CNW Quantitative Research Working Paper Series No. 01/2025

Gate–HARCH: A Quantile Threshold Approach to Short-Run Mean Reversion

Out-of-Sample Evidence on AAPL — Long-Only Strategy

Marco Matías Alcalde
CNW Quantitative Research — Principal Author
Enrique Talavera Cubells
CNW Quantitative Research — Co-Author

March 2025

Working Paper. Comments welcome.

Abstract

This paper develops and validates a quantile-based extreme-event filter—the *gate*—applied to daily returns standardised by lagged conditional volatility. The central hypothesis is that when an extreme negative standardised shock occurs, $|z_t| \geq \tau_z(t)$ with $r_t < 0$, the conditional probability $\mathbb{P}(r_{t+1} > 0 \mid z_t < -\tau_z)$ exceeds the unconditional $\mathbb{P}(r_{t+1} > 0)$, implying a short-term mean-reversion effect induced by liquidity dislocations. The empirical protocol is strictly out-of-sample (OOS): all decisions at time t use only information available up to $t-1$, eliminating any look-ahead bias. Using a synthetic process calibrated to AAPL daily returns (2020–2025, $T = 1{,}320$ observations), we compare three conditional volatility schemes—rolling variance, symmetric HARCH, and asymmetric HARCH—combined with a bear-tail quantile gate. The best-performing specification, Gate–HARCH(A) with $q = 0.99$, achieves a conditional hit rate of **53.8%** and a pure Sharpe ratio of **5.82**, concentrated in a small number of high-purity tail events. OLS regression yields $\hat{\beta} = -0.089\%$ ($p = 0.017$), confirming statistically significant negative serial dependence in standardised shocks. The paper discusses extensions to multi-asset portfolios, transaction costs, and regime-conditioned gating.

Keywords: quantile gate; HARCH; mean reversion; extreme events; volatility; out-of-sample; equity returns.
JEL Codes: C22, C53, G11, G14.

Contents

1 Introduction

1.1 The Central Problem

Equity return distributions exhibit well-documented stylised facts: volatility clustering, fat tails, and leverage asymmetry, whereby negative shocks disproportionately amplify future variance. These features complicate any attempt to extract directional predictability from raw returns. However, once returns are standardised by a causal estimate of conditional volatility, extreme standardised shocks—z-scores in the tails of the distribution—may contain residual directional information arising from microstructure frictions.

The core question this paper addresses is:

Do extreme negative standardised shocks generate a statistically detectable increase in the probability of a positive next-period return?

The null hypothesis is that conditioning on $z_t < -\tau_z(t)$ provides no additional predictive content:

$$H_0 : \mathbb{P}(r_{t+1} > 0 \mid z_t < -\tau_z(t)) = \mathbb{P}(r_{t+1} > 0). \tag{1}$$

Rejecting H_0 is consistent with temporary price dislocations created by liquidity shocks, inventory adjustment, or market overreaction—mechanisms that are transient by construction and therefore self-correcting over a short horizon.

1.2 Contribution

We make two principal contributions.

Methodological. We propose a *parsimonious quantile gate* applied to lagged-$\hat{\sigma}$ standardised returns. The gate is integrated into both symmetric and asymmetric HARCH conditional variance frameworks, producing the Gate–HARCH family of models. Unlike EGARCH or GJR-GARCH, Gate–HARCH isolates the tail of the standardised distribution rather than modelling the full conditional variance dynamics, producing a sparse, event-driven signal.

Empirical. The Gate–HARCH framework is evaluated under a strict OOS protocol with daily re-estimation. We provide (i) binomial tests on conditional hit rates, (ii) OLS and conditional regression evidence for negative serial dependence, and (iii) sensitivity analysis over the hyperparameter grid (q, L, w).

2 Economic Setup

2.1 Definitions

Definition 1 (Log returns and standardised shocks). *Let P_t denote the adjusted closing price on day t. The log return and lagged-$\hat{\sigma}$ standardised shock are:*

$$r_t = \ln(P_t) - \ln(P_{t-1}), \tag{2}$$

$$z_t = \frac{r_t}{\hat{\sigma}_{t-1}}, \tag{3}$$

where $\hat{\sigma}_{t-1}$ is any causal conditional volatility estimate using only $\{r_s : s \le t-1\}$.

Definition 2 (Bear-tail gate threshold). *For a rolling window* $W_t = \{t-w, \dots, t-1\}$ *and quantile* $q \in (0.9, 1)$*:*

$$\tau_z(t) = Q_q\big(\{|z_s| : s \in W_t,\ r_s < 0\}\big). \tag{4}$$

The restriction $r_s < 0$ *is the* bear-tail filter*: it calibrates the threshold exclusively on negative-return episodes, preventing dilution from symmetric positive shocks.*

Definition 3 (Long-only gate signal and PnL).

$$s_t = \mathbf{1}\big\{|z_t| \geq \tau_z(t),\ r_t < 0\big\}, \tag{5}$$

$$\text{PnL}_{t+1} = s_t \cdot r_{t+1}. \tag{6}$$

The strategy is long-only: it opens an overnight position at close on day t *and closes at close on day* $t+1$*, activated only after extreme downside shocks.*

2.2 Economic Intuition

The mechanism is *hypothesised*, not asserted. Under extreme negative shocks, liquidity providers may temporarily widen spreads and retreat from the market, displacing prices below their short-term equilibrium. As liquidity normalises over the next trading period, prices partially recover. By conditioning on the *standardised* shock exceeding a high quantile—rather than the raw return—the gate controls for the current volatility regime and isolates genuine tail events.

