percentile ranks, and thresholds must each be replaced with decision-time estimates. A fully dated specification freeze followed by genuinely unseen observations is needed; no current performance number is asserted to measure the cost of selection bias.

External scope: single evaluation window and limited asset coverage. The seven-year sample contains meaningful regime variation (pandemic crash, recovery, rate shock, expansion), but it does not include a prolonged multi-year bear market or a sustained low-volatility period. Nor does the three-contract design establish universality beyond equity index and energy futures. *Can claim:* the saved comparisons favor the combined layer within the observed sample and design. *Cannot claim:* that the same magnitude of effect will persist under more extreme, more prolonged, or structurally different conditions. *Would resolve:* extension to a fuller market cycle and additional asset classes with independent parameterization.

Implementation and accounting. The simulation permits fractional quantities, close-based signals and exits, possible positions outside entry sessions, and quantity changes without separately charged rebalancing costs. The bar-level and entry-fixed trade ledgers disagree. Fixed cost multipliers do not validate fills, margin capacity, slippage, or impact. A canonical position ledger with explicit execution timing, session exits, integer or micro-contract sizing, and collateral accounting is required before a live interpretation.

Model-choice uncertainty. CL retains residual ARCH effects. Realized-volatility comparisons do not isolate asymmetry, and a matched $\gamma = 0$ test remains necessary. Alternative conditional-volatility models may be useful, including realized-GARCH (Hansen et al., 2012), but additional searches must be included in the trial inventory and evaluated chronologically.

6.7 Future Research

The limitations above define a natural research agenda, ordered by expected impact on the thesis’s conclusions.

Prospective validation. Freeze the entire pipeline, including data transformations, model-estimation windows, score and regime calibration, search grid, accounting, and execution rules. Evaluate it on observations demonstrably unseen during development. Post-2025 data can serve this role only if the relevant design was frozen before those observations were inspected; the calendar date alone does not establish a holdout.

Stronger falsification. The randomized frequency-preserving placebo implemented here is a useful step, but it can be extended further. The most valuable next falsification test would preserve gate frequency and additional dependence structure, such as within-month clustering or transition probabilities, while remaining economically meaningless.

Alternative volatility specifications. The residual ARCH effects in CL represent the most clearly identified model limitation. A realized-GARCH specification (Hansen et al., 2012), which augments the conditional variance equation with high-frequency realized measures, is a natural candidate given the availability of 5-minute data. Markov-switching GARCH and jump-diffusion specifications could accommodate the discrete regime shifts and extreme tail events (e.g., April 2020) that the single-process EGARCH cannot represent. More broadly, testing the framework against stochastic volatility models would clarify whether the observed advantage of the EGARCH-based input is robust to alternative estimators and whether asymmetry has any separate contribution.

Asset-class expansion. The three-contract design leaves open how broadly the framework generalizes. Fixed-income futures (where volatility dynamics are driven by monetary policy rather than risk appetite), non-U.S. equity indices (which would test dependence on U.S. microstructure), and additional commodity classes (metals, agriculture) represent the most informative extensions. Preliminary tests during development found that Treasury futures detracted under the current parameterization, but this may reflect suboptimal asset-specific parameters rather than a fundamental incompatibility.

Adaptive parameters. The current framework uses fixed vol-score and regime thresholds. An adaptive estimation approach, such as expanding-window recalibration of the vol-score shape parameters, could improve robustness across changing market conditions.

7 Conclusion

This thesis studies whether conditional-volatility information can improve a long-only intraday breakout framework in equity index and energy futures. The saved 2019–2025 development results favor the combined volatility filtering and sizing layer: reported Sharpe is 1.412 with the layer and 0.664 without it, with maximum drawdown of 10.6% and 29.3% respectively. The associated paired-bootstrap nonpositive-tail fraction is 0.0038, conditional on those selected return streams.

The evidence is comparative and retrospective. Removing the entire layer changes both trade eligibility and size. Alternative volatility inputs and residual-sign diagnostics do not isolate an EGARCH asymmetry effect. Full-sample volatility calibration uses future information, and the research history does not establish an untouched 2024–2025 holdout. The pre-2019 parameter-selection exercise reports a later Sharpe of 1.195, but the earlier training Sharpe of 0.066 shows substantial period dependence. Neither that exercise nor partial rolling re-selection resolves architecture selection or every information-timing issue.

