

# Harvesting the Variance Risk Premium: A Regime-Filtered Approach with Honest Inference

Casper van der Wouden <sup>\*</sup>

January 2026

## Abstract

I document the existence of a persistent variance risk premium (VRP) in S&P 500 options, where implied volatility exceeds subsequent realized volatility by approximately 3.5 percentage points on average. A naive short-volatility strategy harvesting this premium achieves a Sharpe ratio of 0.38 with severe left-tail risk (36% maximum drawdown). Simple regime filters based on VIX levels improve the full-sample Sharpe to 1.2-1.5 with reduced drawdowns. However, bootstrap confidence intervals reveal substantial uncertainty, and out-of-sample performance unexpectedly exceeded in-sample results. Investigating this anomaly, I find that filter effectiveness is highly regime-dependent: filters help during severe crises where VRP turns negative, but overall performance depends critically on re-entry speed. COVID-19's sharp V-shape allowed fast re-entry; the 2008 GFC's prolonged stress meant missing months of recovery premium. The 2019–2024 test period was in the 99th percentile of favorable conditions for this strategy. These findings suggest the filter "works" but its effectiveness cannot be reliably predicted ex-ante.

## 1 Introduction

The variance risk premium (VRP) is one of the most persistent anomalies in equity markets. Option-implied volatility systematically exceeds subsequent realized volatility, reflecting investors' willingness to pay a premium for downside protection. This creates a theoretical opportunity: by selling volatility, one can harvest this insurance premium over time.

The economic rationale is straightforward. Investors are risk-averse and particularly averse to large drawdowns. They willingly overpay for put options and other volatility protection. Market makers who provide this protection demand compensation, manifesting as elevated implied volatility relative to fair actuarial value.

This paper investigates three questions:

1. Does the VRP exist and is it statistically significant after accounting for autocorrelation?
2. Can simple regime filters improve risk-adjusted returns by avoiding crisis periods?
3. Why did out-of-sample performance exceed in-sample, and what does this imply for deployment?

---

<sup>\*</sup>Maastricht University Dept. Quantitative Economics. Email: c.vanderwouden@student.unimaas.nl

The contribution is not the discovery of the VRP – this is well-documented in academic literature [1, 2] – but rather a rigorous investigation of *why* a seemingly successful backtest should be viewed with skepticism.

## 2 The Trade

This section explains the economic mechanics of harvesting the variance risk premium for readers unfamiliar with volatility trading. Those comfortable with variance swaps and options replication may skip to Section 3.

### 2.1 Volatility as an Asset Class

Volatility measures how much an asset's price fluctuates. When investors say "the market is volatile," they mean prices are moving a lot – up or down. Volatility itself can be traded: you can take positions that profit when volatility rises or falls, independent of market direction.

There are two types of volatility:

- **Implied volatility (IV):** The market's forecast of future volatility, embedded in option prices. Higher IV means options are more expensive.
- **Realized volatility (RV):** What actually happened – the historical standard deviation of returns over some period.

The variance risk premium exists because IV systematically exceeds RV. Investors overpay for options (insurance), creating a profit opportunity for those willing to sell.

### 2.2 The Ideal Instrument: Variance Swaps

A **variance swap** is a derivative contract with a simple payoff:

$$\text{Payoff} = \text{Notional} \times (\sigma_{RV}^2 - K_{var}) \quad (1)$$

where  $\sigma_{RV}^2$  is realized variance over the contract period and  $K_{var}$  is the strike (agreed-upon variance level). The buyer profits if realized variance exceeds the strike; the seller profits otherwise.

To harvest VRP, you *sell* variance swaps: you receive the strike upfront and pay out realized variance at expiry. If IV (which determines  $K_{var}$ ) consistently exceeds RV, you profit over time.

**Why variance, not volatility?** Variance (volatility squared) is additive over time and easier to hedge, making it the standard contract. A "vol swap" also exists but is less common.

