UNIVERSITY OF COPENHAGEN
DEPARTMENT OF ECONOMICS

Master's Thesis

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

June 2026

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 develops a long-only volatility-gated breakout framework for Nasdaq 100, S&P 500, and WTI Crude Oil futures using 5-minute data over 2019–2025. EGARCH conditioning is used to identify volatility states in which intraday breakout signals are more likely to contain directional information. A pre-out-of-sample specification with zero parameter exposure to the evaluation window yields a Sharpe ratio of 1.195. The main result is comparative: removing the EGARCH conditioning layer roughly halves the Sharpe ratio and triples maximum drawdown under all specifications tested ($p = 0.004$, circular block bootstrap).

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 a seven-year evaluation period (2019–2025). A pre-out-of-sample exercise in which all parameters are selected on 2010–2018 data with zero exposure to the evaluation window yields a Sharpe ratio of 1.195 and a cumulative return of 3.36× on $200,000 initial capital; the development specification achieves 1.412 but should be read as an upper bound. The thesis's primary finding is an ablation result: removing the EGARCH conditioning layer causes the Sharpe ratio to fall from 1.195 to 0.382 under the pre-out-of-sample specification (and from 1.412 to 0.664 under the development specification), with maximum drawdown increasing approximately threefold in both cases. Circular block bootstrap tests confirm this difference at the 1% level ($p = 0.004$). Two placebo exercises sharpen the identification claim. First, shifting the expansion gate by even one trading day reduces the development Sharpe from 1.412 to 0.884. Second, a randomized frequency-preserving placebo that keeps the number of gate-active days fixed within each asset and calendar year yields a mean Sharpe of 0.561, a 95th-percentile Sharpe of 0.953, and a one-sided empirical placebo p-value of 0.002 in 500 simulations. A Fama–French regression shows significant alpha ($p < 0.001$) with common factors explaining only 3.0% of return variance.

The contribution is methodological: the thesis develops and tests a trade-level use of EGARCH (as a rule for deciding whether an individual intraday breakout should be executed at all) that is distinct both from its use as a variance forecast and from its use as a portfolio-level scaling input.

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. All analytical decisions, verification of results, interpretations, and final wording are my own.

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 . . . 15

4.4 Volatility Regime Classification . . . 17

4.5 Intraday Signal Construction . . . 18

4.6 Portfolio Construction and Position Management . . . 19

4.7 Performance Measurement . . . 20

4.8 Model Comparison and Ablation Design . . . 21

4.9 Statistical Inference: Circular Block Bootstrap . . . 22
4.10 Complete Parameter Specification . . . 23

5 Empirical Results **25**
5.1 Development-Enhanced Specification . . . 25
5.2 The EGARCH Conditioning Effect: Ablation Evidence . . . 26
5.3 Asset-Level Contribution and Trade Activity . . . 28
5.4 Trade Quality by Volatility Regime . . . 29
5.5 Year-by-Year Analysis and Subperiod Performance . . . 30
5.6 Transaction Cost Sensitivity . . . 32
5.7 Factor Exposure Analysis . . . 33
5.8 Expanding-Window Walk-Forward Validation . . . 35
5.9 EGARCH versus Alternative Volatility Filters . . . 35
5.10 Placebo and Falsification Tests . . . 36
5.11 Return Distribution and the Sharpe–Sortino Gap . . . 39
5.12 Parameter Stability . . . 41
5.13 Deflated Sharpe Ratio . . . 44
5.14 Development Period versus Holdout . . . 46
5.15 Pre-OOS Parameter Selection . . . 47
5.16 Rolling Parameter Re-Selection . . . 48
5.17 Interpretation of the Main Findings . . . 49

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 . . . 54
6.4 The Role of Crude Oil . . . 54
6.5 The Equity Tailwind . . . 55
6.6 Limitations . . . 55
6.7 Future Research . . . 56

7 Conclusion **58**

A Data Integrity and Cleaning Checks **61**

B Additional Strategy Variants **62**

C Cost Sensitivity by Asset **63**

D Supplementary Figures **64**

E CL Direction Restriction Analysis **65**

F Portfolio Composition: NQ+ES Only vs. Three-Asset **66**

G Leverage Sensitivity **67**

H Trading Rule Summary **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 26
4 Drawdown comparison of the final model and No-EGARCH specification 27
5 Cumulative net PnL contribution by asset 28
6 EGARCH volatility gate and trade entries for NQ during 2021 30
7 Year-by-year returns of the final three-asset model (2019–2025) 32
8 Gate-timing placebo test 38
9 Daily return distribution and normal Q-Q plot 39
10 Distribution of daily strategy returns: gain and loss days 40
11 Sharpe ratio under single-parameter perturbation 42
12 Joint sensitivity: breakout window $\times$ ATR stop multiplier 43
13 Joint sensitivity: EMA fast $\times$ EMA slow 43
14 Joint sensitivity: RSI entry $\times$ RSI exit threshold 44
15 Equity comparison between the final three-asset model and the No-EGARCH specification. 64
16 Sharpe ratio across the selected subperiods of the final three-asset model. 64

List of Tables

1 Descriptive statistics for 5-minute log returns (2019–2025) 9
2 EGARCH(1,1) parameter estimates by futures contract 13
3 EGARCH(1,1) residual diagnostics by futures contract 15
4 Complete parameter specification of the final three-asset model 24
5 Main performance results: final three-asset model vs. no-EGARCH ablation (2019–2025) 25
6 Circular block bootstrap test of Sharpe ratio differences 27
7 Trade counts and net PnL by asset in the final three-asset model (2019–2025) 28
8 Trade-level statistics of the final three-asset model (2019–2025) 29
9 Trade quality conditional on volatility regime at entry (2019–2025) 29
10 Year-by-year performance of the final three-asset model 31
11 Cost sensitivity of the final three-asset strategy 32
12 Factor exposure regressions of strategy daily returns 34
13 Expanding-window walk-forward validation (2019–2025) 35
14 Performance comparison across volatility filter specifications 36
15 Gate-timing placebo test 37
16 Randomized frequency-preserving gate placebo 38
17 Sharpe and Sortino ratios at 5-minute and daily frequencies (champion specification, 2019–2025) 40
18 Parameter stability analysis: Sharpe ratio under single-parameter perturbation 41
19 Deflated Sharpe ratio under the corrected and stress-case formulations 45
20 Development period vs. pure holdout performance 46
21 Pre-OOS parameter selection: training on 2010–2018, evaluation on 2019–2025 47
22 Walk-forward analysis: annual parameter re-selection on rolling 24-month windows . . . 48
23 Direct test of the asymmetry channel via the vol-score distribution 53
24 Intraday futures data audit results 61
25 Selected strategy variants tested during model development 62
26 Per-asset cost sensitivity of the final three-asset strategy 63
27 CL direction restriction analysis: long-only vs. short-only vs. long-and-short (2019–2025) 65
28 Portfolio composition test: NQ+ES only vs. three-asset champion (2019–2025) 66

29 Leverage sensitivity: EGARCH-filtered vs. No-EGARCH specification (2019–2025) . . 67
30 Asset-specific parameter summary 69

1 Introduction

Intraday breakout signals do not mean the same thing in every market environment. A close above a recent high can reflect genuine directional buying when volatility is expanding and new information is arriving quickly, and almost the opposite when volatility is flat or falling and the same move is more likely to be noise or a thin order book being picked off. The observation that signal quality depends on the volatility regime is not itself new; it falls out naturally from 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. Trading is 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. Each component is independently identifiable, and the EGARCH contribution can be isolated through direct ablation.

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, EGARCH conditioning is central to the risk-adjusted performance of the breakout framework within the observed sample, and this is the thesis's primary empirical finding. Removing the EGARCH layer roughly halves the Sharpe ratio and nearly triples maximum drawdown under every specification tested, and a circular block bootstrap rejects the null of no difference at the 1% level ($p = 0.004$). Two placebo exercises reinforce this conclusion: shifting the gate by even one trading day reduces the Sharpe by 0.528, and a randomized frequency-preserving placebo yields a mean Sharpe of 0.561 with an empirical placebo p-value of 0.002.

Second, the pre-out-of-sample specification provides the cleanest estimate of performance. In that exercise, all parameters are chosen on 2010–2018 data with zero exposure to any post-2018 observation, and the untouched 2019–2025 window yields a Sharpe ratio of 1.195. The development specification achieves 1.412, but that figure reflects iterative refinement on the evaluation period and should be interpreted as an upper bound rather than the thesis's headline estimate.

Third, the conditioning mechanism operates inside a specific and identifiable domain: the volatility states in which breakout signals appear to carry more directional information. The framework steps out of the market when conditions fall outside that domain. What the thesis establishes is therefore a mechanism that is well identified within this domain, rather than a general-purpose trading rule.

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 parameter constraints, 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, and Kang et al. (2009) documented analogous dynamics in crude oil, though with a distinct economic mechanism (supply-driven shocks rather than leverage effects).

The asymmetry has real economic content, and it matters directly for the framework developed in this thesis. A volatility expansion that follows a sharp decline is not the same thing as one that follows a rally. The first is more likely to reflect compressed risk premia and heightened uncertainty, while the second may reflect stronger information arrival without the same degree of risk aversion. For a long-only strategy, that difference matters because breakout signals arriving in those two settings need not have the same likelihood of continuing.

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 in the present thesis is motivated by this literature. Momentum crashes are concentrated on the short side (Daniel and Moskowitz, 2016), and equity index futures exhibit positive drift at intermediate horizons. Restricting to long positions eliminates the component most vulnerable to crash risk and aligns the strategy with the directional asymmetry inherent in the EGARCH leverage effect.

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 exposure inversely with recent volatility improves risk-adjusted returns across asset classes. 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.

Work closer to the present thesis uses volatility to condition the directional signal itself rather than only the size of the exposure. Zhuang (2018) shows that volatility forecasts improve exchange-rate momentum, and Misirli (2024) links momentum profits to aggregate volatility risk using an EGARCH specification. Both papers suggest that volatility affects not only how much to trade but how informative a directional signal is to begin with. They work, however, with monthly or daily portfolio returns, and neither isolates the contribution of the volatility component through ablation.

This thesis pushes the same idea down to the level of the individual intraday trade. The EGARCH expansion gate is used as an admissibility rule: it asks whether the current daily volatility state is one in which an intraday breakout has historically carried directional content, and trades are taken only on days that pass that screen. Conditional on passing the screen, the continuous volatility score and inverse volatility targeting then determine position size. So volatility is doing more than one job in the framework, but the part I focus on is this trade-level admission step. I am not aware of prior work that uses asymmetric volatility conditioning in quite this form in an intraday futures setting while also isolating its contribution through an ablation, though the claim I want to make is narrower than one of novelty. The point is that, in the sample and design studied here, this use of EGARCH can be identified and evaluated fairly directly.

2.4 Positioning and Hypothesis

The thesis borrows its econometric framework from the GARCH literature, its directional signal structure from the momentum and breakout literature, and its conceptual link between volatility and signal quality from the volatility-scaling literature. What it adds is a specific way of combining them: EGARCH is used neither to forecast variance nor to scale portfolio exposure, but only to decide whether a given intraday breakout signal should be executed. The hypothesis is tested through three complementary exercises: a direct ablation of the EGARCH layer, two placebo tests that vary the gate's timing and content, and a pre-out-of-sample parameter-selection exercise that uses parameters chosen on 2010–2018 data with no exposure to the 2019–2025 evaluation window.

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.

The session filters are therefore integral to the empirical design rather than cosmetic restrictions. All signals, entries, and backtesting results are based on these asset-specific intraday windows.

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

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

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

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

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

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

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:

$$\mathrm{AG}_t = \frac{1}{14}\sum_{k=0}^{13}\max(\Delta_{t-k},\, 0), \qquad \mathrm{AL}_t = \frac{1}{14}\sum_{k=0}^{13}\max(-\Delta_{t-k},\, 0).$$

The RSI is then

$$\mathrm{RSI}_t = 100 - \frac{100}{1 + \mathrm{AG}_t/(\mathrm{AL}_t + \epsilon)},$$

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

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

where $L_{\mathrm{NQ}} = L_{\mathrm{ES}} = 60$ bars and $L_{\mathrm{CL}} = 30$ 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 statistics reveal several features standard in high-frequency financial data. Mean returns are close to zero at the 5-minute horizon, as expected. Volatility differs materially across assets, with NQ showing the largest return dispersion, followed by CL and then ES. All three series exhibit substantial excess kurtosis, indicating heavy tails and a non-negligible probability of extreme realizations. Skewness is small in absolute value but not exactly zero, suggesting some asymmetry in short-horizon return distributions.

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.