The paper’s contribution is a documented empirical framework and a set of historical comparisons that motivate further testing. Its prospective profitability and implementability remain unsettled. A canonical trade and daily-return ledger, historical-only model fitting and calibration, a gate-by-sizing decomposition, a matched symmetric-volatility experiment, and an independently frozen evaluation are the next steps needed to make a stronger claim. The revision retains the original empirical outputs where a correction cannot be supported and identifies those items in Appendix I.

Appendices

A Data Integrity and Cleaning Checks

This appendix summarizes the main data verification steps conducted before the backtesting exercise. Because intraday futures data are vulnerable to structural issues such as duplicate timestamps, missing bars, inconsistent OHLC values, and discontinuities around contract transitions, the cleaned datasets were audited systematically before model estimation.

The audit covered four main checks. First, timestamp integrity was evaluated by verifying that all observations were strictly ordered and that no duplicate timestamps remained in the cleaned files. Second, OHLC consistency was checked to ensure that the high price was not below the open, close, or low, and that the low price was not above the open, close, or high. Third, the data were screened for nonpositive prices and unusually large return spikes. Fourth, the asset-specific trading sessions were inspected for missing bars.

The results indicated that the cleaned files for NQ, ES, and CL were structurally sound. No duplicate timestamps were detected, all timestamp series were monotonic, and no OHLC consistency violations were found. The number of missing bars during the trading sessions was limited relative to the size of the full sample and did not indicate systematic corruption.

Table 24: Intraday futures data audit results

Asset	Obs.	Dup. Time	High Err.	Low Err.	Non-Pos. Px	$ r_t > 1\%$
NQ	1146282	0	0	0	0	258
ES	1149796	0	0	0	0	206
CL	1138547	0	0	0	0	312

Note: Obs. denotes the number of 5-minute observations. Dup. Time denotes duplicate timestamps. High Err. and Low Err. indicate violations of the high-low price consistency checks. Non-Pos. Px counts nonpositive prices. The final column reports the number of bars with absolute returns greater than 1%.

B Additional Strategy Variants

This appendix reports the additional strategy variants considered during model development. The variant sequence followed specific design hypotheses: EGARCH vol-score shape (tent vs flat-middle), regime threshold selection (33/67 vs 42/58), IVT smoothing window (MA25 vs MA40), breakout window optimization, ATR stop multiplier tuning, RSI exit threshold adjustment, session window optimization, and asset-specific parameter differentiation. Economic motivation does not remove the selection effects from comparing these modifications on the same data.

The tested variants included alternative vol-score shapes, different regime thresholds, varying IVT smoothing windows, asset-specific breakout windows, ATR stop multipliers, RSI exit thresholds, and session endpoints. A consistent pattern emerged: the flat-middle vol-score shape, the 42/58 regime thresholds, and MA40 IVT smoothing formed the strongest EGARCH configuration, while asset-specific signal parameters (breakout windows, EMA periods, session hours) provided further gains when tailored to the distinct dynamics of equity index versus energy futures.

Table 25: Selected strategy variants tested during model development

Variant	Sharpe	Max DD (%)	Return Mult.	Neg. Yrs
Baseline (tent vol-score, r33/67, MA25)	1.239	-13.6	3.38×	0
Flat-middle vol-score only	1.296	-13.2	3.52×	0
Regime 42/58 only	1.282	-13.4	3.47×	0
Flat-mid + r42/58 + MA40	1.366	-13.6	3.68×	0
+ NQ/ES bw=60	1.412	-12.4	3.70×	0
+ ATR 2.2 + RSI exit 52 + CL sess 13:00	1.412	-12.4	3.88×	0
No-EGARCH	0.664	-32.5	4.65×	0

Note: Max DD denotes maximum drawdown. Return Mult. denotes terminal equity divided by initial capital. This legacy variant table reports drawdowns of 12.4% and 32.5% for filtered and unfiltered final variants, versus 10.6% and 29.3% in the main table; frequency or accounting differences require reconciliation. Neg. Yrs denotes the number of negative calendar years in the 2019–2025 evaluation period.

C Cost Sensitivity by Asset

The main text reports portfolio-level cost sensitivity (Table 11). This appendix decomposes the analysis by individual contract to identify which assets are most and least tolerant of higher transaction costs.