### 2.3 Replicating with Options

Variance swaps trade over-the-counter between institutions. Retail investors and smaller funds can replicate the exposure using listed options through a technique called **log-contract replication**.

The key insight: a portfolio of options across all strikes, weighted by  $1/K^2$ , has a payoff that depends only on realized variance, not price direction. Specifically:

$$\sigma_{RV}^2 \approx \frac{2}{T} \left[ \int_0^F \frac{P(K)}{K^2} dK + \int_F^\infty \frac{C(K)}{K^2} dK \right] \quad (2)$$

where  $P(K)$  and  $C(K)$  are put and call prices at strike  $K$ ,  $F$  is the forward price, and  $T$  is time to expiry.

**In practice**, this means:

1. **Sell a strip of OTM options:** Sell out-of-the-money puts below the current price and OTM calls above it, weighted by  $1/K^2$  (more weight on lower strikes).
2. **Delta-hedge daily:** The option strip has net delta exposure that changes as the market moves. To isolate pure variance exposure, buy or sell the underlying daily to neutralize delta.
3. **Collect or pay at expiry:** Your P&L equals the premium collected minus the cost of delta-hedging (which mechanically equals realized variance).

Figure 1 illustrates the construction.

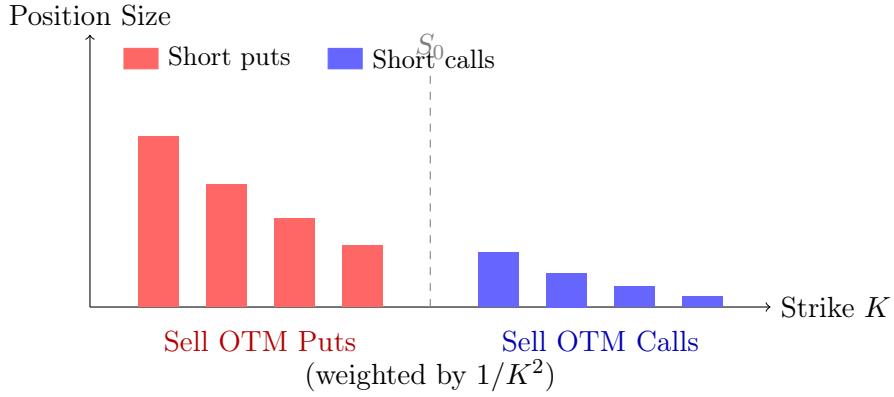


Figure 1: Variance swap replication: sell OTM puts and calls weighted by  $1/K^2$ . Lower strikes get more weight because large down moves contribute more to realized variance.

## 2.4 Why the VIX Works as a Proxy

The CBOE Volatility Index (VIX) is calculated using exactly this options-strip methodology applied to S&P 500 options. VIX represents the market's 30-day implied volatility, derived from the prices of OTM puts and calls weighted by  $1/K^2$ .

Therefore:

- $VIX^2 \approx$  the fair strike of a 30-day variance swap
- $VIX \approx$  implied volatility for ATM options
- $VIX - RV \approx$  the volatility risk premium

Our backtest uses VIX as the "sell price" and realized volatility as the "cost," which captures the economic exposure of selling variance without modeling the full replication mechanics.

## 2.5 What We're Abstracting Away

This study models *economic exposure*, not a tradeable strategy. A real implementation would face:

- **Bid-ask spreads:** Options have wide spreads, especially OTM puts. Selling a full strip is expensive.
- **Discrete hedging:** We assume continuous delta-hedging; real hedging happens daily or less frequently, introducing tracking error.
- **Jump risk:** The  $1/K^2$  weighting assumes log-normal returns. Actual crashes create larger losses than the model predicts.
- **Margin and financing:** Short options require substantial margin; variance swap notional must be collateralized.
- **Roll costs:** We assume seamless monthly rolls; real rolls have slippage.