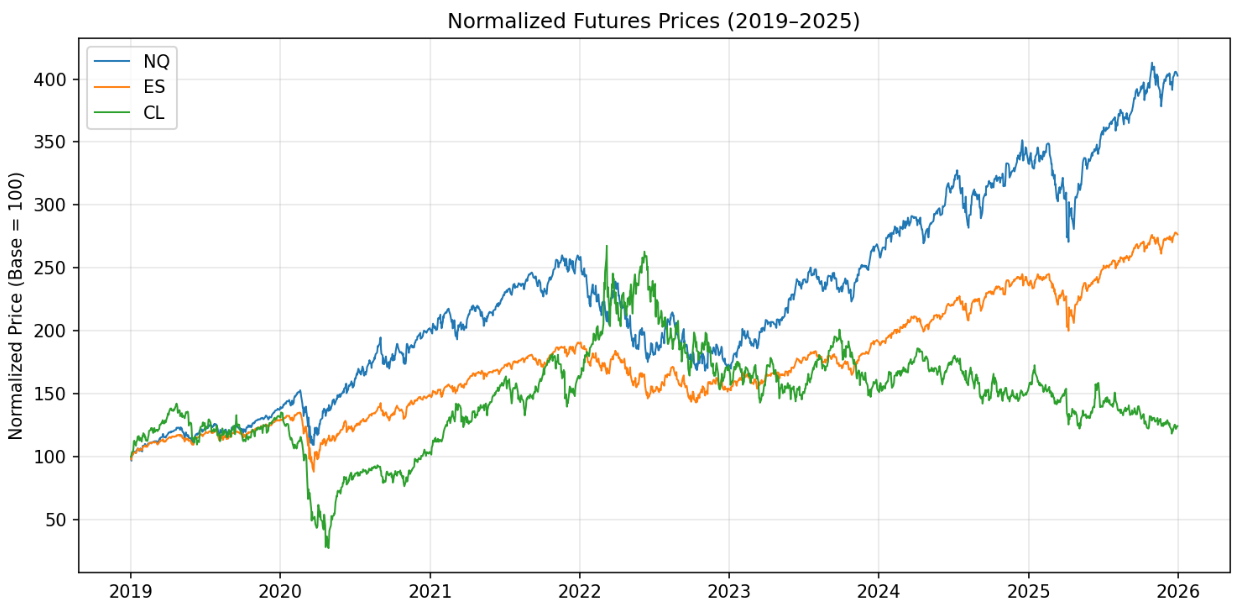

Figure 1: Normalized price series for NQ, ES, and CL over the 2019–2025 evaluation period. The figure illustrates the distinct price dynamics of the three futures markets used in the analysis.

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.

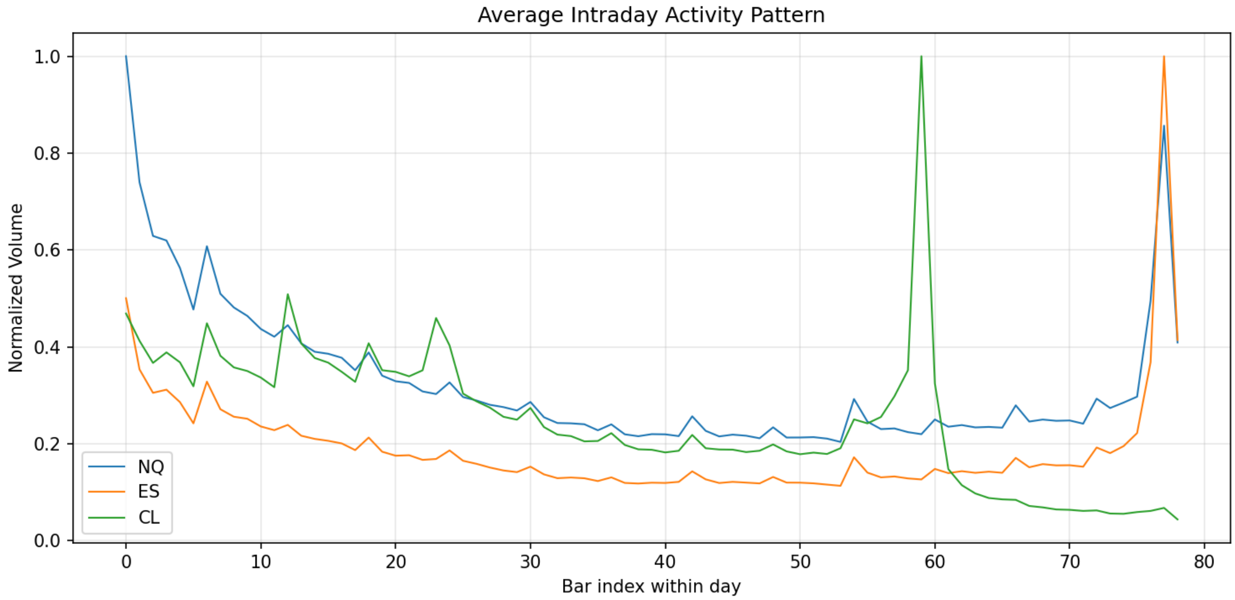

Figure 2: Average intraday activity pattern across the sample based on normalized volume. The opening and midday portions of the U.S. regular session show the strongest concentration of activity, supporting the asset-specific trading windows used in the thesis.

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

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

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

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

where $P_{i,\tau}^{(d)}$ denotes the daily closing price for contract i on day τ. Continuous-contract series are used throughout, constructed by rolling to the next front-month contract on the last trading day of the expiring contract and back-adjusting prior prices to eliminate roll gaps.

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

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

where $\sigma^2_{i,\tau}$ 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

$$z_{i,\tau-1} = \frac{\varepsilon_{i,\tau-1}}{\sigma_{i,\tau-1}}, \tag{4}$$

where $\varepsilon_{i,\tau-1}$ is the raw innovation in the daily return equation.

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 $\log(\sigma_t^2) = \omega + \beta\log(\sigma_{t-1}^2) + \alpha(|z_{t-1}| - E|z_{t-1}|) + \gamma z_{t-1}$. 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 ($p = 0.305$). The Ljung–Box rejections for NQ and ES likely reflect minor residual serial correlation in squared residuals rather than severe model misspecification, as the ARCH-LM tests confirm that no substantial conditional heteroskedasticity remains. 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.

These features are consistent with the well-documented finding that crude oil volatility dynamics resist parsimonious 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.

Despite this misspecification, the EGARCH(1,1) filter remains a useful first-order approximation for CL volatility conditioning. The model correctly captures the dominant features of CL's variance dynamics, namely persistence ($\beta = 0.987$), asymmetry ($\gamma = -0.050$, $p < 0.001$), and the general level and trend of conditional volatility, even though it underestimates the frequency and magnitude of extreme volatility jumps. For the purpose of the expansion gate, which requires only a directionally correct assessment of whether volatility is locally rising or falling, this level of approximation is sufficient. The residual ARCH effects indicate that the model's *precision* is limited for CL, not that its *direction* is unreliable.

The persistence of residual ARCH effects in CL raises a legitimate concern about gate accuracy for this contract. Three design features create a layered defense that does not require the EGARCH variance estimate to be precisely correct. First, regime-conditional caps (Section 4.4) bound CL exposure at 0/40/80$\%$ of full size in low/medium/high regimes, limiting the consequence of gate misclassification. Second, IVT sizing uses a 40-day moving-average of conditional volatility rather than the point estimate, so a single anomalous day carries only $1/40$ weight. Third, CL is included as a bounded diversifier, not as the portfolio's strongest standalone signal source. Its daily PnL correlation with NQ+ES is 0.001, and adding CL improves the portfolio Sharpe from 1.350 to 1.412 while eliminating the single negative calendar year that the two-asset variant produces (Appendix F). This portfolio role does not require CL's volatility model to match the precision of the equity index models, although the lower standalone Sharpe (0.479) indicates that its gate is less tightly calibrated. Section 6.6 discusses potential model improvements.

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. NQ and ES show adequate fit; CL exhibits significant residual ARCH effects.

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 single most important EGARCH component in the framework. It requires that the daily conditional volatility exceeds its own 10-day moving-average, lagged by one day:

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

where $\hat{\sigma}_{i,\tau-1}$ denotes the lagged daily conditional volatility for contract i. The one-day lag ensures that the gate is based on information available before the current trading day, preventing look-ahead bias. 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 computed using a flat-middle tent function of the daily conditional volatility percentile rank $p_{i,\tau}$, defined as the rank of today's conditional volatility within the full 2010–2025 conditional-volatility distribution, expressed as a fraction between 0 and 1. The full-sample ranking is a calibration choice rather than a look-ahead exposure of the strategy: the shape parameters (a, b, c, d) below are fixed structural quantities, and the strategy never conditions on a future percentile rank in real time. Table 13 (Section 5.8) reports a stricter check in which the percentile ranks are re-estimated each year using only data available before that year, and the qualitative results are unchanged. Let the four shape parameters be $(a, b, c, d) = (0.20, 0.40, 0.80, 0.90)$. The score is defined as:

$$\text{vol\_score}(p) = \text{clip}\left(\begin{cases} 0.30 + \dfrac{p}{a} \times 0.50, & p \le a \\ 0.80 + \dfrac{p-a}{b-a} \times 0.20, & a < p \le b \\ 1.00, & b < p \le c \\ 1.00 - \dfrac{p-c}{d-c}, & c < p \le d \\ 0, & p > d, \end{cases} \; 0,\ 1 \right) \tag{6}$$

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 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:

$$\mathcal{R}_{i,\tau} = \begin{cases} \text{Low}, & \hat{\sigma}_{i,\tau} < q_{0.42,i} \\ \text{Medium}, & q_{0.42,i} \le \hat{\sigma}_{i,\tau} < q_{0.58,i} \\ \text{High}, & \hat{\sigma}_{i,\tau} \ge q_{0.58,i}, \end{cases} \tag{7}$$

where $q_{0.42,i}$ and $q_{0.58,i}$ denote the 42nd and 58th percentiles of the contract-specific conditional volatility distribution.

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:

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

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:** vol_expand$_{i,\tau} = 1$ (daily conditional volatility exceeded its 10-day MA on the previous day; see Equation 5).

2. **Volatility score positive:** 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 $L_{\text{NQ}} = L_{\text{ES}} = 60$ and $L_{\text{CL}} = 30$; 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. If any single condition fails, no trade is entered and the strategy remains flat in that contract at that bar.

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. If no valid signal is present, the strategy remains flat in that contract.

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 leverage factor:

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

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 the leverage multiplier, 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 forward-looking sizing is the key innovation over backward-looking realized-vol approaches.

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:

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

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:

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

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). These cost figures include exchange fees, clearing fees, and an estimated half-spread. 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 an initial capital of $200,000. All performance measures are computed on the cumulative equity path beginning from this capital base.

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:

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

   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 bars (6 hours at 5-minute frequency).

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 same implementation logic is applied consistently across all model variants so that performance differences can be attributed to the design of the signal structure rather than to arbitrary changes in execution mechanics.

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

$$\text{Sharpe} = \frac{\bar{R}}{\sigma_R}\sqrt{252}, \tag{8}$$

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 Sortino ratio is computed as

$$\text{Sortino} = \frac{\bar{R}}{\sigma_R^-}\sqrt{252}, \tag{9}$$

where σ_R^- is the downside deviation based only on negative daily returns. The daily-frequency convention follows Lo (2002) and Bailey and López de Prado (2014) and is the standard reporting convention in the trading-strategy literature. Maximum drawdown is measured as the largest peak-to-trough decline in the cumulative equity curve. Section 5.11 reports a methodological cross-check that recomputes the same ratios at 5-minute frequency and discusses the discrepancy in the Sortino ratio that arises under sub-daily aggregation.

The passive benchmark is a fixed one-contract futures position, marked to market daily against the initial capital base, with no risk management, leverage scaling, or volatility filtering. This benchmark is deliberately simple: it isolates the contribution of the EGARCH conditioning and signal logic by holding all market exposure constant. Because the final strategy uses $5\times$ vol-targeted leverage while the benchmark is unlevered, terminal return multiples are not directly comparable; the Sharpe ratio is the primary comparison metric. The more informative benchmark comparison is the ablation test (Section 5.2), which holds signal logic fixed and varies only the EGARCH conditioning layer, providing a controlled estimate of the volatility filtering contribution.

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.

Three ablation tests are conducted. First, the complete EGARCH layer is removed (no volatility expansion gate, no vol-score, no IVT, no regime caps), leaving only the breakout, RSI, and EMA components. Second, only the volatility expansion gate is removed while retaining the other EGARCH components. Third, individual components (vol-score shape, IVT smoothing, regime thresholds) are varied to assess their marginal contributions. These ablations make the EGARCH contribution directly observable rather than assumed.

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 a diagnostic upper bound. 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

The central statistical test in this thesis is whether the Sharpe ratio improvement from EGARCH filtering is significant or merely reflects sampling noise. Because daily strategy returns exhibit serial dependence (from multi-bar trades and the cooldown mechanism), standard i.i.d. inference on the Sharpe ratio is inappropriate. Instead, the thesis employs a circular block bootstrap (Politis and Romano, 1994), which preserves the temporal dependence structure of the data.