Table 26: Per-asset cost sensitivity of the final three-asset strategy

Asset	Cost Mult.	Sharpe	Sortino [†]	Max DD (%)	Return Mult.	Breakeven
NQ	0.5×	1.441	0.413	−10.3	2.86×	
	1.0×	1.398	0.402	−10.6	2.81×	
	2.0×	1.311	0.378	−11.2	2.69×	
	4.0×	1.136	0.331	−12.5	2.47×	
	8.0×	0.786	0.232	−17.1	2.02×	17.5× (\$254)
ES	0.5×	1.431	0.427	−7.9	2.03×	
	1.0×	1.255	0.378	−9.0	1.90×	
	2.0×	0.904	0.276	−12.3	1.65×	
	4.0×	0.206	0.064	−21.2	1.15×	
	8.0×	−1.109	−0.336	−85.0	0.15×	5.0× (\$148)
CL	0.5×	0.515	0.134	−10.0	1.21×	
	1.0×	0.446	0.117	−10.5	1.18×	
	2.0×	0.308	0.082	−11.4	1.13×	
	4.0×	0.034	0.009	−13.5	1.01×	
	8.0×	−0.507	−0.139	−25.2	0.79×	4.5× (\$58)

Note: Each asset is evaluated independently with its champion parameters over 2019–2025. Base round-turn costs: NQ \$14.51, ES \$29.51, CL \$12.80. Breakeven column reports the approximate cost multiplier at which the standalone Sharpe ratio turns negative, with the implied round-turn cost in parentheses. Initial capital: \$200,000.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The decomposition reveals substantial heterogeneity across contracts. NQ is the most cost-tolerant asset, with a standalone breakeven multiplier of approximately 17.5× (implied round-turn cost of \$254), reflecting its combination of high gross profitability and low base transaction costs. ES breaks even at approximately 5.0× (\$148 round-turn), consistent with its higher base cost and lower per-trade profit relative to NQ. CL is the most cost-sensitive contract, reaching breakeven at approximately 4.5× (\$58 round-turn), a consequence of its lower standalone Sharpe ratio and smaller gross profit margin.

The saved portfolio cost table crosses zero near the eight-times cost scenario. The zero-PnL cost multiplier depends on aggregate gross PnL and cost-weighted quantity, not on volatility reduction by diversification. It can exceed the ES and CL standalone thresholds because the portfolio also includes NQ, whose saved threshold is higher. The original explanation attributing breakeven profitability to variance reduction was incorrect.

D Supplementary Figures

Figure 15 compares the final three-asset model with its No-EGARCH counterpart. The visual deterioration of the No-EGARCH specification reinforces the interpretation in the main text that volatility conditioning is a core component of strategy effectiveness.