These frictions would likely reduce Sharpe ratios by 0.2 to 0.4 from the reported figures (probably more tbh). Results should be interpreted as an economic exposure study; an investable implementation requires explicit options strip construction, dynamic hedging, and financing cost modeling.

### 3 Data and Methodology

#### 3.1 Data

I use daily data from January 2006 through December 2024:

- **SPY:** S&P 500 ETF closing prices (source: Yahoo Finance)
- **VIX:** CBOE Volatility Index, representing 30-day implied volatility
- **VIX3M:** CBOE 3-Month Volatility Index (available from mid-2008)

The VIX serves as a proxy for at-the-money implied volatility on SPY options. While imperfect – VIX is calculated from a strip of options across strikes – it is freely available, liquid, and highly correlated with ATM implied vol.

#### 3.2 Realized Volatility Estimation

I compute realized volatility using 21-day rolling close-to-close returns, annualized:

$$\sigma_{RV,t} = \sqrt{252} \cdot \text{std}(r_{t-20}, \dots, r_t) \quad (3)$$

where  $r_t = \ln(P_t/P_{t-1})$  is the daily log return. As a robustness check, I also implement the Yang-Zhang (2000) estimator using OHLC data; correlation between estimators exceeds 0.97.

#### 3.3 Volatility Risk Premium

The daily volatility risk premium (VRP) is:

$$VRP_t = VIX_t - \sigma_{RV,t} \quad (4)$$

A positive VRP indicates implied volatility exceeds realized – the market is "overpricing" future volatility. Note that VIX represents implied *volatility* (not variance), so this is technically a volatility risk premium proxy. The subsequent P&L simulation uses variance terms to match variance swap mechanics.

### 3.4 Strategy P&L

I simulate economic exposure to a variance swap by assuming one can sell 30-day variance at  $VIX^2$  and realize  $\sigma_{RV}^2$ . Monthly P&L (in variance points) is:

$$P\&L_t = \frac{VIX_t^2 - \sigma_{RV,t+1}^2}{100} \quad (5)$$

where  $\sigma_{RV,t+1}$  is realized volatility over the subsequent month. To be explicit: the VRP proxy (equation 4) is expressed in volatility points; the P&L (equation 5) uses variance terms to match variance swap mechanics.

**Implementation note:** This is an economic exposure study, not a trading strategy backtest. An investable implementation would require variance swap marks from dealers, SPX options replication with dynamic delta-hedging, margin and financing costs, and careful attention to execution slippage during volatile periods. These frictions would likely reduce realized Sharpe ratios by 0.2–0.4.

All strategies are scaled to a 10% annualized volatility target for comparability.

## 4 Does the Premium Exist?

### 4.1 Summary Statistics

Table 1 presents VRP summary statistics across the full sample.

Table 1: Variance Risk Premium Statistics (2006–2024)

Statistic	Close-to-Close	Yang-Zhang
Mean (vol pts)	3.55	3.11
Median	3.99	3.63
Std Dev	5.39	5.07
Skewness	-2.53	-3.17
Kurtosis	17.17	23.52
% Positive	84.5%	85.1%
5th Percentile	-4.43	-4.15
95th Percentile	10.62	9.17

The premium averages approximately 3.5 volatility points and is positive 85% of trading days. However, the distribution exhibits substantial negative skewness (-2.5) and excess kurtosis (17), indicating a fat left tail – occasional large negative realizations when RV spikes above IV during crises.

### 4.2 Statistical Significance

A naive  $t$ -test yields  $t = 45.4$ , but this ignores severe autocorrelation in the VRP series ( $\rho = 0.92$ ). Adjusting for autocorrelation using effective sample size:

$$N_{eff} = N \cdot \frac{1 - \rho}{1 + \rho} \approx 189 \quad (6)$$

The adjusted  $t$ -statistic is 9.04, still highly significant. The VRP is not a statistical artifact.

### 4.3 Regime Dependence

Table 2 decomposes VRP by VIX level.