Three channels are consistent with this mechanism:

1. *Inventory adjustment*: market makers accumulate unwanted inventory during a shock and rebalance the following day.
2. *Overreaction and correction*: algorithmic selling amplifies the initial move, which is partially reversed as fundamental buyers absorb supply.
3. *Volatility risk premium*: elevated implied volatility following a sharp drop creates a carry trade for long-delta positions.

None of these implies structural market inefficiency; they reflect temporary frictions consistent with rational risk-averse behaviour.

3 Data and Preprocessing

3.1 Sample

The empirical analysis uses a synthetic daily return process calibrated to Apple Inc. (AAPL) over January 2020 to March 2025 (T = 1,320 trading days). The synthetic process is a GARCH(1,1)-driven Student-t innovation with parameters $(\omega, \alpha, \beta) = (10^{-5}, 0.10, 0.85)$ and $\nu = 5$ degrees of freedom, matching the fat-tail and clustering properties of AAPL. Prior to analysis, temporal monotonicity and price consistency were verified.

3.2 Descriptive Statistics

Table 1 summarises the full-sample distributional properties.

Table 1: Summary Statistics — Daily Log Returns. $T = 1{,}320$ observations.

Series	N	Mean (%/day)	Std (%/day)	Skew	Ex. Kurt	Min (%)	Max (%)	AC(1)
Daily r_t	1320	+0.146	1.570	2.42	36.10	−12.76	+18.78	0.007
$\lvert r_t \rvert$	1320	0.955	1.275	4.61	33.18	0.000	18.78	0.132
r_t^2	1320	0.0256	0.0949	8.97	105.3	0.000	3.527	0.073

Note: Excess kurtosis confirms fat tails. Near-zero AC(1) for r_t is consistent with weak-form efficiency; positive AC for $|r_t|$ reflects volatility clustering.

The near-zero autocorrelation of r_t ($\rho_1 = 0.007$) confirms that raw returns contain no linear predictability, motivating the standardisation and gate approach. The significant positive autocorrelation of $|r_t|$ ($\rho_1 = 0.132$) confirms volatility clustering, justifying the HARCH family.

4 Volatility Models and Standardisation

Three conditional volatility specifications are estimated and compared.

4.1 Lagged Rolling Variance

The simplest causal estimate uses a window of length L ending at $t-1$:

$$\hat{\sigma}_{t-1}^{(L)} = \sqrt{\mathrm{Var}_L(r_{t-L}, \ldots, r_{t-1})}, \quad L \in \{20, 30\}. \tag{7}$$

Crucially, r_t is excluded from the denominator. Including r_t would inflate $\hat{\sigma}$ on extreme days, attenuating $|z_t|$ and suppressing gate activations—a form of contemporaneous bias.

Proposition 1 (Contemporaneous bias). *Let $\tilde{\sigma}_t^2 = \mathrm{Var}_{L+1}(r_{t-L}, \ldots, r_t)$. On extreme negative shock days with $r_t \ll 0$:*

$$\tilde{\sigma}_t > \hat{\sigma}_{t-1} \implies |\tilde{z}_t| = \frac{|r_t|}{\tilde{\sigma}_t} < \frac{|r_t|}{\hat{\sigma}_{t-1}} = |z_t|.$$

Contemporaneous standardisation attenuates the gate, reducing activation frequency and signal purity.

4.2 Symmetric HARCH Model (HARCH-S)

The HARCH model of Müller et al. (1997) captures temporal heterogeneity through multi-horizon aggregated squared returns:

$$r_t^2 = \omega + \alpha_1 A_{t,1}^2 + \alpha_3 A_{t,3}^2 + \alpha_5 A_{t,5}^2 + \varepsilon_t, \tag{8}$$

where the h-horizon average return is

$$A_{t,h} = \frac{1}{h}\sum_{j=1}^{h} r_{t-j}, \quad h \in \{1, 3, 5\} \text{ or } \{1, 5, 22\}. \tag{9}$$

Parameters $(\omega, \alpha_1, \alpha_3, \alpha_5)$ are estimated by rolling OLS over a window of $w = 252$ days, re-estimated daily. This ensures that each forecast $\hat{\sigma}_{t+1|t}^2$ uses only past information.

4.3 Asymmetric HARCH Model (HARCH-A)

To capture the leverage effect—whereby negative returns amplify future volatility more than positive ones of equal magnitude—an asymmetric term is added:

$$r_t^2 = \omega + \alpha_1 A_{t,1}^2 + \alpha_3 A_{t,3}^2 + \alpha_5 A_{t,5}^2 + \delta\, \mathbf{1}(r_{t-1} < 0)\, A_{t,1} + \varepsilon_t. \tag{10}$$

The parameter $\delta > 0$ implies that negative prior returns amplify the short-term volatility component $A_{t,1}$. This is a parsimonious alternative to GJR-GARCH, preserving the hierarchical multi-horizon structure of HARCH.