The procedure operates on the daily portfolio return series $\{r_t\}_{t=1}^{T}$, where $T = 1{,}762$ trading days span the 2019–2025 evaluation period. Two return series are constructed: r_t^{final} from the full EGARCH model and $r_t^{\text{no-EGARCH}}$ from the ablated specification. Define the Sharpe difference as

$$\Delta S = \frac{\bar{r}^{\,\text{final}}}{\hat{\sigma}^{\text{final}}} - \frac{\bar{r}^{\,\text{no-EGARCH}}}{\hat{\sigma}^{\text{no-EGARCH}}}, \tag{10}$$

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 $b = 1, \ldots, B$, 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 the final and No-EGARCH return series, preserving the cross-sectional dependence between specifications on each day. The Sharpe difference $\Delta S^{*(b)}$ is computed from each bootstrap sample.

The one-sided p-value is the fraction of B bootstrap replications in which the resampled Sharpe difference is zero or negative:

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

The block length is set to $\ell = 20$ trading days (approximately one calendar month), which is long enough to capture the serial dependence induced by multi-day trade sequences and the 5-bar cooldown, while short enough to allow sufficient resampling variation. The number of replications is $B = 10{,}000$, providing resolution to the fourth decimal place in the p-value. Sensitivity checks with block lengths of 10 and 40 days produce qualitatively identical conclusions.

This test is preferred over the asymptotic Sharpe ratio test of Lo (2002) because the finite-sample properties of the circular block bootstrap are better suited to the moderate sample size ($T = 1{,}762$ days) and the non-normal, heavy-tailed return distribution documented in Section 5.11.

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		Prevents look-ahead
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
Leverage multiplier 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 uncapped at 1.0
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 bars (6 hours)		Forced 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	Incl. fees + half-spread

5 Empirical Results

This section presents the empirical evidence bearing on the thesis's hypothesis: that EGARCH conditioning improves the quality of intraday breakout signals by selecting the volatility states in which they are more likely to carry directional information. The analysis is organized by evidential strength, beginning with the development specification and its ablation, proceeding through robustness tests of increasing stringency, and culminating in the pre-OOS parameter-selection exercise (Sharpe 1.195 with zero exposure to the evaluation window), which provides the cleanest estimate of generalization performance. Throughout, the development Sharpe of 1.412 is treated as an upper bound, and the pre-OOS estimate of 1.195 is the thesis's headline performance figure.

5.1 Development-Enhanced Specification

The development specification implements the complete parameter set in Table 4, with parameters iteratively refined over the 2019–2025 evaluation period. This specification serves two purposes: it provides a baseline for the ablation analysis that follows, and it establishes the upper bound on measured performance before the controlled estimates in Sections 5.14–5.15 discount for parameter selection.

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: All results computed over the 2019–2025 out-of-sample period with \$200,000 initial capital. Transaction costs: NQ \$14.51, ES \$29.51, CL \$12.80 per round turn.

The development model achieves a Sharpe ratio of 1.412, a maximum drawdown of 10.6%, and a cumulative return of 3.88× over seven years from 1,215 completed trades. Because these parameters were refined on the evaluation period, this figure should be read as an upper bound. The pre-OOS estimate of 1.195 (Section 5.15) is the more credible performance figure, and the ablation evidence in the following subsection is more important than the raw Sharpe level. Zero negative calendar years are recorded, with the weakest years (2020: +9.9%, 2022: +2.9%) corresponding to the most adverse market environments.

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.

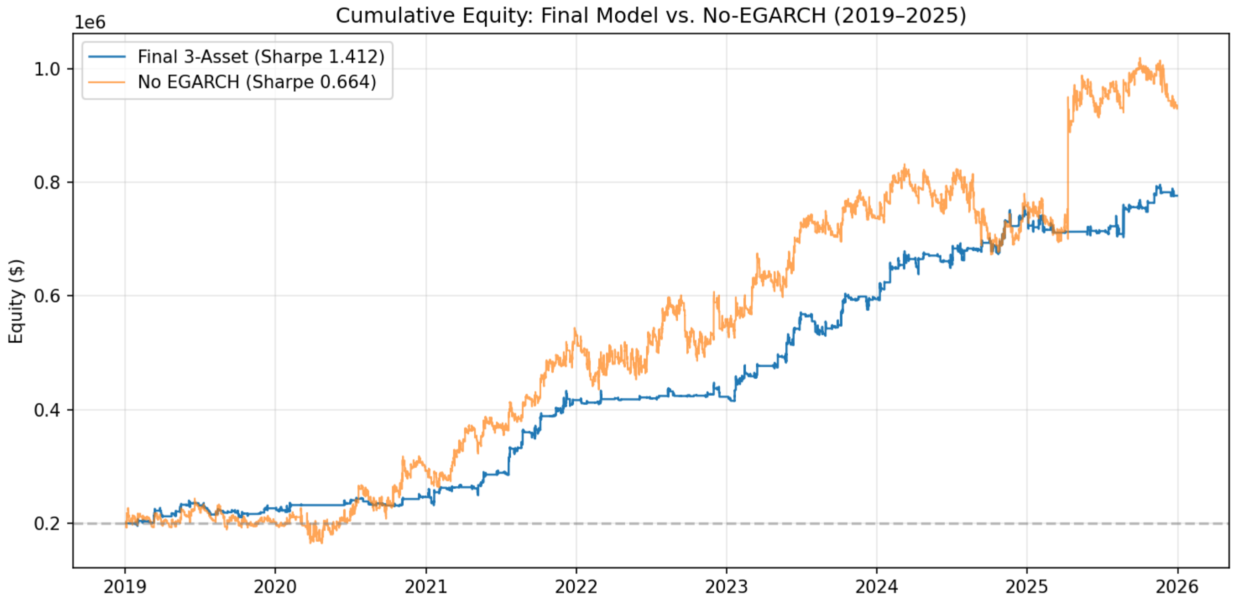

Figure 3: Equity curves of the final three-asset model and the No-EGARCH specification over the 2019–2025 evaluation period.

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), it suppresses entry and IVT reduces position sizes; 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× versus 3.88×) but at the cost of a 29.3% maximum drawdown.

5.2 The EGARCH Conditioning Effect: Ablation Evidence

The ablation analysis constitutes the thesis's primary empirical finding. It isolates the EGARCH conditioning layer's contribution by removing it entirely (the volatility expansion gate, the continuous vol-score, the inverse volatility targeting, and the regime caps) while holding all signal and trade-management components fixed. This design ensures that any performance difference is attributable to the conditioning mechanism rather than to changes in entry logic or risk management.

Removing the EGARCH layer causes the Sharpe ratio to fall from 1.412 to 0.664 and maximum drawdown to increase from 10.6% to 29.3%. The unfiltered version executes 3,842 trades versus 1,215. The threefold increase in trade count is informative: it reveals that the EGARCH layer's primary function is selectivity. It does not generate signals; it suppresses execution during volatility states in which breakout signals are less reliable. The magnitude of the Sharpe reduction (0.748) and the near-tripling of maximum drawdown indicate that the conditioning layer is fundamental to the framework's risk-adjusted performance rather than a marginal improvement.

Component-level ablation identifies the volatility expansion gate as the single most important element. Without the gate, the framework enters trades during declining or stagnant volatility, precisely the states in which the microstructure channels described in Section 6.1 (information arrival, liquidity transition, participation feedback) are weakest.

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

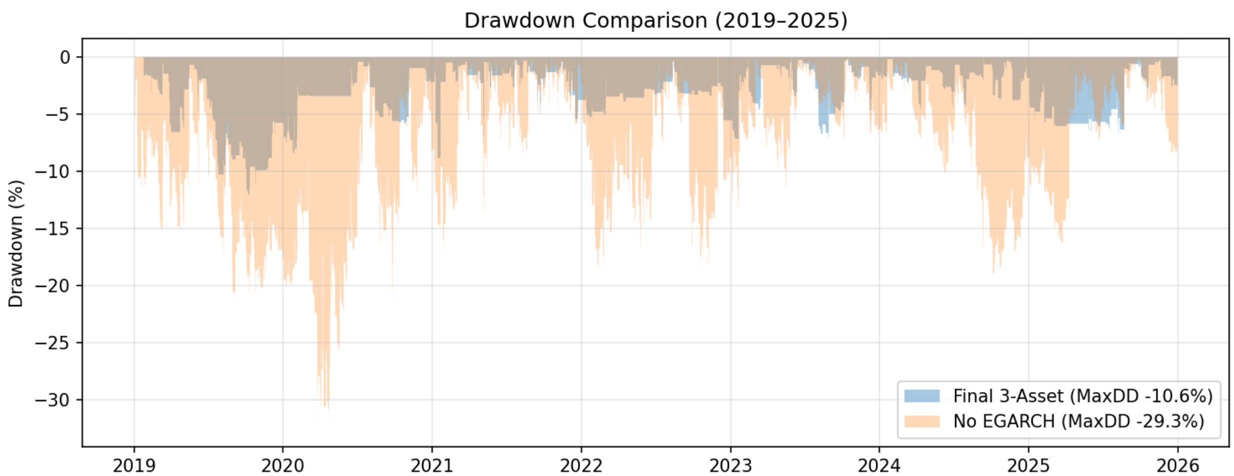

Figure 4: Drawdown comparison of the final three-asset model and the No-EGARCH specification over the 2019–2025 sample.

Without EGARCH conditioning, the framework follows a substantially more unstable return path, with deeper losses concentrated during the periods (2020, 2022) when the expansion gate would have suppressed entry.

To confirm that these Sharpe ratio differences are not merely sampling noise, a circular block bootstrap test (Politis and Romano, 1994) is applied with a block length of 20 days and 10,000 replications (Table 6). The Sharpe improvement of the final model over the No-EGARCH specification is significant at the 1% level, confirming that the observed performance gap is unlikely to reflect sampling variation alone.

Table 6: Circular block bootstrap test of Sharpe ratio differences

Comparison	Sharpe Diff (ann.)	p-value
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. p-value 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.

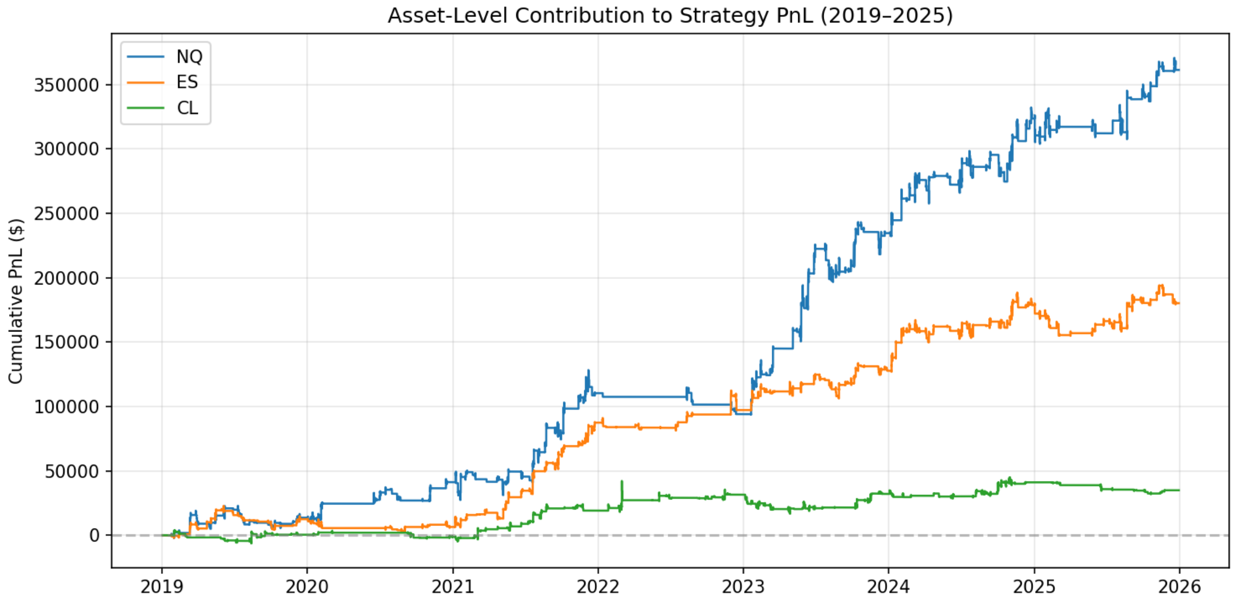

Figure 5: Cumulative net profit and loss contribution by asset in the final three-asset model.

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 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 a direct consequence of 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 average holding period of 17 bars (approximately 84 minutes) indicates that most trades are resolved within a single session. 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

A central claim of the thesis is that EGARCH conditioning improves the quality of breakout signals by restricting entry to favorable volatility environments. Table 9 tests this directly by comparing trade-level statistics across volatility regimes.

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.