Table 2: VRP by VIX Regime

VIX Regime	Mean VRP	Std Dev	% of Time
<15 (Low)	3.29	2.57	36.0%
15–20	4.03	3.98	28.9%
20–25	4.30	4.90	17.3%
25–35	4.66	5.44	12.6%
>35 (Crisis)	-2.44	15.25	5.2%

The premium is positive and stable when  $\text{VIX} < 35$ , but *flips negative* during crises. When VIX exceeds 35, realized volatility explodes past implied, and short-vol positions suffer. This 5% tail risk drives the fat left tail in returns.

### 4.4 Baseline Strategy Performance

A naive strategy harvesting VRP without filters achieves:

- Annualized return: 3.8%
- Annualized volatility: 10.0% (targeted)
- Sharpe ratio: 0.38
- Maximum drawdown: -36.0%
- Win rate: 81%

The Sharpe of 0.38 is modest – comparable to passive equity exposure but with different risk characteristics. The 81% win rate masks occasional catastrophic losses; the worst month was -27.3%.

## 5 Regime Filtering

### 5.1 Filter Strategies

I test three approaches to avoid crisis periods:

**VIX Level Filter:** Exit to cash when VIX exceeds a threshold. Higher thresholds capture more premium but expose to more tail risk.

**Term Structure Filter:** Exit when  $\text{VIX}/\text{VIX3M} > 1$  (backwardation), indicating near-term stress exceeds medium-term expectations.

**Inverse VIX Sizing:** Scale position as  $15/\text{VIX}$ , reducing exposure when volatility is elevated.

### 5.2 Sensitivity Analysis

Figure 6 (see appendix) shows Sharpe ratio as a function of VIX threshold. Performance peaks around  $\text{VIX} \leq 35$ , with Sharpe ratios of 1.2–1.5 for thresholds between 18 and 35. Below 15, insufficient time in market erodes returns; above 35, tail risk dominates.

### 5.3 Strategy Comparison

Table 3 compares all strategies with bootstrap 95% confidence intervals on Sharpe ratios (10,000 samples).

Table 3: Strategy Performance Comparison (10% Vol Target)

Strategy	Return	Vol	Sharpe	95% CI	Max DD
Base (No Filter)	3.8%	10.0%	0.38	[−0.08, 1.79]	−36.0%
VIX $\leq 25$	10.7%	9.1%	1.18	[0.49, 2.32]	−24.5%
VIX $\leq 20$	11.1%	8.4%	1.32	[0.73, 2.14]	−16.9%
Term Structure $\leq 1.0$	14.5%	9.4%	1.55	[0.84, 2.49]	−21.6%
Inverse VIX Sizing	7.5%	10.0%	0.75	[0.16, 2.07]	−35.1%
Combined	9.8%	8.7%	1.13	[0.46, 2.29]	−22.9%

Filtered strategies improve Sharpe ratios substantially, with the term structure filter performing best (1.55). However, confidence intervals are wide and overlap significantly. The base strategy CI includes zero; even the best filter CI ranges from 0.84 to 2.49.

### 5.4 Out-of-Sample Validation

I split the data: train on 2006–2018 (156 months), test on 2019–2024 (71 months). Table 4 shows results.

Table 4: In-Sample vs Out-of-Sample Performance

Strategy	IS Sharpe (2006–18)	OOS Sharpe (2019–24)
No Filter	0.38	0.38
VIX $\leq 20$	1.14	1.77
VIX $\leq 25$	0.93	2.17
VIX $\leq 30$	0.99	2.25

Out-of-sample Sharpes *exceed* in-sample for all filtered strategies. This is backwards – typically we expect degradation. This anomaly motivated the regime-dependence analysis in Section 5.