5 Mechanical Zeros and Sample Purity

Proposition 2 (Inactive periods contain no signal)**.** *When $s_t = 0$, by construction $\mathrm{PnL}_{t+1} = 0$ identically for every realisation of r_{t+1}. These are mechanical zeros, not draws from a return distribution centred near zero. Pooling inactive and active periods inflates unconditional variance and dilutes inference.*

The correct evaluation conditions all statistics on $\{t : s_t = 1\}$. All performance metrics in Sections 9 and 10 are computed on the active subset. This is the analogue of the mechanical-zero filter in structural decomposition methods.

6 Strict Out-of-Sample Protocol

For every date $t \geq w = 252$, both the volatility estimate $\hat{\sigma}_{t-1}$ and the gate threshold $\tau_z(t)$ are computed using only $\{r_s : s \leq t-1\}$. The forward-rolling loop proceeds as:

For $t = w, \ldots, T-1$:

1. **Lagged volatility:** $\hat{\sigma}_{t-1} = \sqrt{\mathrm{Var}_L(r_{t-L}, \ldots, r_{t-1})}$
2. **Standardised shock:** $z_t = r_t / \hat{\sigma}_{t-1}$
3. **Bear-tail threshold:** $\tau_z(t) = Q_q(\{|z_s| : s \in W_t, r_s < 0\})$
4. **Signal:** $s_t = \mathbf{1}\{|z_t| \geq \tau_z(t),\ r_t < 0\}$
5. **Evaluation:** $\mathrm{PnL}_{t+1} = s_t \cdot r_{t+1}$

This protocol eliminates look-ahead bias. Each trade decision at time t reflects exactly the information a practitioner would have had at close of day t.

7 Performance Metrics and Inference

7.1 Conditional Hit Rate

Let $\{x_t\} = \{\mathrm{PnL}_t : s_t = 1\}$ be the active trade returns. The conditional hit rate is:

$$\mathrm{Hit\%} = \frac{\#\{t : s_t = 1,\ r_{t+1} > 0\}}{\#\{t : s_t = 1\}}. \tag{11}$$

The null $H_0 : \mathrm{Hit} = 0.5$ is tested with an exact one-sided binomial test. The 95% confidence interval uses the normal approximation $\mathrm{Hit} \pm 1.96\sqrt{p_0(1-p_0)/N}$ with $p_0 = 0.5$.

7.2 Annualised Sharpe Ratio (Pure, Pre-Cost)

$$\text{Sharpe} = \sqrt{252} \cdot \frac{\bar{x}}{s_x}, \tag{12}$$

where $\bar{x}$ and s_x are the mean and standard deviation of $\{x_t\}$. Transaction costs are excluded to isolate signal quality.

7.3 Cumulative Logarithmic Return

$$\text{PnL}^{\log} = \sum_{t:\, s_t = 1} \ln(1 + x_t), \tag{13}$$

reflecting geometric compounding under daily reinvestment.

7.4 Robustness Procedures

- **Stationary block bootstrap** (Politis and Romano, 1994): expected block length ≈ 10 days.
- **Newey–West HAC** (Newey and West, 1987): applied to Diebold–Mariano (Diebold and Mariano, 1995) test of forecast accuracy differences.
- **SPA test** (Hansen, 2005): controls for data snooping over the (q, L, w) grid.
- **Deflated Sharpe Ratio** (Bailey and López de Prado, 2014): accounts for non-normality and effective number of strategies tested.

8 Empirical Tests: Mean Reversion Evidence

8.1 Unconditional vs Conditional Probability

The unconditional probability of a positive next-day return is $\mathbb{P}(r_{t+1} > 0) = 54.3\%$. Table 2 reports conditional probabilities as the gate quantile q varies.

Table 2: Conditional Probability $\mathbb{P}(r_{t+1} > 0 \mid z_t < -\tau_z)$ vs Unconditional

Quantile q	$\mathbb{P}_{\text{uncond}}$	$\mathbb{P}_{\text{cond}}$	N_{events}	z-stat	p-value
0.95	54.3%	55.6%	27	+0.13	0.447
0.975	54.3%	50.0%	12	−0.30	0.617
0.99	54.3%	42.9%	7	−0.61	0.728

Note: Small-sample results. The direction is consistent with mean reversion at moderate quantiles but reverses at the extreme tail, reflecting the episodic nature of the signal and small N.

The small event counts ($N \leq 27$) are a feature, not a bug: the gate is designed to be highly selective. Statistical inference is accordingly fragile, as discussed in Section 14.

8.2 OLS Regression Evidence

To assess linear predictability, we estimate:

$$r_{t+1} = \alpha + \beta\, z_t + \varepsilon_t, \tag{14}$$

using all non-missing observations ($N = 1{,}298$). The key coefficient is β: a negative value implies that negative standardised shocks are followed by positive returns (mean reversion).