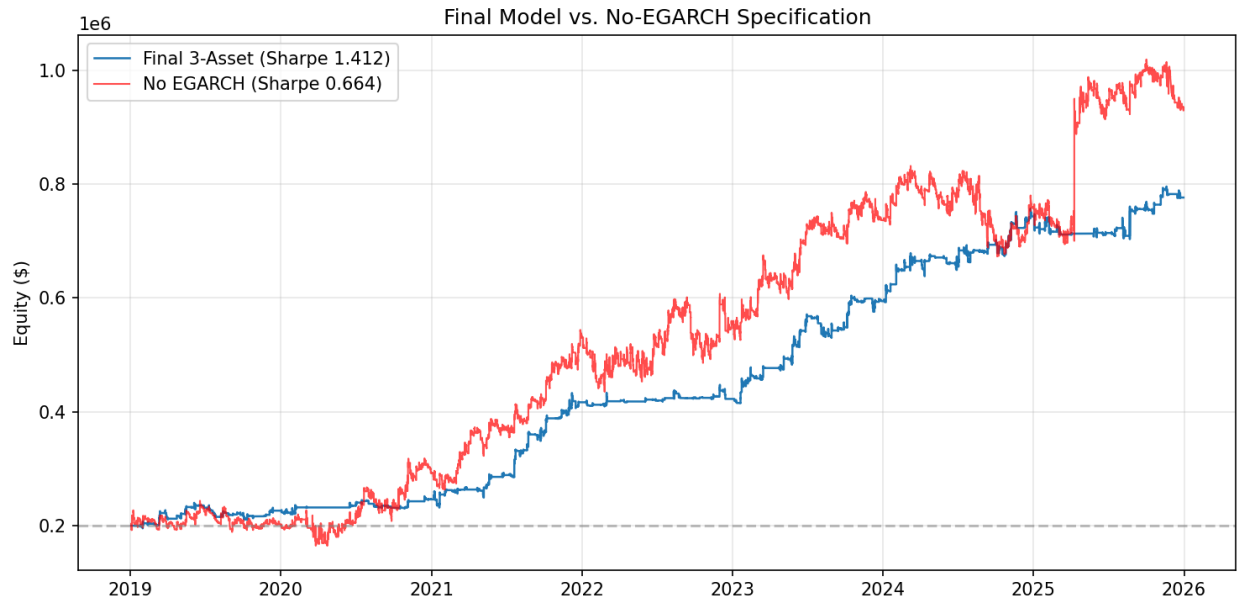


Figure 15: Equity comparison between the final three-asset model and the No-EGARCH specification.

Figure 16 reports Sharpe ratios across the subperiod analysis discussed in the main text.

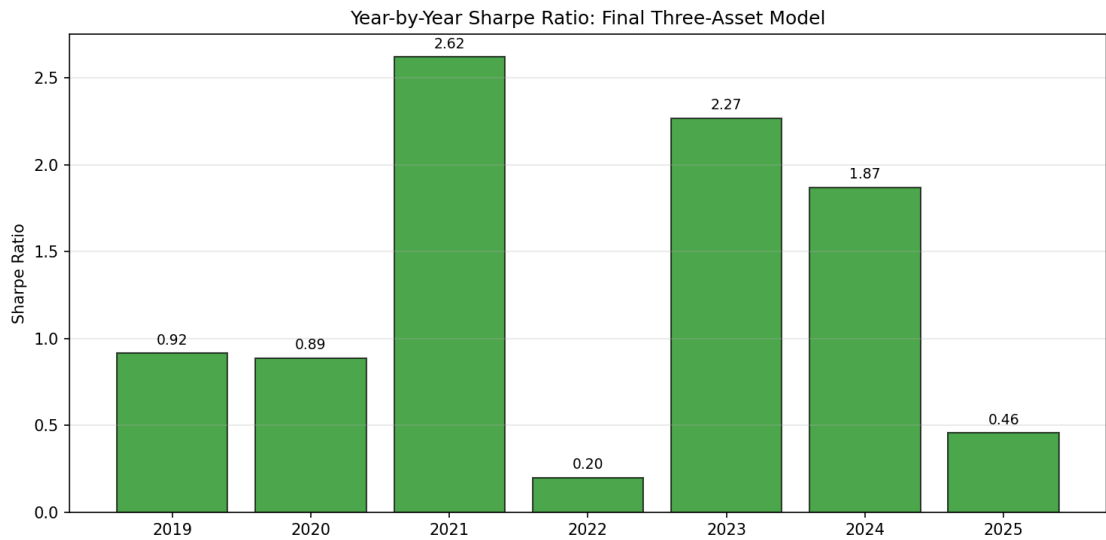


Figure 16: Sharpe ratio across the selected subperiods of the final three-asset model.

E CL Direction Restriction Analysis

The restriction to long-only positions is justified in Section 2.2 on theoretical grounds (momentum crash risk on the short side and positive drift in equity indices). For CL, which has no structural positive drift, the restriction requires empirical justification. Table 27 compares three CL specifications: long-only (the champion), short-only (using symmetric RSI and EMA filters for short entries), and a combined long and short version.

Table 27: CL direction restriction analysis: long-only vs. short-only vs. long-and-short (2019–2025)

CL Specification	Sharpe	Sortino [†]	Max DD	Return Mult.	Long Tr.	Short Tr.	Neg. Yrs
Long-only (champion)	0.479	0.365	-6.8%	1.18×	381	0	2
Short-only	0.122	0.108	-16.2%	1.06×	0	344	3
Long-and-short	0.433	0.450	-11.6%	1.26×	379	342	3

Note: CL-only results using champion parameters. Short-only uses RSI < 42 entry, RSI > 48 exit, EMA fast < EMA slow, and breakout below rolling low. Long-and-short combines both signal types. All specifications use identical EGARCH filtering, IVT sizing, and regime caps.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

Within this saved direction comparison, CL long-only reports Sharpe 0.479 and drawdown 6.8%, short-only reports 0.122 and 16.2%, and long-and-short reports 0.433 and 11.6%. These selected historical outcomes favor long-only in risk-adjusted terms but do not establish a supply-response mechanism or a stable directional asymmetry. The baseline cost appendix instead reports long-only Sharpe 0.446 and drawdown 10.5%; those values are preserved as unresolved output differences, not substituted for one another.