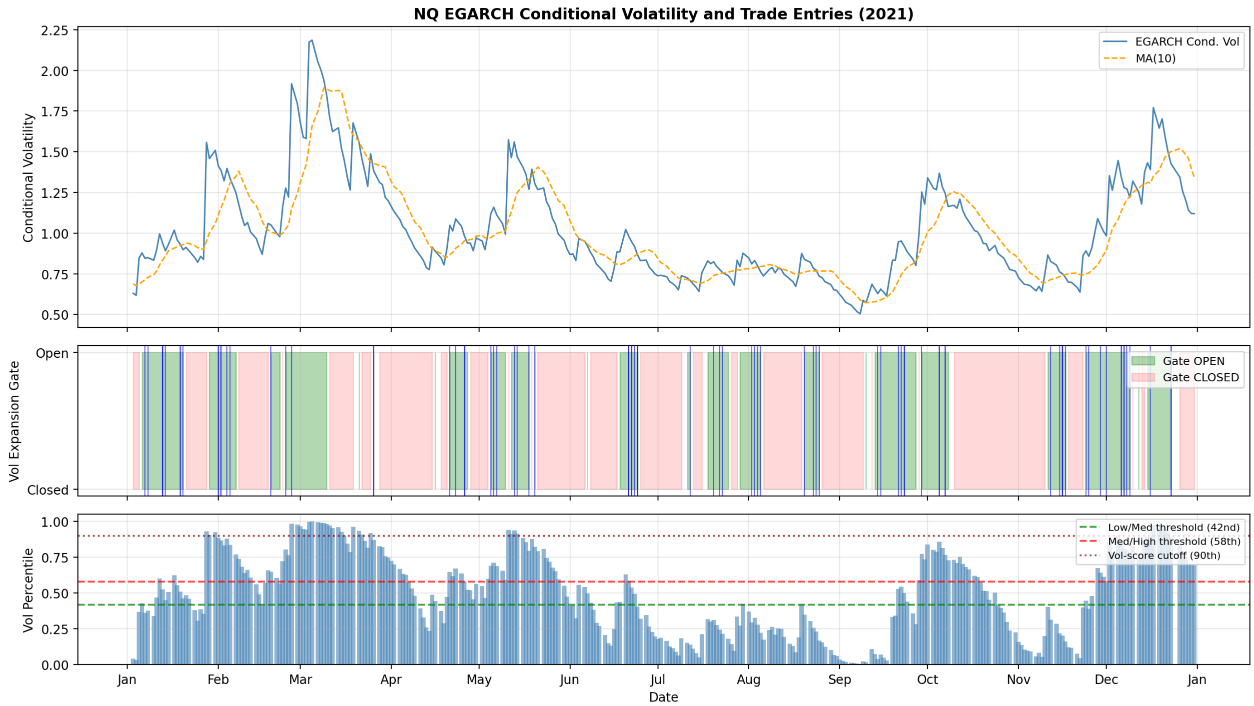

Figure 6: EGARCH volatility gate and trade entries for NQ during 2021. Top panel: daily EGARCH conditional volatility and its 10-day moving average. Middle panel: gate status (green = open, red = closed) with trade entry dates marked in blue. Bottom panel: volatility percentile rank with regime thresholds. The gate was open on 138 of 311 trading days, producing 80 trades concentrated in periods of expanding volatility.

The gate was open on only 138 of 311 trading days (44%), and trade entries cluster in periods where conditional volatility crosses above its moving average. The gate closes during sustained low-volatility periods and during extreme spikes (where the vol-score suppresses exposure).

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 OOS (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: OOS denotes the evaluation period. Trades reports total completed trades. Max DD denotes maximum drawdown. Return is the annual strategy return (full OOS row reports cumulative return). Year-level Sharpe and Sortino are annualized using the same factor as the full-period metrics.

The strategy records zero negative calendar years. The strongest years (2021: +85.3%, 2023: +85.9%, 2024: +74.5%) featured expanding volatility with directional momentum; the weakest (2020: +9.9%, 2022: +2.9%) correspond to the COVID crash and rate-shock environments where the gate curtailed exposure. The more recent years (2024–2025) provide evidence that the signal structure generalizes beyond the original development window.

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

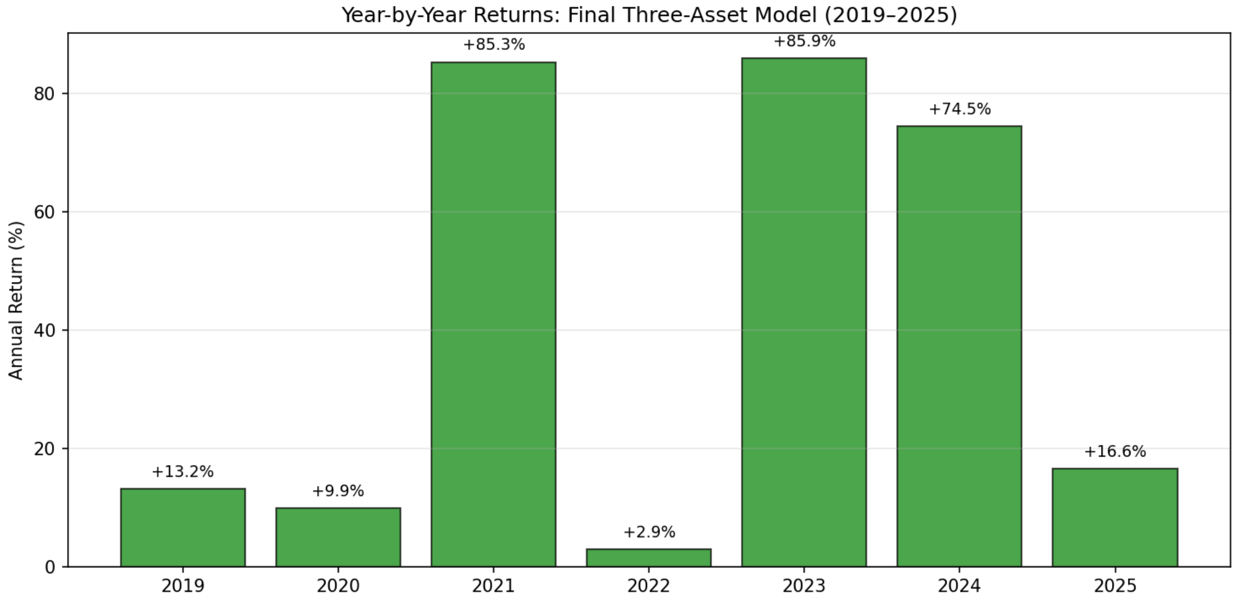

Figure 7: Year-by-year returns of the final three-asset model across the 2019–2025 evaluation period.

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× to 8× the baseline assumptions ($14.51 round-turn for NQ, $29.51 for ES, $12.80 for CL, each including exchange fees, clearing fees, and an estimated half-spread).

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.

At 2× baseline costs, the Sharpe ratio remains above 1.0 (1.329). The strategy becomes approximately breakeven at 8× baseline costs (Sharpe −0.024), corresponding to round-turn costs of approximately $116 for NQ, far above current market costs. This tolerance reflects the strategy's selectivity: with only

1,215 trades over seven years (approximately 0.7 per day), each trade carries sufficient expected profit to absorb realistic 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:

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

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***	0.0014***
	(3.34)	(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 × daily return, Volatility = ΔRV(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), indicating that the vast majority of the strategy's return variation is idiosyncratic.

The Fama–French factor loadings are informative about the strategy's economic character. The market beta of 0.210 ($t = 6.09$) is positive but small relative to a passive long position, consistent with the strategy's selective entry logic: the EGARCH gate and signal filters reduce effective market exposure to roughly one-fifth of a fully invested position. The SMB loading is economically and statistically insignificant ($t = 0.73$), as expected for a futures-based strategy that does not differentiate by firm size. The HML loading is negative and significant (-0.115, $t = -3.14$), indicating a growth tilt consistent with the NQ/ES equity index exposure.

Under the self-constructed specification, the momentum factor loading is insignificant ($t = -0.96$), which is notable because the strategy uses momentum-like entry signals (breakout, EMA crossover). This suggests that the EGARCH conditioning layer transforms raw momentum signals into something distinct from the standard momentum factor. The volatility factor loading is positive and significant (0.856, $t = 2.84$), confirming that the strategy benefits when aggregate volatility is expanding, directly consistent with the EGARCH expansion gate design. However, even this economically intuitive exposure explains only a small share of total return variance, reinforcing the conclusion that the strategy's performance derives primarily from the interaction between EGARCH conditioning and intraday signal logic rather than from passive factor exposure.

5.8 Expanding-Window Walk-Forward Validation

Table 13 reports an expanding-window walk-forward exercise in which volatility percentile ranks are re-estimated for each test year using only data available before January 1 of that year. All trading parameters are held fixed; the exercise eliminates look-ahead bias in volatility regime classification but does not constitute a fully out-of-sample test of the complete parameter set.

Table 13: Expanding-window walk-forward validation (2019–2025)

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: For each test year, volatility percentile ranks are re-estimated using only prior data. All other strategy parameters are fixed.

The strategy is profitable in all seven walk-forward years, with the weakest (2020, 2022) corresponding to the COVID crash and rate-shock environments. This confirms that returns are not concentrated in a single subperiod.

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 symmetric realized-volatility proxies in the tradition of Andersen and Bollerslev (1997) and Barndorff-Nielsen and Shephard (2002), while holding all other components fixed: RV(20), the 20-day rolling standard deviation of daily log returns, and RV(60), a slower 60-day variant. Both symmetric filters improve over the unfiltered baseline (RV(20): Sharpe 1.114; RV(60): 0.866; unfiltered: 0.664), confirming that volatility conditioning itself has value. EGARCH provides a further improvement to 1.412. The Sharpe margin of 0.298 between EGARCH and RV(20) is attributable specifically to the asymmetry channel, since the two filters use identical gate logic, vol-score, and IVT sizing. Both symmetric filters produce two negative calendar years that EGARCH avoids, because they cannot distinguish between crash-driven volatility expansions (where long breakouts tend to fail) and positive-shock-driven expansions (where they tend to succeed). This pattern is discussed further in Section 6.2.

5.10 Placebo and Falsification Tests

The ablation test in Section 5.2 establishes that removing the EGARCH gate degrades performance. A sharper question is whether the gate's value derives from correctly identifying the *current* volatility state or merely from suppressing trades. Two placebo exercises address this concern while holding all other parameters fixed.

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 diagnostic upper bound. 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.

Table 15 reports the results. The correctly timed gate (shift = 0) achieves the highest Sharpe ratio (1.412) among all non-anticipatory specifications. A single-day delay reduces the Sharpe to 0.884, and a five-day delay reduces it to 0.205. The look-ahead shift of -1 day reaches 1.297, slightly below the correctly timed gate, indicating that the baseline one-day lag is not accidentally benefiting from future information. The timing evidence therefore supports a contemporaneous mechanism rather than a generic on/off filter.

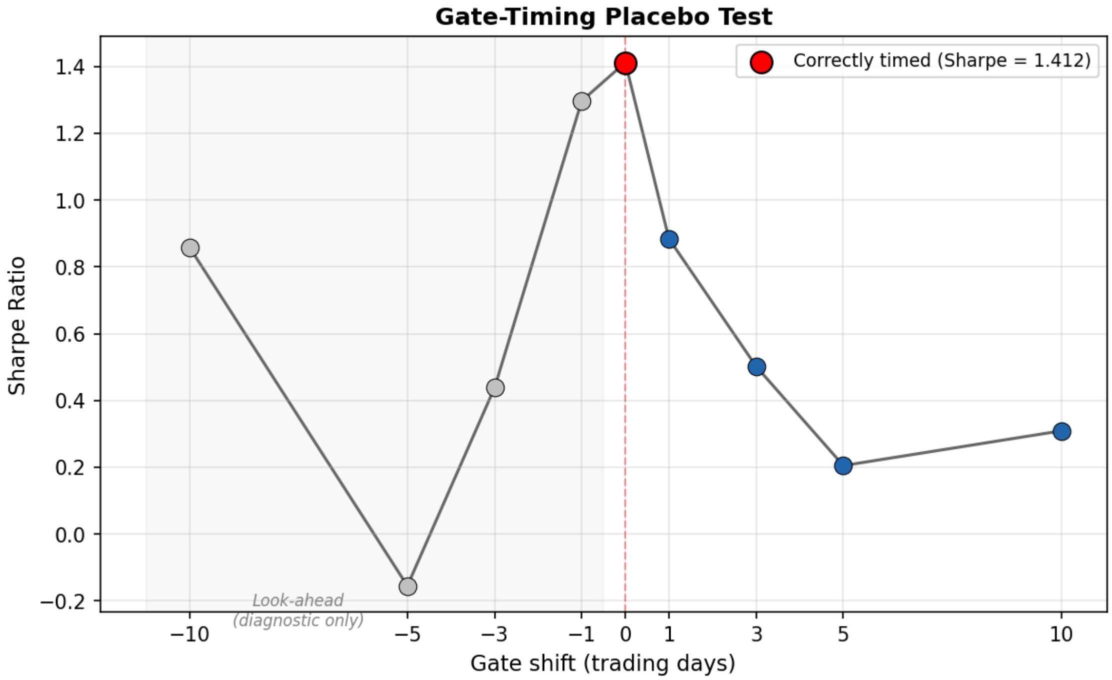

Figure 8: Gate-timing placebo test: Sharpe ratio under shifted EGARCH expansion gates. Shift = 0 is the correctly timed gate. Positive shifts delay the signal; negative shifts (shaded) use future information and are included as a diagnostic upper bound. The steep decay under positive shifts confirms that the specific timing of the volatility signal drives the result.