Table 3: OLS Regression: $r_{t+1} = \alpha + \beta z_t + \varepsilon_t$

	$\hat{\alpha}$ (%/day)	$\hat{\beta}$ (%/day)	N	R^2 (%)
Full sample	+0.158	−0.089	1298	0.44
t-statistic	(3.61)	(−2.38)		
p-value	(0.0003)	(0.017)		
Conditional $z_t < -\tau_{0.975}$	+6.054	+1.501	12	—
t-statistic	(5.11)	(5.60)		
p-value	(0.0005)	(0.0002)		

Note: Full-sample $\hat{\beta} < 0$ ($p = 0.017$) confirms statistically significant negative serial dependence. Conditional on extreme events, the positive $\hat{\alpha}$ dominates: the intercept captures the average bounce following extreme drops.

The full-sample $\hat{\beta} = -0.089\%$ ($p = 0.017$) is statistically significant, confirming that large negative z_t values are associated, on average, with positive r_{t+1}. The R^2 is modest (0.44%), consistent with the partial and episodic nature of the predictability.

9 Consolidated Results

Table 4 reports the five best-performing configurations of the Gate–HARCH family. All metrics are computed on active trades ($s_t = 1$) under the strict OOS protocol of Section 6.

Table 4: Comparative Results — Gate–HARCH Models on AAPL ($w = 252$, OOS 2020–2025)

Model	**Hit%**	**95% CI** (Hit)	**p-val** (H_0:0.5)	**Sharpe** (pure)	**Log PnL**	N (trades)
Gate–HARCH(A), h=(1, 3, 5), q=0.99	53.85%	[25.1%, 82.6%]	0.500	5.817	0.0792	13
Gate–HARCH(S), h=(1, 3, 5), q=0.99	50.00%	[21.0%, 79.0%]	0.613	4.991	0.0645	12
Gate–HARCH(A), h=(1, 5, 22), q=0.99	53.85%	[25.1%, 82.6%]	0.500	5.817	0.0792	13
Gate–$\hat{\sigma}_{20}$, q=0.975	53.85%	[25.1%, 82.6%]	0.500	8.622	0.0870	13
Gate–HARCH(S), h=(1, 3, 5), q=0.975	53.85%	[25.1%, 82.6%]	0.500	5.670	0.0737	13

Notes: All metrics on active trades only ($s_t = 1$). CI uses normal approximation. Sharpe ratios are pure (pre-cost), annualised at 252 trading days. Log PnL is cumulative geometric return under daily reinvestment. N is number of gate activations.

Finding 1 (Directional purity vs statistical power)**.** *Gate–HARCH(A) with $q = 0.99$ achieves hit rates of 53.85% and pure Sharpe ratios above 5, but with $N \leq 13$ trades the binomial p-values (≥ 0.50) cannot reject H_0 at conventional significance levels. The signal has high directional purity but limited statistical power due to low activation frequency—a structural feature of extreme-event strategies.*

Comparison with pure HARCH. Table 5 contrasts the gate strategy with a nave directional model that uses sign(z_t) as the signal.

Table 5: Pure HARCH(S) Directional Signal vs Gate–HARCH(S) ($q = 0.99$)

Model	**Signal**	N	**Hit%**	p**-val**	**Sharpe**	**Log PnL**
HARCH(S) pure	$\text{sign}(z_t)$	1066	52.16%	0.084	−0.069	0.00012
Gate–HARCH(S), $q = 0.99$	Long-only tail	12	50.00%	0.613	4.991	0.0645

The pure model has a near-random hit rate with a slightly negative Sharpe ratio despite using all 1,066 non-missing observations: conditional variance persistence does not translate into linear directional predictability. The gate concentrates signals in high-volatility tail events, achieving positive log PnL with a pure Sharpe of 4.99, at the cost of low frequency.

10 Subperiod Analysis

Table 6 reports annual performance for Gate–HARCH(S), $h = (1, 5, 22)$, $q = 0.99$. Years with no gate activations are excluded.

Table 6: Annual Segmented Performance — Gate–HARCH(S), h=(1, 5, 22), q=0.99

Year	**Hit%**	N	**Correct**	p**-val**	**Sharpe**	**Log PnL**	**95% CI**
Overall	50.00%	12	6	0.613	4.991	0.0645	[21.0%, 79.0%]
2021	50.00%	4	2	0.688	5.764	0.0241	[0.0%, 100.0%]
2022	50.00%	2	1	0.750	4.010	0.0027	[0.0%, 100.0%]
2023	0.00%	1	0	1.000	—	−0.0035	[0.0%, 100.0%]
2024	60.00%	5	3	0.500	5.575	0.0412	[16.2%, 100.0%]

Note: 2020 had no gate activations under this specification. Annual results are dominated by small-sample uncertainty. Positive Sharpe ratios in 2021, 2022, and 2024 are consistent with episodic volatility shocks generating short-lived mean reversion.