F Portfolio Composition: NQ+ES Only vs. Three-Asset

A natural question is whether CL earns its place in the portfolio on a risk-adjusted basis, given its modest PnL contribution relative to NQ and ES. Table 28 compares the two-asset (NQ+ES) and three-asset (NQ+ES+CL) portfolios.

Table 28: Portfolio composition test: NQ+ES only vs. three-asset champion (2019–2025)

Portfolio	Sharpe	Sortino [†]	Max DD	Return Mult.	Trades	Neg. Yrs
NQ + ES only	1.350	1.346	-12.0%	3.71×	834	1
NQ + ES + CL (champion)	1.412	1.579	-10.6%	3.88×	1215	0

Note: Both specifications use identical EGARCH filtering, IVT sizing, and champion parameters. Daily PnL correlation between the equity index basket (NQ+ES) and CL is 0.001, indicating low sample linear correlation, without establishing tail or nonlinear independence.

[†] Historical ratio using the standard deviation of negative returns; not conventional target downside deviation (Section 4.7).

The three-asset champion achieves a marginally higher Sharpe ratio (1.412 vs. 1.350) and, critically, zero negative calendar years versus one for the two-asset version (2022: −3.3%). The daily PnL correlation between the equity index basket and CL is 0.001, indicating near-zero sample linear correlation, without establishing tail or nonlinear independence. CL’s portfolio role is therefore diversification under bounded exposure rather than raw return maximization: in 2022, when the equity index strategy posted a loss, CL’s positive contribution was sufficient to keep the overall portfolio in positive territory. Maximum drawdown is also slightly lower with CL included (10.6% vs. 12.0%). This is the relevant inclusion test for CL; it need not dominate on standalone Sharpe or volatility-model fit to justify its place in the portfolio.

G Contract-Count Scaling Sensitivity

Table 29 varies the contract-count scaling parameter L from 1 to 15 while retaining the saved signal settings. The original “leverage” labels refer to this parameter, not measured account notional or margin leverage. The table is a historical sizing sensitivity, not evidence of a feasible margin budget.

Table 29: Scale L sensitivity: EGARCH-filtered vs. No-EGARCH specification (2019–2025)

Scale L	With EGARCH				Without EGARCH			
	Sharpe	Max DD	Ret. Mult.	Ret/DD	Sharpe	Max DD	Ret. Mult.	Ret/DD
1×	1.412	−3.0%	1.58×	19.33	0.664	−8.7%	1.73×	8.39
2×	1.412	−4.7%	2.15×	24.47	0.664	−13.1%	2.46×	11.15
3×	1.412	−6.8%	2.73×	25.44	0.664	−18.9%	3.19×	11.59
5×	1.412	−10.6%	3.88×	27.17	0.664	−29.3%	4.65×	12.46
7×	1.412	−13.9%	5.04×	29.06	0.664	−38.2%	6.11×	13.38
10×	1.412	−18.2%	6.77×	31.70	0.664	−49.6%	8.30×	14.72
15×	1.412	−23.9%	9.65×	36.19	0.664	−64.5%	11.95×	16.98

Note: The bold baseline uses contract-count scale $L = 5$. This varies simulated quantity, not measured account leverage. Ret/DD is $(\text{Return Mult.} - 1)/|\text{Max DD}|$. No margin or integer-contract model is supplied. The unfiltered variant removes the gate, score, inverse-volatility sizing, and regime caps and uses fixed quantities at the stated scale. Reported invariance applies only to this saved scaling exercise.

The reported ratios are invariant across this saved scaling exercise. Exact invariance holds for a positive linear rescaling of an unchanged PnL stream; it need not hold after integer rounding, clipping, margin constraints, impact, or changing transaction-cost structures. The historical negative-return ratio has the same scaling property but is not the conventional Sortino statistic.