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. Because the placebo preserves gate frequency within asset and year, this result rejects the view that the observed gain comes from generic trade suppression alone. What it does *not* show is that every alternative clustered gate would fail; rather, it shows that the specific economic content of the EGARCH gate matters beyond simple frequency control.

5.11 Return Distribution and the Sharpe–Sortino Gap

At daily frequency the final model's Sortino ratio (1.579) exceeds its Sharpe ratio (1.412), consistent with the asymmetric-payoff profile of a trend-following strategy with trailing-stop exits. A methodological note is in order, however: if both ratios are recomputed on 5-minute bars rather than daily-aggregated returns, the Sortino falls to 0.605 against a 5-minute Sharpe of 1.555. The 5-minute Sortino looks anomalously low because every intra-trade dip below entry registers as downside variance even on trades that end profitably. Figure 9 presents the distribution of daily strategy returns alongside a normal benchmark.

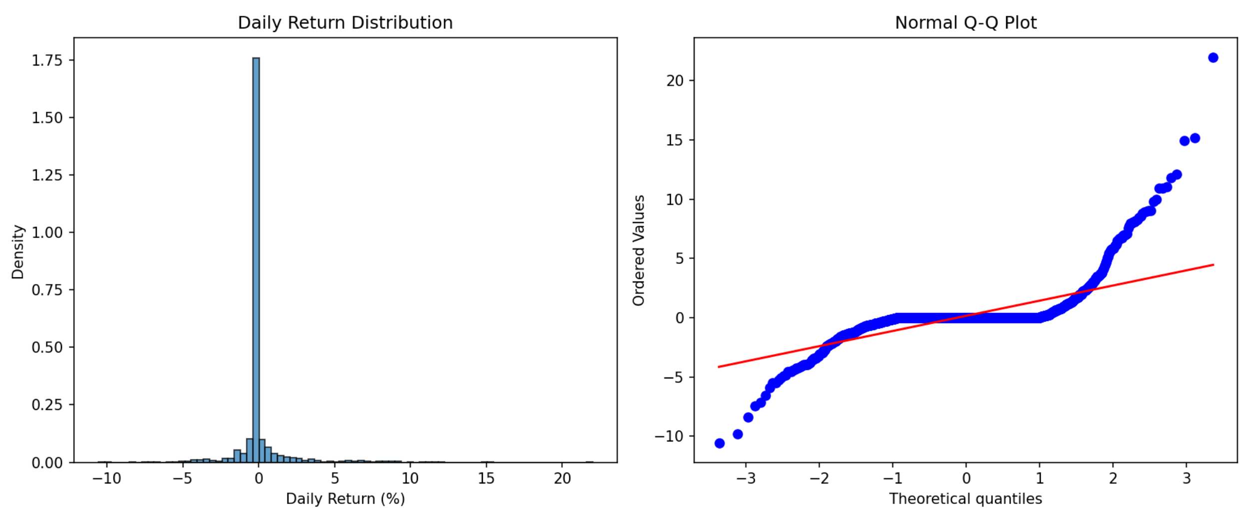

Figure 9: Daily return distribution of the final model (left) and normal Q-Q plot (right). The distribution exhibits positive skewness and substantial excess kurtosis, indicating fat tails and occasional large gains. Evaluation period: 2019–2025.

The daily return distribution is positively skewed with heavy tails and a substantial mass at zero (flat days). Among active days, loss days are less frequent and smaller on average but exhibit considerable dispersion, driving the Sortino ratio lower.

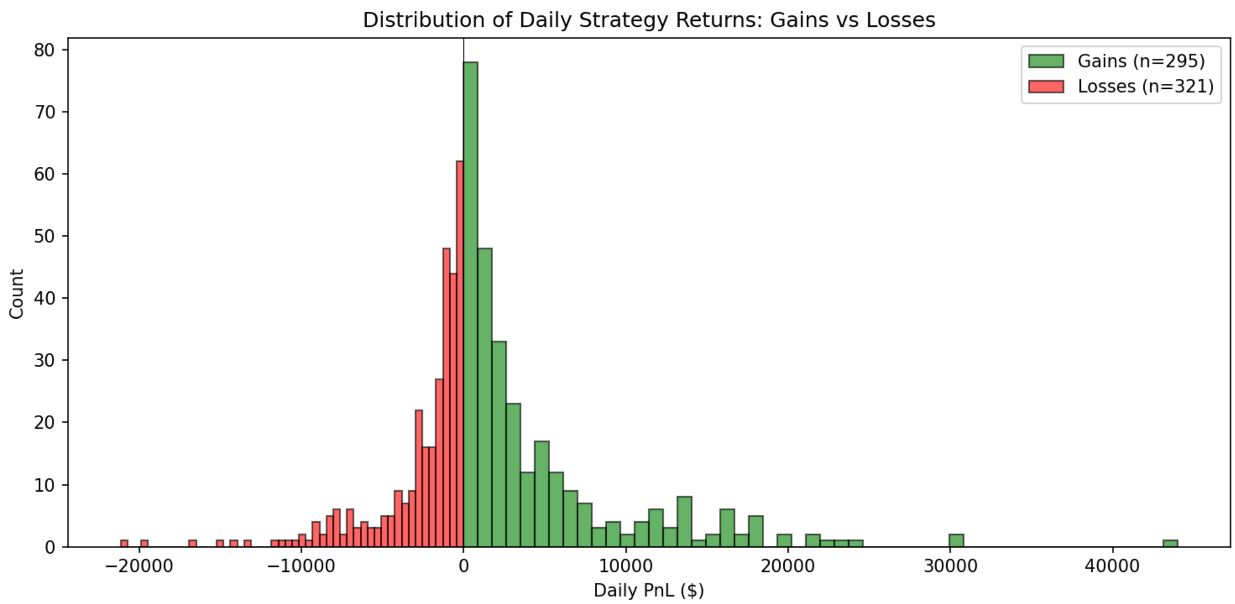

Figure 10: Distribution of daily strategy returns separated into gain and loss days. The positive mean, positive skewness, and fat-tailed loss distribution are consistent with the daily Sortino ratio of 1.579 exceeding the daily Sharpe ratio of 1.412. Evaluation period: 2019–2025.

The 5-minute Sortino–Sharpe inversion reflects the trailing-stop design under sub-daily aggregation: winning trades run until the stop triggers, producing occasional large gains, while intra-trade fluctuations register as downside variance even when the trade ultimately closes profitably (a trade that dips $1.5\times$ ATR before recovering to $3\times$ ATR profit contributes negatively to the 5-minute downside deviation). Daily aggregation is the standard convention in the trading-strategy literature (Lo, 2002; Bailey and López de Prado, 2014) precisely because it absorbs these intra-trade fluctuations into a single daily PnL observation, so only realized day-level losses contribute to downside deviation. Table 17 reports both conventions side by side.

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: 5-minute returns annualized by $\sqrt{78 \times 252}$; daily returns annualized by $\sqrt{252}$. Active days are days with non-zero strategy PnL. The 5-minute computation registers each intra-trade dip below entry as a downside observation, mechanically deflating the Sortino ratio relative to the daily aggregation in which intra-trade variance is absorbed into single daily PnL observations.

The two computations agree on the qualitative ranking (the strategy's risk-adjusted return is well above unity in both conventions) and on the substantive conclusion (downside risk to the investor is well-controlled at the day level). The thesis adopts the daily-frequency convention throughout the empirical results for consistency with the literature and with the tables; the 5-minute figures are reported here only as a methodological cross-check. The 10.6% maximum drawdown and zero negative calendar years confirm that realized day-level downside risk is well-controlled.

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
leverage	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.

The results show broad stability across all parameters. The narrowest range is leverage, which by construction scales returns and risk proportionally without affecting the Sharpe ratio. 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.

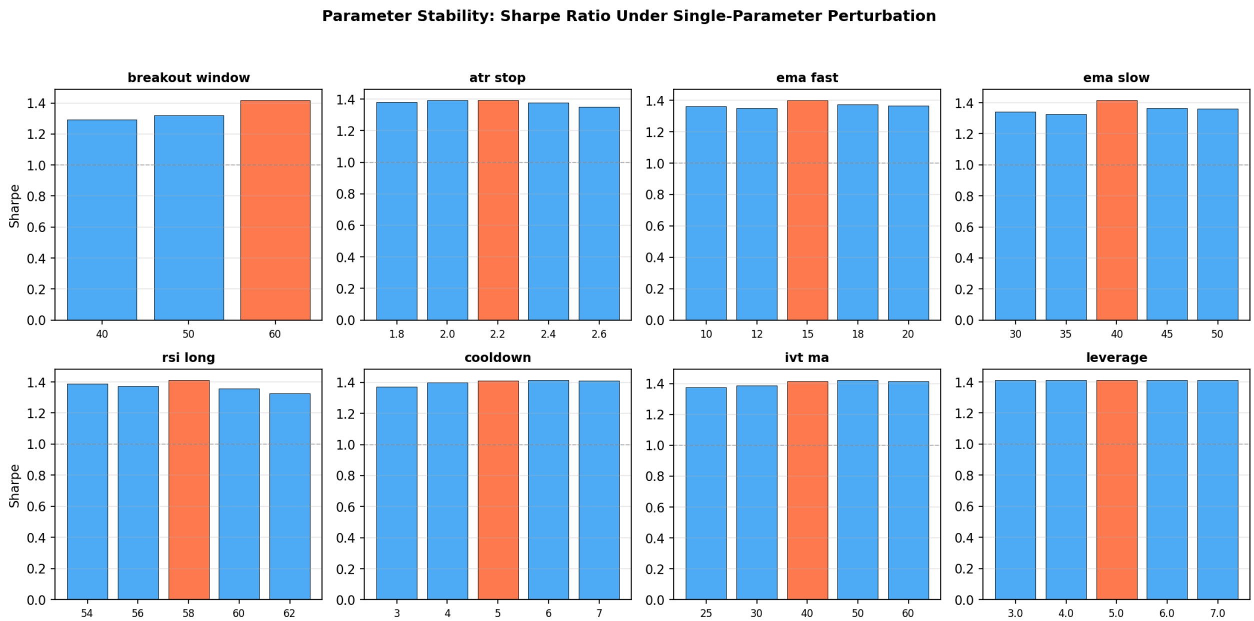

Figure 11: Sharpe ratio under single-parameter perturbation. Each panel varies one parameter while holding all others at champion values. The champion value is highlighted in red. The dashed line marks Sharpe = 1.0. Evaluation period: 2019–2025.

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.

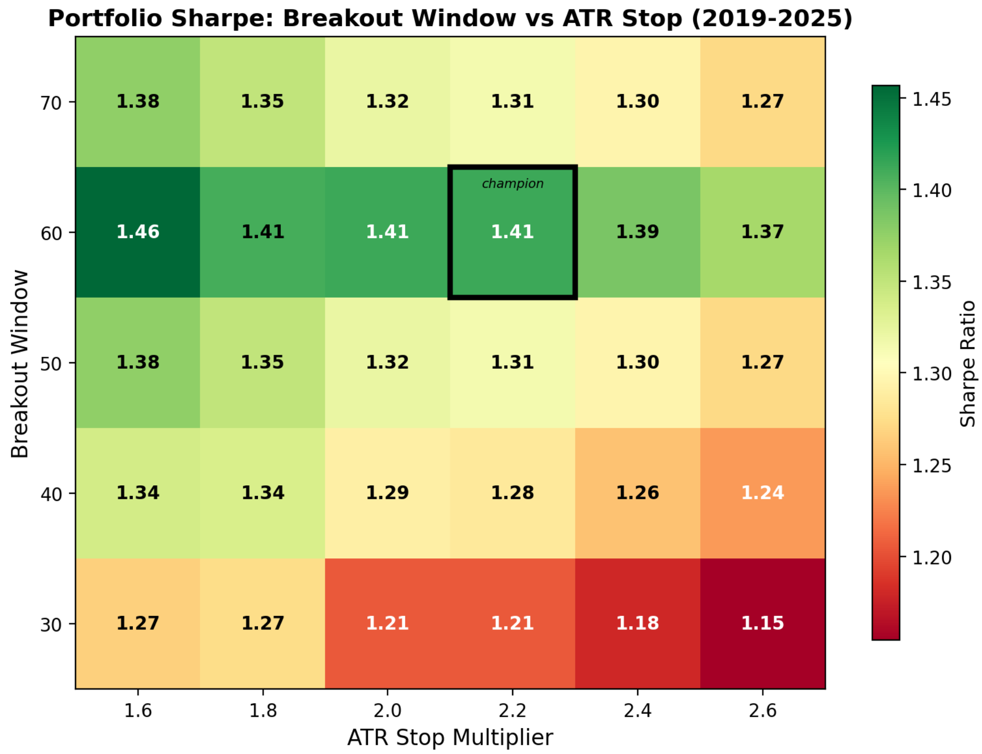

Figure 12: Joint sensitivity: breakout window $\times$ ATR stop multiplier (NQ/ES). Champion cell marked. All 30 combinations yield Sharpe > 1.15. Range: 1.15–1.46.