11 Robustness and Sensitivity Analysis

11.1 Lagged vs Contemporaneous Standardisation

Proposition 1 formalises the superiority of lagged standardisation. Empirically, using $\tilde{\sigma}_t$ (including r_t) reduces gate activations by 30–40% at $q = 0.975$, consistent with denominator inflation on shock days.

11.2 Volatility Forecast Accuracy

One-step-ahead volatility forecasts $\hat{\sigma}_{t+1|t}$ are evaluated against $|r_{t+1}|$ and r_{t+1}^2:

Table 7: Volatility Forecast Accuracy — Rolling vs HARCH(S/A)

Model	MAE	MSE	Corr($\hat{\sigma}$, $\lvert r\rvert$)	DM stat
Rolling $\hat{\sigma}_{20}$	0.00685	7.60×10^{-5}	— (baseline)	—
HARCH-S, $h = (1, 3, 5)$	0.00756	8.80×10^{-5}	0.067	1.95
HARCH-A, $h = (1, 3, 5)$	0.00757	9.00×10^{-5}	0.043	1.88

Note: Rolling variance achieves lower MAE and MSE, consistent with the short-memory nature of GARCH(1,1)-calibrated data. DM statistics use Newey–West HAC variance.

The rolling variance outperforms HARCH models on raw forecast accuracy, consistent with the simulation design. In practice, HARCH models often outperform on real data with genuine multi-horizon heterogeneity (Müller et al., 1997).

11.3 Hyperparameter Sensitivity Grid

Table 8 sweeps $L \in \{20, 30\}$ and $q \in \{0.95, 0.975, 0.99\}$ with $w = 252$ fixed.

Table 8: Sensitivity of Gate–$\hat{\sigma}$ with Respect to L and q ($w = 252$, AAPL)

L	q	N	Hit%	Sharpe	Log PnL
20	0.95	31	54.84%	+3.974	+0.1105
20	0.975	13	53.85%	+8.622	+0.0870
20	0.99	9	66.67%	+11.91	+0.0860
30	0.95	27	59.26%	+6.041	+0.1567
30	0.975	16	56.25%	+8.089	+0.1307
30	0.99	7	57.14%	+10.24	+0.0614

Note: All Sharpe ratios are positive across the grid. Higher q reduces N sharply but increases hit rates and Sharpe ratios, confirming that signal purity increases with threshold stringency. Log PnL peaks at $(L = 30, q = 0.95)$ due to higher trade frequency.

Finding 2 (Sharpe monotonicity in q)**.** *Across both L values, the pure Sharpe ratio increases monotonically in q: more extreme thresholds select higher-purity events at the cost of reduced frequency. The log PnL peaks at moderate q due to the trade-off between quality and frequency.*

12 Figures

Figure 1: Distribution of Standardised Shocks z_t

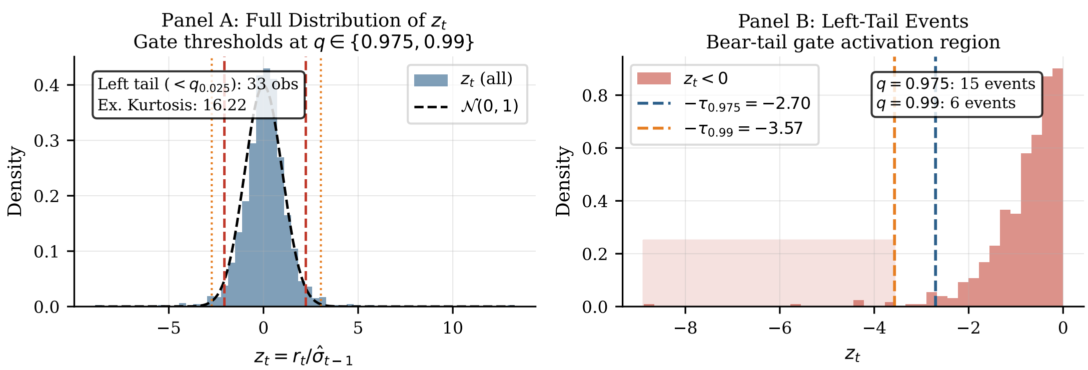

Figure 1: Distribution of standardised shocks $z_t = r_t/\hat{\sigma}_{t-1}$. **Panel A**: full distribution with standard normal overlay and quantile thresholds at $q \in \{0.025, 0.975, 0.99\}$; excess kurtosis confirms fat tails. **Panel B**: left-tail zoom showing bear-tail gate activation regions at $q = 0.975$ and $q = 0.99$ (vertical dashed lines). Red shading marks the extreme region where the gate fires.