For the EGARCH model, maximum drawdown increases less than proportionally with L within the tested range: tripling L from 5 to 15 increases drawdown from 10.6% to 23.9%, a factor of 2.25 rather than 3.0. For the No-EGARCH specification, a similar pattern holds: the same tripling increases drawdown from 29.3% to 64.5%, a factor of 2.20. The more consequential difference is the *level*: at every tested L , the EGARCH model’s maximum drawdown is substantially lower than that of the No-EGARCH version. At $L = 10$, for example, the EGARCH model produces a 18.2% drawdown compared with 49.6% for the unfiltered specification.

The return-to-drawdown ratio provides a complementary perspective. At contract-count scale $L = 5$, the EGARCH model achieves a Ret/DD of 27.17 compared with 12.46 for the unfiltered version. This approximately twofold advantage in drawdown efficiency persists across all tested scaling levels, suggesting that EGARCH conditioning produces a materially more favorable quantity-scaling and drawdown relationship within the scope of the backtest. Both specifications record zero negative calendar years at all tested leverage levels, but the EGARCH model achieves this with substantially less peak-to-trough exposure.

No margin model supports a numerical safe-leverage threshold. In particular, the original suggestion that risk becomes problematic only beyond 7–10 times is withdrawn. Contract prices, point values, equity, simultaneous positions, broker requirements, and liquidation rules determine financing feasibility. The observed drawdown at $L = 5$ does not establish operational robustness.

H Trading Rule Summary

For clarity, this appendix restates the final three-asset strategy in compact rule form.

Final Three-Asset Model

The final model is applied to NQ, ES, and CL futures using 5-minute intraday data over the 2019 to 2025 evaluation period. Entries are restricted to asset-specific sessions: 09:30 to 14:00 for NQ and ES, and 09:30 to 13:00 for CL. The strategy takes long positions only. Position sizing uses EGARCH inverse vol targeting (IVT) with contract-count scale $L = 5$: the base contract count equals the pre-OOS median conditional volatility divided by the 40-day MA of lagged conditional volatility, multiplied by the contract-count scale, and further scaled by the flat-middle continuous volatility score.

A trade is eligible only if the EGARCH volatility expansion gate is active (daily conditional volatility exceeds its 10-day moving-average, lagged one day). For CL, regime-conditional caps further limit exposure:

$$\text{CL: } w_{\text{low}} = 0.00, w_{\text{medium}} = 0.40, w_{\text{high}} = 0.80.$$

NQ and ES do not use regime-based caps.

Conditional on the volatility gate, a long trade is entered when the close breaks above the rolling breakout high (60-bar lookback for NQ and ES, 30-bar for CL), RSI exceeds 58, and the short EMA is above the medium EMA (EMA15/40 for NQ and ES, EMA20/50 for CL).

Once a trade is entered, a flat ATR trailing-stop-loss governs the exit: $2.2 \times$ ATR for NQ and ES, $2.5 \times$ ATR for CL. No fixed profit target is used; instead, the trailing ATR stop continuously updates the stop level as price moves favorably, allowing winning trades to run. Long trades are also exited if RSI falls below 52. Positions are further subject to a maximum holding period of 72 observed bars and a cooldown period of 5 bars after exit. The archived loop does not itself force session-end liquidation.

Asset-Specific Parameters

Table 30: Asset-specific parameter summary

Parameter	NQ	ES	CL
Session	09:30–14:00	09:30–14:00	09:30–13:00
Breakout window	60 bars	60 bars	30 bars
EMA (short/medium)	15/40	15/40	20/50
ATR stop multiplier	$2.2 \times$	$2.2 \times$	$2.5 \times$
Regime caps	None	None	Low=0, Med=0.4, High=0.8

Shared Parameters

RSI entry threshold: 58. RSI exit threshold: 52. Maximum holding period: 72 bars. Cooldown: 5 bars. Contract-count scale: $L = 5$. IVT smoothing: MA40. Vol-score shape: flat-middle (0.20, 0.40, 0.80, 0.90). Regime thresholds: 42nd/58th percentile. Longs only.

I Reconciliation and Reproducibility Register

This editorial revision does not rerun empirical models. It preserves reported results where the saved evidence does not identify a definitive correction. The following register separates resolved presentation errors from analysis that must be recomputed under a frozen protocol.

Resolved numerical and mathematical presentation. The twice-cost Sharpe is 1.215 and the eight-times-cost Sharpe is -0.022 , as printed in the saved cost table. Terminal equity multiples are distinguished from net returns. The Sharpe-difference equation now includes $\sqrt{252}$. EGARCH centering follows the archived software convention $\sqrt{2/\pi}$. The residual-sign note now gives the correct negative direction for score differences. Published bibliography metadata and the circular-bootstrap citation have been corrected.