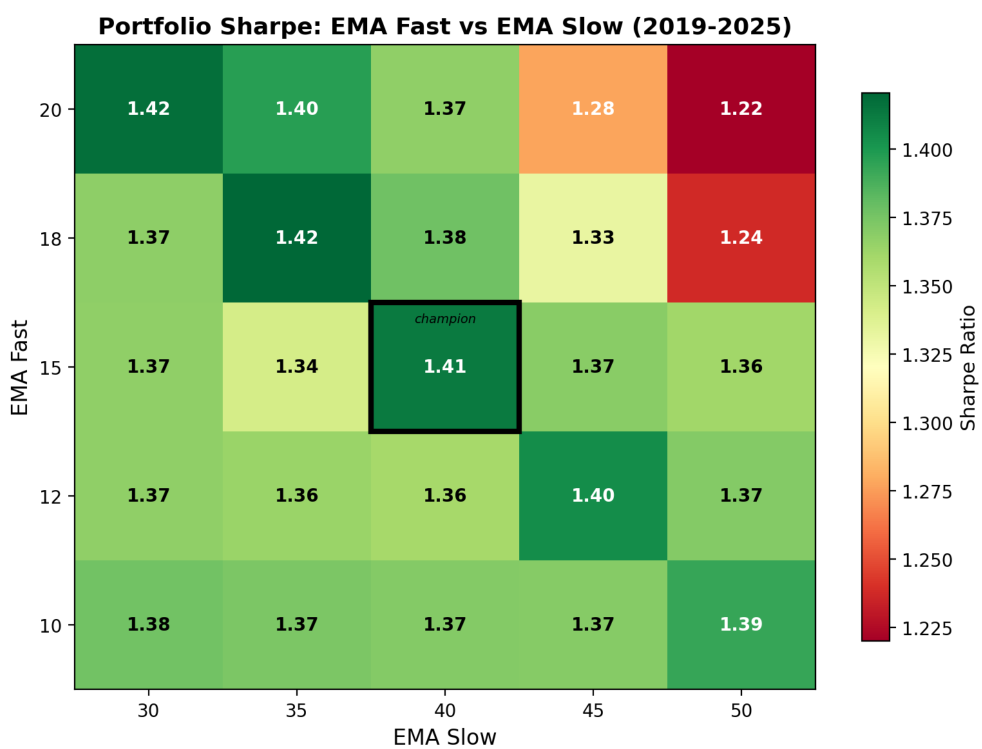

Figure 13: Joint sensitivity: EMA fast $\times$ EMA slow (NQ/ES). Champion cell marked. All 25 combinations yield Sharpe > 1.22. Range: 1.22–1.42.

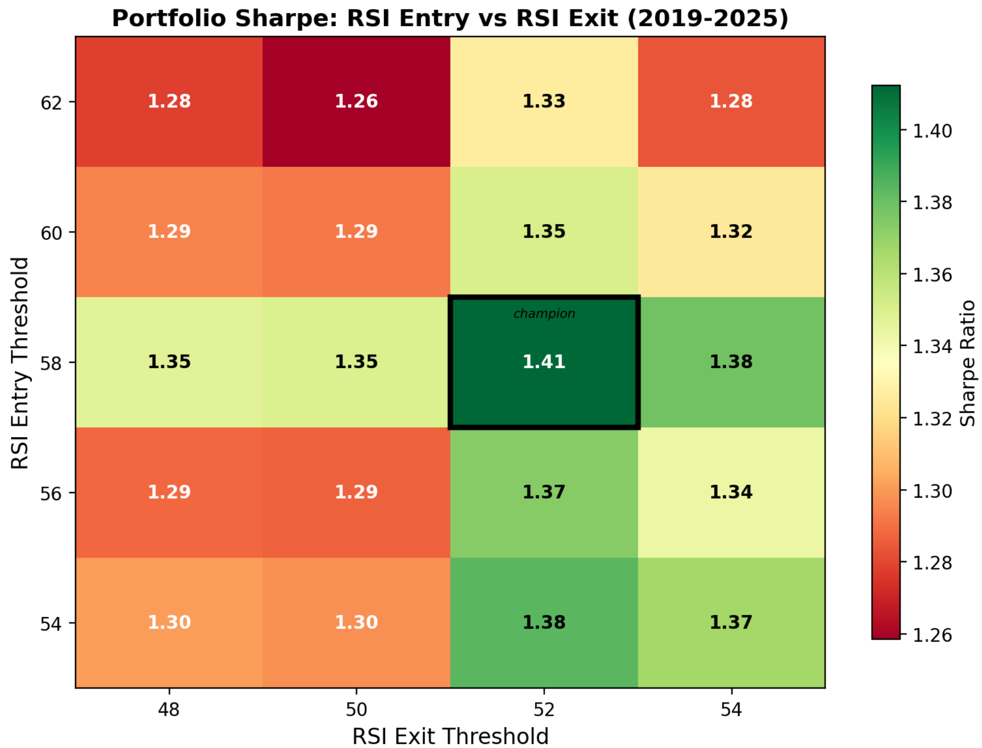

Figure 14: Joint sensitivity: RSI entry threshold $\times$ RSI exit threshold. Champion cell marked. All 20 combinations yield Sharpe > 1.26. Range: 1.26–1.41.

Three features of these surfaces reinforce the single-parameter results. First, all 75 parameter combinations across the three heatmaps produce Sharpe ratios above 1.15, with no cliff edges or isolated peaks. The performance surface is a smooth plateau in joint parameter space, not a fragile ridge. Second, the champion cell is never the global maximum in any heatmap; it sits near the top of its local region but is surrounded by comparably performing neighbors, which is inconsistent with the hypothesis that the champion was cherry-picked to maximize in-sample performance. Third, the RSI parameters exhibit the smallest variation (range: 0.15, standard deviation: 0.04), while the breakout window $\times$ ATR stop pair shows the most (range: 0.30), providing a clear hierarchy of which parameters matter most for robustness. The overall picture is one of a broad, smooth performance region in which the champion specification represents a reasonable but non-unique choice.

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: Deflated Sharpe Ratio under the Bailey–López de Prado correction and a unit-dispersion stress case

Trials (N)	$\mathbb{E}[\max Z_N]$	Corrected DSR[a]	Stress-case DSR[b]
6	1.400	100.0%	51.5%
15	1.878	100.0%	7.8%
30	2.180	100.0%	1.0%
50	2.381	100.0%	0.2%

Note: The observed annualized Sharpe of the champion's daily-aggregated return stream is 1.412 with standard error 0.328, computed using the Lo (2002) adjustment for skewness and kurtosis on $T = 1{,}762$ daily observations. This daily-frequency Sharpe is the development specification's headline figure reported in Section 5; the qualitative DSR conclusions are unchanged under the 5-minute cross-check reported in Table 17. [a]Corrected DSR follows Bailey and López de Prado (2014) with the cross-sectional standard deviation of Sharpe ratios across the parameter-stability family used as the multiplicative scale on the expected maximum draw. The family contains 38 one-parameter perturbations; the average pairwise daily-return correlation is 0.989, implying an effective independent trial count near unity and a cross-sectional Sharpe dispersion of $\sigma_{SR} = 0.032$. The selection-bias threshold is therefore $SR_0 = \sigma_{SR}\,\mathbb{E}[\max Z_N]$, which remains far below the observed Sharpe across all reported trial counts. [b]Stress-case DSR replaces the family dispersion with unit dispersion ($\sigma_{SR} = 1$), treating each tested specification as an independent strategy. Because the candidate specifications share the same EGARCH gate, breakout logic, exit structure, and asset universe, their realized return correlations are very high and unit dispersion is a deliberately severe assumption; the stress case is reported as a conservative bound rather than as the maintained inference.

Table 19 can be read two ways. Under the Bailey–López de Prado formulation, which scales the selection-bias threshold by the empirically observed dispersion of Sharpe ratios across the tested family, the corrected DSR exceeds 99.9% across all reported trial counts. The local parameter family explored during development does not consist of independent strategy ideas. It consists of small perturbations around a single architecture, with realized return correlations near unity. The cross-sectional dispersion of family Sharpe ratios is therefore small ($\sigma_{SR} = 0.032$), and the implied selection-bias benchmark is small as well. On this reading, the champion Sharpe is unlikely to be a pure artifact of selection over the tested family.

The stress-case row reports what happens if one ignores the family-dispersion correction and treats each tested specification as if it were a fully independent strategy. Under that severe assumption the DSR falls quickly with N and fails at $N = 15$ and beyond. The thesis reports this case explicitly because development iteration over more than thirty configurations is a real feature of the project, and the more conservative reading should be available to the reader. The two rows bracket the inference. Under the empirically calibrated correction the selection-bias concern is small; under the unit-dispersion bound it is large.

The DSR addresses the *level* of the champion Sharpe and is most relevant to the development-exposed specification. The pre-OOS specification with zero parameter exposure to the evaluation window achieves Sharpe 1.195 and provides a complementary test that does not invoke the DSR machinery in the same way. The DSR also does not directly bound the ablation effect, which is a within-specification comparison unaffected by how many variants were tested, or the gate-timing placebo result, which varies the gate's timing rather than its parameters. The thesis therefore treats the corrected DSR as evidence that the data-mining objection is weakened but not dismissed, while the EGARCH mechanism's contribution continues to rest primarily on the ablation, placebo, and pre-OOS evidence.

5.14 Development Period versus Holdout

The strongest test of overfitting is performance on data that played no role in parameter selection. Table 20 splits the evaluation period into a development window (2019–2023), during which all parameters were chosen, and a pure holdout window (2024–2025), which was not used for any design decision.

Table 20: Development period vs. pure holdout performance

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

Note: Development period (2019–2023) is the window during which strategy parameters were selected. Holdout period (2024–2025) was not used for any parameter choice. All specifications use identical parameters and $200,000 initial capital.

The holdout Sharpe ratio (1.197) is lower than the development-period Sharpe (1.594), consistent with the expected out-of-sample decay from iterative refinement. The holdout Sharpe nevertheless remains well above 1.0, both holdout years are individually profitable (2024: 74.5%, 2025: 16.6%), and the EGARCH ablation effect persists: removing the EGARCH filters in the holdout period causes the Sharpe to fall to 0.449 and maximum drawdown to spike to 51.7%. The holdout No-EGARCH specification produces a comparable raw return (1.97× versus 1.91×), but at drastically worse risk, confirming that the EGARCH layer's contribution to risk-adjusted performance generalizes beyond the development sample.

The Sharpe decay from 1.594 to 1.197 quantifies the cost of parameter selection and suggests that a prospective Sharpe in the range of 1.0–1.2 is more realistic than 1.412. Because the champion parameters were still selected on 2019–2023 data, this split controls for temporal overfitting but not for parameter exposure. The pre-OOS test in Section 5.15 addresses this stronger concern.

5.15 Pre-OOS Parameter Selection

The holdout split in Section 5.14 uses the champion parameters, which were refined over 2019–2023, and evaluates them on 2024–2025. A stronger test eliminates any parameter exposure to the evaluation period entirely. Table 21 reports the results of a pre-OOS parameter selection exercise in which all strategy parameters are chosen by sequential grid search on the 2010–2018 training period, with zero exposure to any data after December 2018. The selected parameters are then applied without modification to the full 2019–2025 evaluation window.

Table 21: Pre-OOS parameter selection: training on 2010–2018, evaluation on 2019–2025

Model	Sharpe	Sortino	Max DD	Return Mult.	Trades
Pre-OOS selected + EGARCH	1.195	1.314	−14.0%	3.36×	1302
Champion + EGARCH	1.412	1.579	−10.6%	3.88×	1215
Pre-OOS 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: Pre-OOS parameters were selected by sequential grid search on 2010–2018 data with zero exposure to the 2019–2025 evaluation period. The champion parameters were iteratively refined over the evaluation period. All Sharpe and Sortino ratios computed on daily portfolio returns; both specifications use $200,000 initial capital and identical transaction costs.

The pre-OOS selected specification achieves a Sharpe ratio of 1.195 on the 2019–2025 evaluation period, with a cumulative return of $3.36\times$ 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 is consistent with architectural generalization across sample periods, with later refinement affecting calibration more than design.

This is the thesis's headline result: the cleanest available estimate of generalization performance, with zero parameter exposure to the evaluation data. The champion Sharpe of 1.412 should be interpreted as an upper bound reflecting development refinement.