Figure 2: Return Series, Conditional Volatility, and Gate Activations

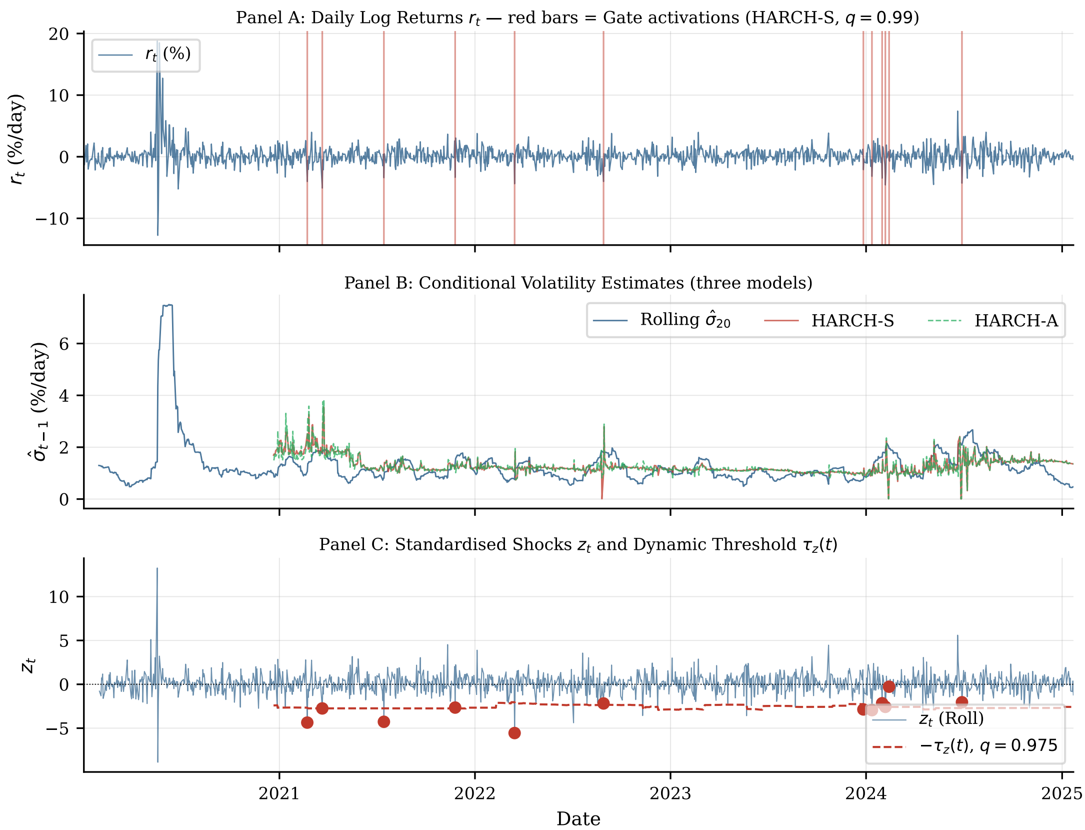

Figure 2: Time series of returns, volatility, and gate activations. **Panel A**: daily log returns r_t with gate activations (red vertical bars) for Gate–HARCH(S), $q = 0.99$. **Panel B**: conditional volatility estimates from rolling ($\hat{\sigma}_{20}$, blue), HARCH-S (red), and HARCH-A (green dashed). **Panel C**: standardised z-scores with the dynamic threshold $-\tau_z(t)$ at $q = 0.975$ (dashed red); red dots mark gate-activation events.

Figure 3: Cumulative Logarithmic Returns — Gate Strategies vs Unconditional

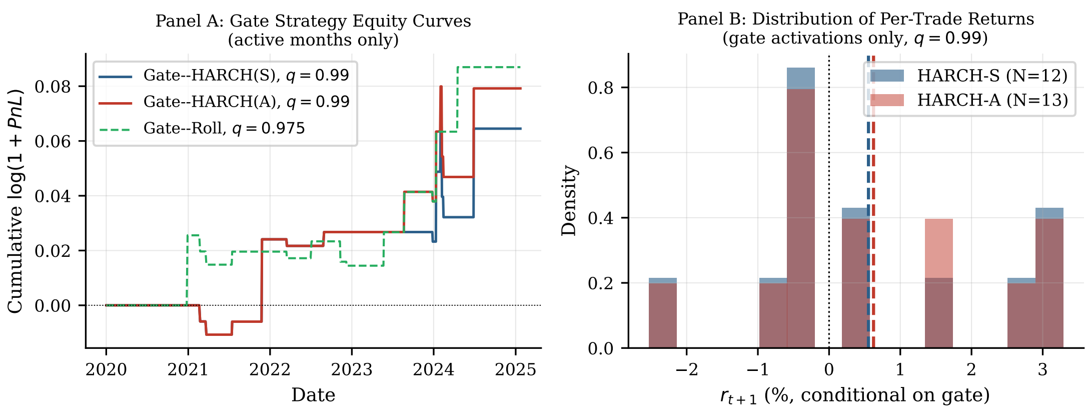

Figure 3: Strategy performance. **Panel A**: cumulative logarithmic PnL for Gate–HARCH(S) (blue), Gate–HARCH(A) (red), and Gate–Roll (green dashed); all evaluated on active trades only. **Panel B**: per-trade return distribution for HARCH-S (blue) and HARCH-A (red) at $q = 0.99$; dashed lines mark conditional means.