Calendar and inference inputs. The archived champion daily-return file contains 1,805 unique dates and 616 nonzero observations, with annualized Sharpe 1.412277 under its saved convention. This supports the frequency table's count, but does not establish the separate bootstrap sample of 1,762 claimed in the original text. The 2021 gate graphic reports 311 observations, and the residual-sign table leaves 2 ES, 1 NQ, and 3 CL days outside its sign buckets. Establish the exchange-session calendar, holiday and zero-PnL treatment, and exact aligned inputs before updating inferential results.

PnL and trade reconciliation. Bar-level asset totals are \$361,429, \$180,235, and \$34,888 (total \$576,552); the trade-level totals are \$360,504, \$180,699, and \$36,257 (total \$577,460). The trade-level regime totals agree with the latter. The original annual trade counts sum to 1,214 versus 1,215 for the full sample. Rebuild a single dated position-and-cash-flow ledger with a declared quantity convention, identify final open trades and boundary attribution, and generate every table from it.

CL and frequency conventions. The cost appendix reports CL Sharpe 0.446 and drawdown 10.5%; the direction appendix reports 0.479 and 6.8%. The historical statistic labeled Sortino uses the standard deviation of negative returns about their conditional mean. Preserve those outputs as historical diagnostics, then recompute conventional target-downside-deviation Sortino and drawdowns using one frequency, calendar, and equity path. Do not infer missing numbers from neighboring tables.

Chronological implementation. Refit volatility coefficients only on past data, compute daily ranks and thresholds using a past-only empirical distribution, and ensure the sizing anchor predates each test interval. Preserve a complete dated strategy freeze. The annual table originally labeled walk-forward is not substantiated by its examined generation block; the later rolling two-parameter search also retains future-informed calibration. Neither is an end-to-end leakage-free test.

Identification and search. Separate a gate-only change from a sizing-only change on matched candidate signals, and compare EGARCH with a matched $\gamma = 0$ restriction. Recompute paired uncertainty for each named contrast using common dates. Inventory all searched assets, directions, sessions, indicators, models, thresholds, and reporting choices before estimating trial-adjusted statistics. Randomization should be tested for sensitivity to serial clustering and cross-asset dependence.

Execution and feasibility. Specify whether decisions formed at a close execute at that close or the next observable price. Model stop fills, session-end liquidation, quantity adjustments, integer or micro-contract feasibility, transaction costs on every adjustment, margin, and collateral interest. Entry-session filters and a 72-observed-bar limit are not sufficient to establish a strictly intraday executable strategy.

Source provenance. The published PDF has SHA-256 prefix d846e4adfafddf40; the matching `masters_thesis.tex` and `thesis_final_draft.tex` have prefix 451ecfb8f9c6c82c. The accompanying revision package records full hashes, copied dependencies, source changes, static code evidence, and build and page-review results. Later numbered drafts are not substituted for the published version.

References

- Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2–3):115–158.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- arch Developers (2026). EGARCH model documentation. <https://arch.readthedocs.io/en/latest/univariate/generated/arch.univariate.EGARCH.html>. Accessed 14 October 2026.
- Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.
- Bailey, D. H. and López de Prado, M. (2014). The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5):94–107.
- Bollen, N. P. B. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance*, 59(2):711–753.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Brock, W., Lakonishok, J., and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5):1731–1764.
- Daniel, K. and Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2):221–247.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.
- Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.
- Kang, S. H., Kang, S.-M., and Yoon, S.-M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, 31(1):119–125.
- Lo, A. W., Mamaysky, H., and Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance*, 55(4):1705–1765.
- Misirli, E. U. (2024). Aggregate volatility risk and momentum returns. *European Financial Management*, 30(4):1994–2032.
- Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.
- Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.
- Park, C.-H. and Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21(4):786–826.
- Politis, D. N. and Romano, J. P. (1991). A circular block-resampling procedure for stationary data. Technical Report EFS NSF 370, Stanford University, Department of Statistics.
- Szakmary, A. C., Shen, Q., and Sharma, S. C. (2010). Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking and Finance*, 34(2):409–426.
- Zhuang, C. (2018). Improving performance of exchange rate momentum strategy using volatility information. *Physica A: Statistical Mechanics and its Applications*, 510:741–753.