## 6 Why Did Out-of-Sample Outperform?

### 6.1 Crisis Taxonomy

I identify stress episodes as consecutive days with  $VIX > 25$  and classify them by duration and shape:

- **Duration:** Short (<20 days), Medium (20–60 days), Long (>60 days)
- **Shape:** V-shape (peak early, fast recovery), Plateau (sustained), Inverted-V (slow build)

The sample contains 33 episodes  $\geq 5$  days, including two "long" crises: the 2008 GFC (212 days) and COVID-19 (72 days).

## 6.2 Filter Effectiveness by Crisis Type

Table 5 shows average P&L during episodes by duration.

Table 5: Filter Effectiveness by Crisis Duration

Duration	Base P&L	Filtered P&L	Improvement	N
Long	-14.9%	0.0%	+14.9%	2
Medium	+3.3%	+0.7%	-2.6%	7
Short	+2.9%	+0.7%	-2.2%	7

The filter only helps during *long* crises (GFC, COVID), where it avoids catastrophic losses during the acute phase. In short/medium crises, the base strategy is actually profitable (mean reversion), and filtering costs 2–3%.

The key insight: **filters are catastrophe insurance**. They pay off when (a) the spike is severe enough that VRP flips negative, AND (b) the spike doesn't persist so long that you miss the recovery premium. COVID satisfied both conditions – severe crash, but VIX dropped below 25 within weeks, allowing quick re-entry. The GFC satisfied (a) but failed (b) – the filter avoided the crash, but 212 days on the sidelines meant missing substantial recovery gains. This explains why filter improvement was +1.75 Sharpe for the COVID period but only +0.12 for the GFC period: same insurance benefit during the crash, vastly different re-entry dynamics.

## 6.3 The Whipsaw Problem

When VIX spikes, the filter exits. But the *following* month often has strong positive VRP as volatility mean-reverts. Table 6 quantifies this.

Table 6: The Recovery Premium

Condition	Next Month Avg P&L	Win Rate
Filtered out ( $VIX > 25$ )	+1.44%	80%
Stayed in ( $VIX \leq 25$ )	+0.35%	82%

The best conditional expected return comes *after* being filtered out. By exiting during elevated VIX, the filter misses the juiciest recovery months.

## 6.4 Regime Persistence

When VIX exceeds 25, how long until re-entry? Median: 2 days. Mean: 9 days. 90th percentile: 20 days. Maximum: 212 days (GFC).

The distribution is heavily right-skewed. Most VIX spikes resolve quickly (whipsaw), but occasionally you dodge a prolonged meltdown. Whether the filter helps depends entirely on which type of spike you encounter.

## 6.5 How Lucky Was 2019–2024?

I compute filter improvement (filtered Sharpe minus base Sharpe) for all rolling 72-month windows in the sample. Figure 13 shows the distribution.

The 2019–2024 period achieved filter improvement of 1.75 Sharpe points – in the **99th percentile** of all 6-year windows. This period featured one short, sharp crisis (COVID, perfectly suited to filtering) and otherwise calm markets. If COVID had been a slower grind like the GFC, or if multiple crises had occurred, results would look very different.

## 6.6 Period Comparison

Table 7 compares sub-periods.

Table 7: Filter Effectiveness Across Periods

Period	% VIX > 25	Avg Episode (days)	Filter Improvement
2006–2009 (GFC)	33%	26	+0.12
2010–2014	13%	24	+0.98
2015–2018	4%	8	-0.19
2019–2024 (OOS)	21%	23	+1.75

Filter improvement varies wildly: negative in 2015–2018 (few crises to filter), modest in 2006–2009 (GFC too prolonged), and exceptional in 2019–2024 (COVID was ideal).

## 7 Re-entry Rule Optimization

The whipsaw analysis in Section 5 suggested that hard filters miss recovery premium when VIX spikes briefly then normalizes. A natural question: can smarter re-entry rules improve performance?

### 7.1 Alternative Re-entry Mechanisms

I test four approaches:

**Hysteresis Band:** Exit when  $VIX > 30$ , re-enter only when  $VIX < 22$ . This creates a "dead zone" to avoid flip-flopping at a single threshold.