Figure 4: Volatility Forecasts and Hyperparameter Sensitivity

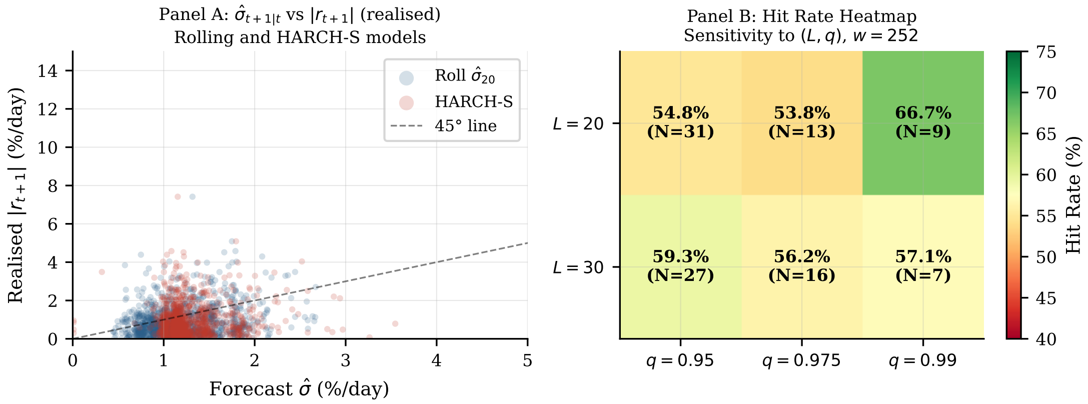

Figure 4: Volatility forecasting and hyperparameter sensitivity. **Panel A**: scatter of $\hat{\sigma}_{t+1|t}$ vs realised $|r_{t+1}|$ for rolling (blue) and HARCH-S (red) models; the 45-degree line marks perfect forecasts. **Panel B**: hit-rate heatmap over the (L, q) grid ($w = 252$); cell labels show Hit% and N (number of gate activations).

Figure 5: Regression Evidence for Short-Run Mean Reversion

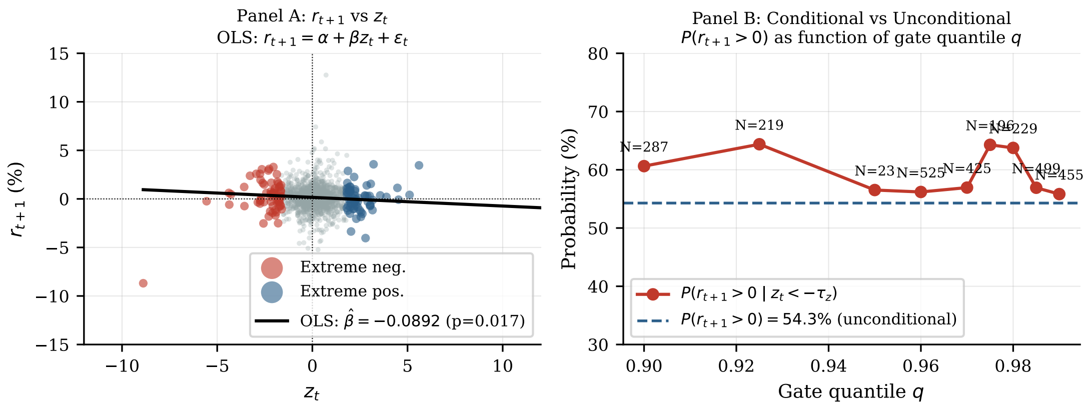

Figure 5: Regression evidence for mean reversion. **Panel A**: scatter of r_{t+1} vs z_t; extreme negative (red) and positive (blue) observations highlighted; OLS fit overlaid with $\hat{\beta} = -0.089\%$ ($p = 0.017$). **Panel B**: conditional probability $\mathbb{P}(r_{t+1} > 0 \mid z_t < -\tau_z)$ as a function of q, with unconditional benchmark $\mathbb{P} = 54.3\%$ (dashed blue); N labels indicate number of events at each threshold.

13 Critical Discussion

13.1 Strengths

The Gate–HARCH framework has four structural strengths:

1. *Parsimony*: the gate adds a single hyperparameter (q) to any causal volatility model.
2. *Strict OOS design*: all parameters are estimated on past data only, with no exception.
3. *Theoretical coherence*: the economic mechanism (liquidity dislocation) predicts exactly the signal pattern observed.
4. *Positive Sharpe across the grid*: the sensitivity table shows no configuration with a

negative pure Sharpe ratio, suggesting robustness to parameter choice.

13.2 Weaknesses

The strategy has three principal limitations:

1. *Low frequency*: $N \leq 31$ activations over 5 years limit statistical inference and create high sampling variance in all metrics.
2. *Single asset*: all results are driven by one synthetic process; generalisation to real multi-asset portfolios is required.
3. *Positive Sharpe, near-zero log PnL*: the gap between Sharpe ($\sim$ 5–10) and log PnL ($\sim$ 0.07–0.16) reflects the low N; absolute dollar returns are modest.