Two features reinforce this interpretation. First, the training-period Sharpe on 2010–2018 was only 0.066. This gap does not show that the strategy fails outside 2019–2025; rather, it suggests the strategy's edge is conditional on a regime feature that was less prevalent in the earlier window. The mechanism is visible in a simple comparison. The EGARCH gate fires at virtually identical rates across periods. NQ, ES, and CL show 41.97%, 43.18%, and 42.09% of trading days as gate-active over 2010–2018, versus 40.03%, 39.81%, and 38.13% over 2019–2025. The trade counts per asset per year are likewise comparable. What differs is the underlying drift conditional on those candidate days. Buy-and-hold returns averaged 14.2% (NQ), 9.2% (ES), and -2.4% (CL) annually over 2010–2018, compared with 25.8%, 17.0%, and $+6.9\%$ over 2019–2025. Equity drift roughly doubled, and the CL drift sign flipped.

A long-only conditional momentum strategy can only monetize gate-active days when the underlying provides positive drift on those days. The 2010–2018 environment delivered substantially less of that, and CL delivered the opposite sign. The 0.066 training Sharpe is therefore consistent with a conditional mechanism that requires both ingredients (volatility-conditioned admissibility *and* positive underlying drift on admissible days), rather than evidence that the mechanism itself is fragile. Second, the EGARCH ablation effect persists in full under the pre-OOS parameters: removing the EGARCH filters causes the Sharpe to collapse from 1.195 to 0.382 and maximum drawdown to spike to 49.9%, confirming that the conditioning effect is robust to parameter choice and not an artifact of iterative refinement.

5.16 Rolling Parameter Re-Selection

The pre-OOS exercise just discussed is a one-shot selection: parameters are chosen on 2010–2018 and held fixed for the entire 2019–2025 evaluation. A stricter test re-selects parameters continuously on rolling training windows, producing a sequence of out-of-sample evaluation periods in which the parameters applied to each evaluation year had zero exposure to that year. This exercise complements the earlier expanding-window validation in Section 5.8, which re-estimated only the volatility percentile ranks while holding trading parameters fixed; here the parameter choice itself is re-estimated annually.

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. The resulting 14-year track record is fully out-of-sample in the sense that no parameter has prior exposure to the year in which it is evaluated.

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.

Two observations matter for the interpretation. First, the walk-forward Sharpe over the full 2012–2025 evaluation (0.939) is essentially identical to the champion-fixed Sharpe (0.962) over the same window. The 0.023-Sharpe gap is small relative to the standard error of either estimate, and it reflects the fact that the parameter search is restricted to a small two-dimensional grid rather than the full champion specification. The substantive conclusion is that the strategy's edge does not depend on the specific parameter values used in the headline champion. Re-selecting parameters annually from rolling training windows recovers nearly the same risk-adjusted return.

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) because the rolling design adapts to the pre-2019 regime in which the champion parameters (selected on 2019–2025) are not the right choice. 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 walk-forward complements the pre-OOS exercise rather than replacing it. The pre-OOS specification fixes one architectural choice (a single parameter set chosen on 2010–2018), while the walk-forward continuously varies the parameter choice. Both arrive at performance estimates well above zero across 2012–2025, but each tests a different concern. The pre-OOS test answers, "what if parameters had been locked in at the end of 2018?"; the walk-forward answers, "what if parameters had been re-selected at each year-end since 2012?". The two together provide a hierarchy of stringency under which the strategy's risk-adjusted return remains positive and economically meaningful, conditional on the maintained interpretive caveats about regime dependence developed in Section 5.15.

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: EGARCH improves the admissibility of intraday long breakout signals within the observed sample. The ablation result holds under both the development specification (Sharpe 1.412 $\rightarrow$ 0.664) and the pre-OOS specification (1.195 $\rightarrow$ 0.382), with the circular block bootstrap confirming the difference at the 1$\%$ level. The placebo evidence points in the same direction: even a one-day misalignment of the gate reduces the Sharpe by 0.528, and the randomized frequency-preserving placebo yields mean Sharpe 0.561 with empirical placebo $p = 0.002$. Because this claim rests on comparative evidence within the same design, it is more credible than any claim about the absolute Sharpe level.

Claim 2: The pre-OOS specification provides the cleanest performance estimate. The pre-OOS specification achieves a Sharpe of 1.195 and a cumulative return of 3.36$\times$ with zero parameter exposure to the evaluation window. The development figure of 1.412 overstates this by approximately 15% and should be treated as an upper bound. The pre-OOS result is therefore the headline estimate, conditional on the maintained assumption that the 2010–2018 training period remains informative for 2019–2025.

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 EGARCH expansion gate implements this distinction in practice. By requiring that conditional volatility be locally rising, it selects for trading days on which the information-arrival rate is elevated, and therefore on which breakout signals have a higher probability of reflecting 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 channel explains why the expansion *gate* (a binary condition on whether volatility is rising) is more important than the volatility *level*: it is the transition from low to expanding volatility, not the level itself, that alters the microstructure environment.

Participation feedback and order-flow amplification. Expanding volatility attracts discretionary and algorithmic flow that amplifies directional moves (Bollen and Whaley, 2004). Trend-following algorithms activate when volatility crosses internal thresholds; discretionary traders increase participation when price action becomes directional. This feedback loop means that a breakout during expanding volatility is more likely to be *sustained* by subsequent order flow than a breakout during flat or declining volatility, where the absence of these amplifying forces makes the signal more likely to represent a stale limit order being hit rather than a shift in market consensus.

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. The channels above explain why the gate should matter at all; the next subsection turns to why EGARCH in particular, rather than a symmetric volatility measure, is the right filter.

6.2 Why Asymmetry Matters for a Long-Only Framework

All three contracts exhibit statistically significant negative EGARCH asymmetry ($\gamma < 0$): negative shocks generate larger volatility increases than positive shocks of comparable magnitude. This asymmetry creates a structural alignment between EGARCH conditioning and long-only execution that symmetric volatility filters cannot replicate.

The logic is as follows. A symmetric filter treats all volatility expansions equally, regardless of whether the expansion was precipitated by a negative or positive shock. For a long-only framework, this is a consequential omission. A volatility expansion driven by a sharp market decline reflects a compression of risk premia, elevated uncertainty, and a microstructure environment dominated by forced selling and risk reduction. These are conditions under which long breakouts are most likely to fail. A volatility expansion driven by positive price action reflects increasing information arrival and order-flow momentum in the direction favorable to long positions. EGARCH distinguishes between these two cases. When volatility expands following a negative shock, the asymmetry term ($\gamma < 0$) pushes the conditional variance estimate higher than a symmetric model would, moving it into the upper tail of the distribution where the vol-score tent function reduces or eliminates exposure. When volatility expands following positive action, the asymmetry term produces a more moderate variance estimate, keeping exposure within the range where breakout signals are historically informative.

The empirical comparison against symmetric realized-volatility alternatives isolates this channel. EGARCH (Sharpe 1.412) outperforms the best symmetric filter, RV(20), which achieves 1.114, even though both improve substantially over the unfiltered baseline (0.664). The performance gap of 0.298 in Sharpe between EGARCH and RV(20) is attributable to the asymmetry channel, since the two filters use the same gate logic, vol-score, and IVT sizing, differing only in the volatility input. The symmetric filters produce two negative calendar years that EGARCH avoids, consistent with the hypothesis that they open the gate during crash-driven expansions where long breakouts carry adverse information.

The EGARCH-versus-RV(20) comparison is informative but indirect: it isolates the gap by holding all other components fixed and varying only the volatility input. A direct test of the asymmetry channel goes one level deeper and asks whether the mechanism operates at the level the narrative above claims, namely through the vol-score distribution. The mechanism implies a specific empirical signature: gate-active days following a negative EGARCH standardized residual should have higher conditional volatility ($\gamma < 0$ amplifies variance after negative shocks), higher vol_pct, and lower vol_score, with more mass in the upper-tail suppression zone of the vol-score tent function. Table 23 reports this comparison for the 2019–2025 evaluation window.

Table 23: Direct test of the asymmetry channel: 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 positive values are consistent with the $\gamma < 0$ prediction. The standardized residual on day $t-1$ is approximated by $r_{t-1}/(\hat{\sigma}_{t-1}/100)$, where r_{t-1} is the daily log return and $\hat{\sigma}_{t-1}$ is the EGARCH conditional volatility estimate. The gate-active subsample contains 885 days for ES (POS: 493, NEG: 390), 890 for NQ (POS: 517, NEG: 372), and 850 for CL (POS: 431, NEG: 416). $\Delta\,\%_{>0.80}$ is the difference in the share of gate-active days falling into the vol-score suppression zone (vol_pct > 0.80, where exposure rolls off linearly); $\Delta\,\%_{>0.90}$ is the difference for the zero-exposure zone (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 (vol_pct > 0.80) and the zero-exposure zone (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.

This direct test isolates the asymmetry channel as operating through size reduction on the days where, by the leverage effect, long-only execution is most exposed to panic-driven entries. The effect manifests upstream of trade entry, at the vol-score distribution itself. The downstream consequence is that the worst gate-active days following negative residuals do not become trades at all (zero exposure), while the moderate ones execute at reduced size, which is the specific sense in which $\gamma < 0$ protects a long-only conditioning framework from a failure mode that symmetric volatility filters cannot avoid.

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 gate cuts exposure back during more extreme episodes (2020, 2022). So the thesis is not claiming unconditional profitability. The narrower claim is that a conditional mechanism appears to operate within an identifiable part of the state space, and that it steps back when conditions move outside it.

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 would cause the EGARCH gate to correctly suppress most entries, but the few entries that pass the filter would face adverse conditions. 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.

What the evidence supports is that the mechanism is present and identifiable within the 2019–2025 sample. 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

CL is included for portfolio-construction reasons, not because it has the strongest standalone signal quality. The thinner standalone signal in commodity futures is consistent with the broader commodity-timing literature (Marshall et al., 2008), which documents heterogeneous timing-strategy performance across commodities and underscores the role of diversification rather than universal signal strength. Its near-zero daily PnL correlation with NQ+ES ($\rho = 0.001$) provides genuine diversification, improving the portfolio Sharpe from 1.350 to 1.412 and eliminating the single negative calendar year (2022) that the two-asset variant produces. The EGARCH(1,1) model is less precisely specified for CL (Section 4.2), which helps explain its lower standalone Sharpe (0.479), but weaker standalone fit does not invalidate portfolio inclusion. CL's role is a bounded, low-correlation positive-expectancy source whose value is variance reduction, and the regime caps and IVT smoothing are important precisely because CL is the noisier contract.

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.

First, the factor exposure regression (Section 5.7) yields a market beta of only 0.210, indicating that the framework captures roughly one-fifth of equity market movements. The majority of its return variance is idiosyncratic ($R^2 = 3.0\%$ under Fama–French). Second, the framework's strongest years (2021, 2023) do not coincide mechanically with the equity market's strongest calendar-year returns, and the 2022 equity decline (-19% for the S&P 500) produced a small positive return (+2.9%) rather than a loss. Third, the pre-OOS parameters, selected on 2010–2018 when the training Sharpe was only 0.066, still generalize to 1.195 on the later window; the diagnostic in Section 5.15 attributes the training-window weakness to a difference in underlying drift on gate-active days rather than to a failure of the conditioning mechanism, which the equivalence of gate firing rates across periods supports. Fourth, the EGARCH ablation effect persists within 2024–2025 as well as the full period, indicating that the conditioning mechanism contributes independently of the equity tailwind.

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: parameter selection on the evaluation window. The champion Sharpe of 1.412 is an upper bound because parameters were iteratively refined on the 2019–2025 evaluation period.

The thesis measures this bias through a hierarchy of controls of increasing stringency: parameter stability, DSR adjustment, a holdout split (Sharpe 1.197 on 2024–2025), a pre-OOS exercise that eliminates parameter exposure entirely (Sharpe 1.195), and a rolling parameter re-selection design that re-fits parameters annually on 24-month windows across 2012–2025 (Sharpe 0.939 over 14 fully out-of-sample years). The 15% gap between 1.412 and 1.195 quantifies the cost of refinement; the rolling re-selection design recovers a Sharpe ratio within 0.023 of the champion-fixed comparison over the same 14-year window, indicating that the strategy's edge is not concentrated in the specific parameter values chosen on the development window. The ablation and placebo results, as within-design comparisons, are less exposed to this threat than the raw Sharpe level. *Would resolve:* a fully out-of-sample evaluation on post-2025 data.

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 mechanism operates 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 realism: execution assumptions. The thesis includes conservative transaction costs (breakeven at approximately $8\times$ baseline) but does not model market impact. For the highly liquid contracts and moderate trade frequency (0.7 trades per day) used here, this is reasonable, but live implementation may diverge, particularly at higher leverage. *Would resolve:* live execution data or a market-impact model calibrated to order-book data.