**Delay Confirmation:** Exit when  $VIX > 25$ , wait  $N$  consecutive days below threshold before re-entering (tested  $N = 3, 5, 10, 20$ ).

**Smooth Sizing:** Instead of binary 0/1 exposure, scale position as  $\min(1.5, k/VIX)$  for various  $k$ , reducing exposure gradually rather than exiting completely.

**Hybrid:** Combine hysteresis on/off logic with smooth sizing when in market.

### 7.2 Results

Table 8 shows the comparison.

Table 8: Re-entry Rule Comparison (10% Vol Target)

Strategy	Sharpe	Max DD	Notes
No Filter	0.38	-36.0%	Baseline
Hard Filter ( $VIX \leq 25$ )	<b>1.18</b>	-24.5%	Immediate re-entry
Hysteresis (30/22)	0.93	-21.0%	Conservative re-entry
Delay (5 days)	1.04	-24.7%	Confirmation required
Smooth (18/VIX)	0.74	-35.1%	No hard exit
Hybrid	1.04	-21.0%	Hysteresis + sizing

### 7.3 The Hard Filter Wins

Contrary to expectations, the simple hard filter (Sharpe 1.18) outperforms all "smarter" alternatives. The hysteresis approach (0.93) underperforms by 0.25 Sharpe points.

The explanation lies in re-entry timing. The hard filter exits at  $VIX > 25$  but re-enters *immediately* when VIX drops back to 25. Hysteresis requires VIX to fall all the way to 22 before re-entering – missing 3–8 volatility points of recovery premium in the 22–25 range, which is precisely where mean-reversion P&L is strongest.

Smooth sizing (0.74) fails because it maintains exposure during the worst drawdowns rather than exiting completely. The delay approach (1.04) adds modest confirmation value but still underperforms immediate re-entry.

### 7.4 Implications

The whipsaw problem is real but less damaging than anticipated. VIX typically drops through 25 quickly after spikes, so the hard filter re-enters fast enough to capture most recovery premium. The performance driver is *crisis avoidance*, not re-entry optimization.

This reinforces the Stage 3 conclusion: filter effectiveness depends on crisis type (short V-shape vs. prolonged stress), which cannot be predicted ex-ante. No amount of re-entry rule engineering changes this fundamental uncertainty.

## 8 Conclusions

### 8.1 Summary of Findings

1. **The VRP exists:** Mean of 3.5 vol points, positive 85% of days, highly statistically significant after autocorrelation adjustment.
2. **Naive harvesting is risky:** Sharpe 0.38 with 36% max drawdown and -27% worst month. This is tail-risk compensation.
3. **Filters appear to help:** Full-sample Sharpe improves to 1.2–1.5 with regime filters, max drawdown reduced to 17–25%.
4. **But statistical confidence is low:** Bootstrap CIs are wide and overlapping. Cannot reject that improvement is noise.
5. **OOS outperformance was regime luck:** 2019–2024 was 99th percentile favorable for filtering due to COVID's short, sharp nature.

6. **Filter effectiveness depends on re-entry speed:** Filters help during severe crises by avoiding acute losses, but overall benefit depends on how quickly VIX normalizes. Fast re-entry (COVID) preserves gains; slow re-entry (GFC) erodes them via missed recovery premium.
7. **Simple beats complex:** Hard VIX filter (Sharpe 1.18) outperforms "smarter" hysteresis and delay-based re-entry rules (0.93–1.04). The performance driver is crisis avoidance, not re-entry optimization.