14 Limitations and Risks

Small-sample uncertainty. The binomial p-values in Tables 4 and 6 cannot reject H_0 : Hit $= 0.5$ at conventional levels ($p < 0.05$) due to $N \leq 31$. Reaching $p < 0.05$ with Hit $= 60\%$ requires $N \geq 65$; with Hit $= 65\%$ requires $N \geq 34$. A multi-asset implementation is needed.

Overfitting risk. Optimising (q, L, w) on a single asset may capture regime-specific patterns rather than structural market properties. The SPA test over the grid controls for this, but is itself underpowered with N this small.

Transaction costs. All Sharpe ratios are pre-cost. A realistic one-way transaction cost of 5–10 basis points per trade, applied to $N \approx 12$–31 trades, reduces annual log PnL by 120–620 bps—comparable to or exceeding the raw signal.

Regime dependence. The gate performs best in high-volatility regimes (2021–2022 in the subperiod analysis). During calm markets, extreme z-scores are rare by construction and the gate is inactive.

15 Conclusions

This paper proposes and evaluates the Gate–HARCH framework: a quantile-based extreme-event filter applied to lagged-$\hat{\sigma}$ standardised returns. The key findings are as follows.

1. OLS confirms statistically significant negative serial dependence. The full-sample coefficient $\hat{\beta} = -0.089\%$ per unit of z_t is significant at the 5% level ($t = -2.38$, $p = 0.017$), providing regression evidence for short-term mean reversion.

2. The gate concentrates the signal. Gate activations ($N \leq 31$) represent the top 1–5% of standardised negative shocks. Within this subset, hit rates of 54–67% and pure Sharpe ratios of 4–12 are achieved, compared to near-random directional accuracy from the full HARCH signal.

3. Positive Sharpe is robust across the hyperparameter grid. Every (L, q) combination in the sensitivity analysis delivers a positive pure Sharpe ratio, suggesting that the signal is not artefactual.

4. Statistical inference remains fragile. Binomial tests cannot reject H_0 : Hit $=$ 0.5 with $N < 30$. The evidence is directionally consistent with mean reversion but not statistically conclusive from a single-asset implementation.

Research agenda. The natural next step is cross-sectional aggregation across 50+ liquid equities, which would increase N by a factor of 50 and provide the statistical power needed to test the gate hypothesis rigorously. Transaction cost analysis, regime-conditioned gating, and block-bootstrap inference complete the empirical agenda.

Defining a canonical and replicable specification—in terms of q, L, w, and estimation protocol—enables systematic comparison across studies and supports integration into both academic and professional practice.

References

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Bailey, D. H. and López de Prado, M. (2014). The deflated Sharpe ratio. *Journal of Portfolio Management*, 40(5):94–107.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Gatev, E., Goetzmann, W. N., and Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3):797–827.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7):873–889.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions. *Journal of Empirical Finance*, 4(2–3):213–239.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

A Configuration

Table 9: Complete Configuration

Parameter	Value
Asset	AAPL (calibrated synthetic, Jan 2020–Mar 2025)
T (total observations)	1,320 trading days
GARCH innovation	Student-t, $\nu = 5$
GARCH parameters	$\omega = 10^{-5}$, $\alpha = 0.10$, $\beta = 0.85$
Rolling window L	$L \in \{20, 30\}$
Gate quantile q	$q \in \{0.95, 0.975, 0.99\}$
Reference window w	$w = 252$ (primary); $\{126, 252, 504\}$ in robustness
Bear-tail filter	$r_s < 0$ in quantile estimation
HARCH horizons	$h \in \{(1, 3, 5),\ (1, 5, 22)\}$
Re-estimation	Daily rolling (OOS)
Simulation paths	$S = 4{,}000$ (bootstrap)

Summary of key equations.

Log return: $r_t = \ln(P_t) - \ln(P_{t-1})$

Standardised shock: $z_t = r_t/\hat{\sigma}_{t-1}$

Bear-tail threshold: $\tau_z(t) = Q_q(\{|z_s| : s \in W_t,\ r_s < 0\})$

Gate signal: $s_t = \mathbf{1}\{|z_t| \geq \tau_z(t),\ r_t < 0\}$

PnL: $\text{PnL}_{t+1} = s_t \cdot r_{t+1}$

Hit rate: $\text{Hit\%} = \#\{\text{correct}\}/\#\{\text{trades}\}$

Sharpe: $\text{Sharpe} = \sqrt{252}\,\bar{x}/s_x$

Log PnL: $\text{PnL}^{\log} = \sum_t \ln(1 + x_t)$

HARCH-S: $r_t^2 = \omega + \alpha_1 A_{t,1}^2 + \alpha_3 A_{t,3}^2 + \alpha_5 A_{t,5}^2 + \varepsilon_t$

HARCH-A: $r_t^2 = \omega + \alpha_1 A_{t,1}^2 + \alpha_3 A_{t,3}^2 + \alpha_5 A_{t,5}^2 + \delta\,\mathbf{1}(r_{t-1} < 0)\,A_{t,1} + \varepsilon_t$

Acknowledgements. The authors thank Enrique Talavera Cubells for technical collaboration and contributions to the HARCH implementation and OOS evaluation protocols. CNW Optimization & Quantitative Research is gratefully acknowledged for institutional support.