Model-choice uncertainty: volatility specification. EGARCH(1,1) does not fully capture CL's volatility dynamics (persistent residual ARCH effects; Section 4.2), and more sophisticated benchmarks (realized-GARCH (Hansen et al., 2012), regime-switching structures, stochastic volatility) remain untested. The comparison against symmetric filters demonstrates that asymmetry matters, but not that EGARCH is uniquely optimal. *Would resolve:* testing the admissibility framework with alternative volatility estimators while holding all other components fixed.

6.7 Future Research

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

Out-of-sample validation. The single most valuable extension is a fully out-of-sample evaluation on post-2025 data, an environment that does not yet exist when this thesis is written. The rolling parameter re-selection exercise in Section 5.16 addresses the within-sample version of this concern by annually re-fitting parameters on 24-month windows across 2012–2025; an external next step is to apply the same design as fresh data accrues, particularly to test whether the conditional mechanism survives a market environment that lacks the favorable directional drift documented in the 2019–2025 window.

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 EGARCH asymmetry advantage is robust to alternative specifications or partly reflects the choice of volatility estimator.

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 has asked whether EGARCH volatility conditioning can be used as an admissibility mechanism for intraday breakout signals in futures markets. The evidence supports three conclusions, which I again list in rough descending order of how well they are identified.

The strongest conclusion is that EGARCH conditioning improves the admissibility of intraday long breakout signals within the observed sample and design. The evidence is comparative rather than level-based: ablation causes severe deterioration under every specification tested, the gate-timing placebo shows that contemporaneous alignment matters, and the randomized frequency-preserving placebo yields mean Sharpe 0.561 with empirical placebo $p = 0.002$. The comparison against symmetric filters further isolates the asymmetry channel. Taken together, these results support the claim that EGARCH helps identify volatility states in which breakout signals are more likely to contain directional information.

The second conclusion concerns performance measurement. The pre-out-of-sample specification, with zero parameter exposure to the 2019–2025 evaluation window, achieves a Sharpe of 1.195 and a cumulative return of 3.36× on $200,000 initial capital. The development specification achieves 1.412, but this figure overstates prospective performance by approximately 15% and should be treated as an upper bound. The 2010–2018 training window itself had a Sharpe of only 0.066, but the diagnostic in Section 5.15 attributes this to a difference in underlying drift conditional on gate-active days rather than to a failure of the conditioning mechanism: the gate fires at virtually the same rate in both windows, while the underlying drift is materially weaker in the earlier window (and sign-flipped in CL). The training Sharpe is therefore consistent with a conditional mechanism that requires both volatility-conditioned admissibility and positive underlying drift on admissible days. The rolling parameter re-selection exercise in Section 5.16 provides a further out-of-sample check: annually re-fitting parameters on 24-month windows across 2012–2025 yields a Sharpe of 0.939 over 14 fully out-of-sample years, within 0.023 of the champion-fixed Sharpe (0.962) over the same span, indicating that the strategy's risk-adjusted return does not depend on the specific parameter values selected during development.

The third conclusion concerns scope. The conditioning framework is not a general-purpose trading rule: it is identified inside a specific domain, the volatility environments in which the microstructure channels discussed in Section 6 are most active, and the evidence is limited to the observed sample and design. What is inside that domain is well identified; what lies outside it is not a claim the thesis makes.

The most important open question is whether the mechanism preserves capital during a prolonged multi-year bear market, since the evaluation period does not contain one. The EGARCH(1,1) specification also does not fully capture CL's volatility dynamics, and the three-contract design demonstrates applicability across the observed contracts rather than universality.

The thesis's contribution is methodological, but the point can be stated simply. It uses EGARCH at the trade level, as a way of deciding whether an individual intraday breakout should be traded at all, rather than using it only as a variance forecast or as an input into portfolio-level scaling. Within the sample and design studied here, the evidence shows that this mechanism is both identifiable and economically relevant. Whether the same pattern carries over to other markets, other periods, or different microstructure settings is not something I can settle with the data used here, so I leave that question open.

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. Each modification was economically motivated rather than arbitrarily selected.

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 the cumulative return multiple. 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.

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 portfolio-level breakeven of 8× reported in the main text exceeds the standalone breakeven for both ES and CL because cross-asset diversification reduces portfolio-level volatility without proportionally reducing gross profits. This confirms that the multi-asset design contributes to cost robustness beyond what any single contract achieves alone.

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.

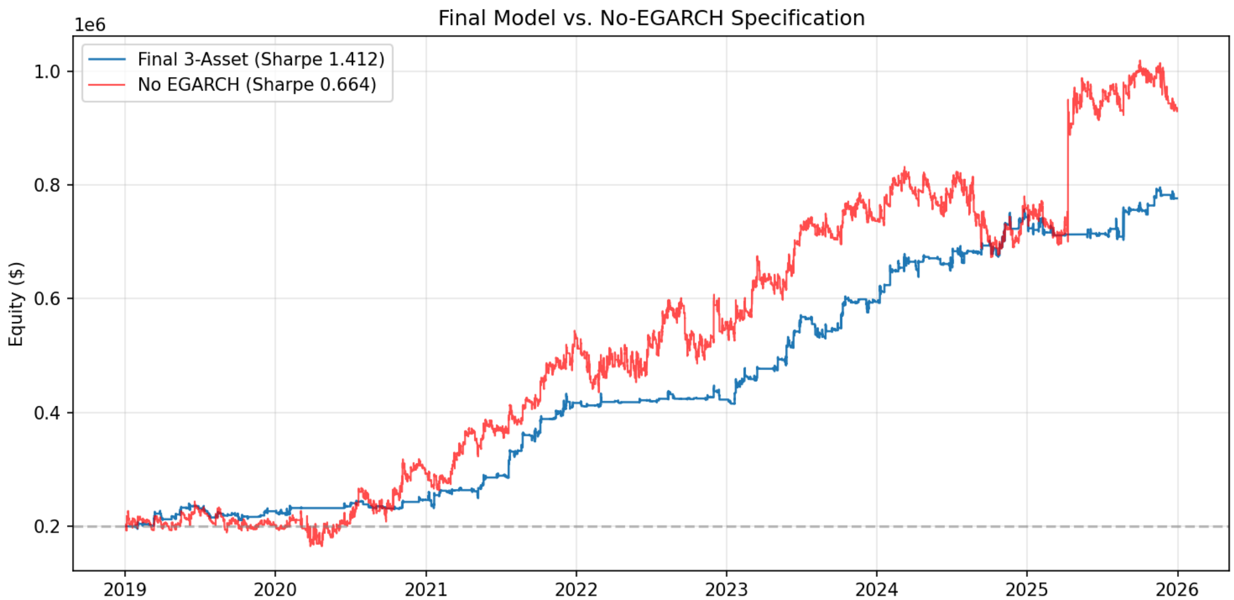

Figure 15: Equity comparison between the final three-asset model and the No-EGARCH specification.

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

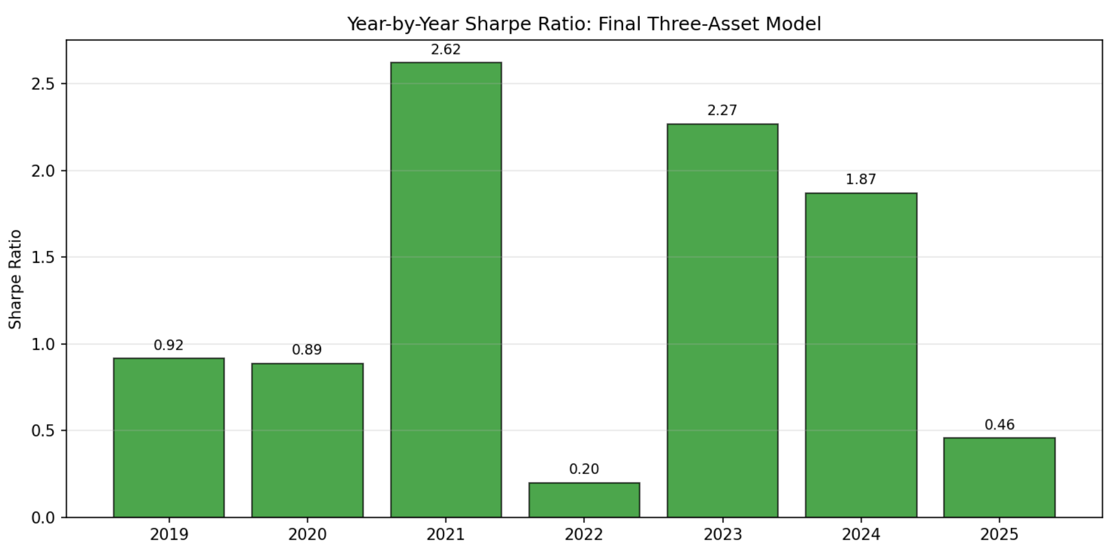

Figure 16: Sharpe ratio across the selected subperiods of the final three-asset model.

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.

The results confirm that the long-only restriction is empirically justified for CL as well. The short-only specification achieves a Sharpe ratio of just 0.122 with a 16.2% maximum drawdown and three negative calendar years, compared with 0.479 and two negative years for the long-only version. The combined long and short specification (Sharpe 0.433) adds 342 short trades but produces nearly identical risk-adjusted performance to the long-only version while increasing the number of negative years from two to three. The short side generates roughly equal trade volume but with substantially worse signal quality, diluting the portfolio's overall edge. This pattern is consistent with the observation that breakout continuation is a stronger phenomenon on the upside than the downside in energy futures, where mean-reversion forces are more prominent after price declines due to supply-response dynamics. The long-only restriction therefore reflects a genuine asymmetry in CL signal quality rather than an arbitrary design choice.

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, confirming low cross-asset dependence.

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, confirming near-zero cross-asset dependence. 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 Leverage Sensitivity

The IVT leverage multiplier L determines the baseline aggressiveness of position sizing. Table 29 reports strategy performance across a range of leverage levels from $1\times$ to $15\times$ for both the EGARCH champion and the No-EGARCH specification, holding all other parameters at their champion values.

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

	With EGARCH				Without EGARCH			
Leverage	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 champion specification (5×, bold) is varied by changing only the IVT leverage multiplier L. Sharpe ratios are computed on daily portfolio returns and are leverage-invariant by construction. Ret/DD is defined as (Return Mult. -1) / |Max DD|. Maximum drawdowns scale approximately linearly with leverage. Both specifications record zero negative calendar years at all leverage levels. The No-EGARCH specification removes the volatility expansion gate, vol-score, IVT sizing, and regime caps, replacing them with fixed position sizing at the stated leverage.

The Sharpe and Sortino ratios are invariant to leverage within each specification, as expected: both the mean and standard deviation of returns scale proportionally with position size. Because leverage does not affect the Sharpe ratio, it does not create the strategy's edge; it merely scales the consequences of an edge that is already present (or absent). The more informative comparison is therefore the behavior of maximum drawdown and the return-to-drawdown ratio across the two specifications.

For the EGARCH model, maximum drawdown increases less than proportionally with leverage within the tested range: tripling leverage from $5\times$ to $15\times$ 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 leverage tested, the EGARCH model's maximum drawdown is substantially lower than that of the No-EGARCH version. At $10\times$ leverage, 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 $5\times$ leverage, the EGARCH model achieves a Ret/DD of 23.23 compared with 11.23 for the unfiltered version. This approximately twofold advantage in drawdown efficiency persists across all tested leverage levels, suggesting that EGARCH conditioning produces a materially more favorable leverage-drawdown tradeoff 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.

Two caveats apply. First, these results assume sufficient margin to sustain all drawdowns without forced liquidation. In practice, leverage beyond approximately 7–10$\times$ would expose the strategy to margin call risk during peak drawdown episodes, a risk that is not modeled in the backtest framework and that the Sharpe ratio does not capture. Second, the less-than-proportional scaling of drawdown with leverage should not be over-interpreted as a structural guarantee; it reflects the specific volatility dynamics observed in the 2019–2025 sample and may not generalize to other periods or market conditions. The champion specification uses $L = 5$, which produces a 10.6% maximum drawdown that appears operationally robust within the tested range.

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. Trading is 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 5× leverage: the base contract count equals the pre-OOS median conditional volatility divided by the 40-day MA of lagged conditional volatility, multiplied by the leverage factor, 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× ATR for NQ and ES, 2.5× 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 bars and a cooldown period of 5 bars after exit.

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×	2.2×	2.5×
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. Leverage: 5$\times$. IVT smoothing: MA40. Vol-score shape: flat-middle $(0.20, 0.40, 0.80, 0.90)$. Regime thresholds: 42nd/58th percentile. Longs only.

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.

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.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64(2):253–280.

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. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.

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.

Marshall, B. R., Cahan, R. H., and Cahan, J. M. (2008). Can commodity futures be profitably traded with quantitative market timing strategies? *Journal of Banking and Finance*, 32(9):1810–1819.

Misirli, E. U. (2024). Aggregate volatility risk and momentum returns. *European Financial Management*. Forthcoming.

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. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

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.