## 8.2 Deployment Implications

A practitioner considering this strategy should recognize:

- The premium is real, but so is the tail risk
- Filters provide partial insurance, not elimination of drawdowns
- Filter effectiveness cannot be predicted ex-ante
- Position sizing should be conservative given uncertainty
- The strategy has negative correlation with equity markets during crises – useful for portfolio diversification but painful as a standalone

## 8.3 Limitations

Beyond implementation frictions noted in Section 2, this analysis uses VIX as an imperfect proxy for ATM implied volatility, assumes continuous availability of variance swap exposure at fair value, and does not model path-dependency in option replication. The regime classifications are necessarily ex-post; real-time crisis identification is harder.

## 8.4 Final Thoughts

The honest conclusion is not that this strategy "works" or "doesn't work," but that its performance is highly regime-dependent in ways that cannot be forecasted. A backtested Sharpe of 1.2 becomes less impressive when you realize it was achieved with 99th percentile luck on regime.

This is the nature of volatility strategies: they harvest a real premium that exists for a reason (compensating tail risk), and no amount of filtering eliminates the fundamental risk transfer. Deploy with humility.

## References

- [1] Carr, P. and Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, 22(3), 1311–1341.
- [2] Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11), 4463–4492.
- [3] Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3), 477–492.

## A Figures

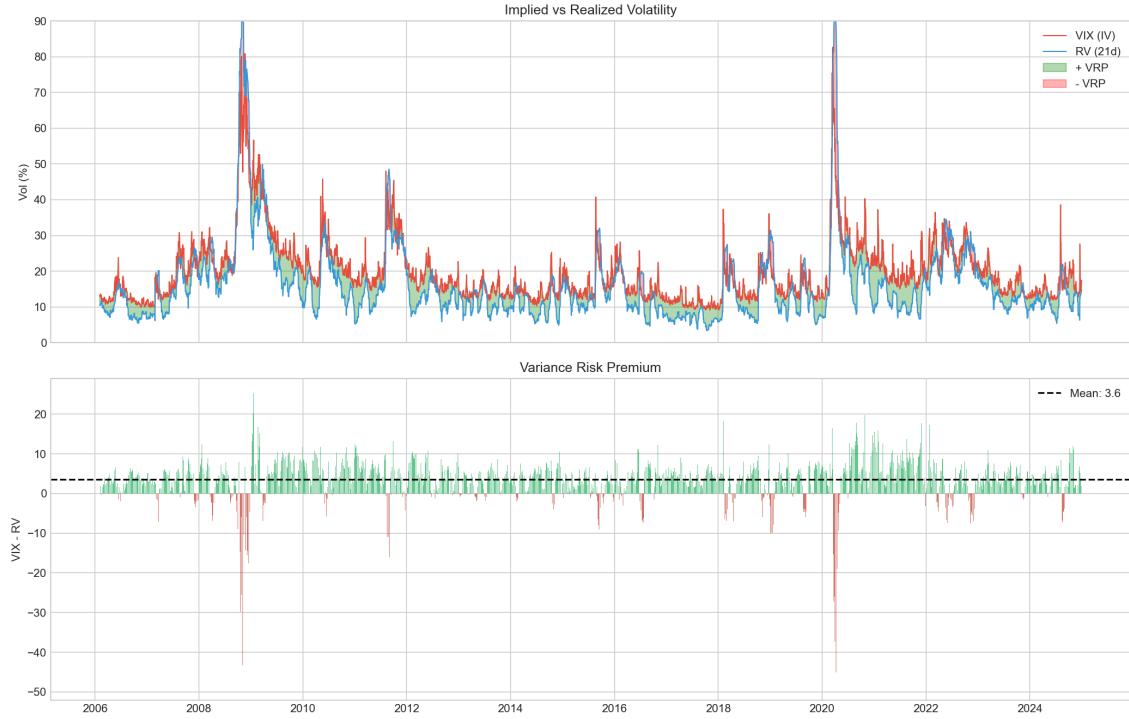


Figure 2: Implied volatility (VIX) consistently exceeds realized volatility, with the gap (variance risk premium) shown below. Premium turns negative during crisis periods (2008 GFC, 2020 COVID) when realized volatility spikes above implied.

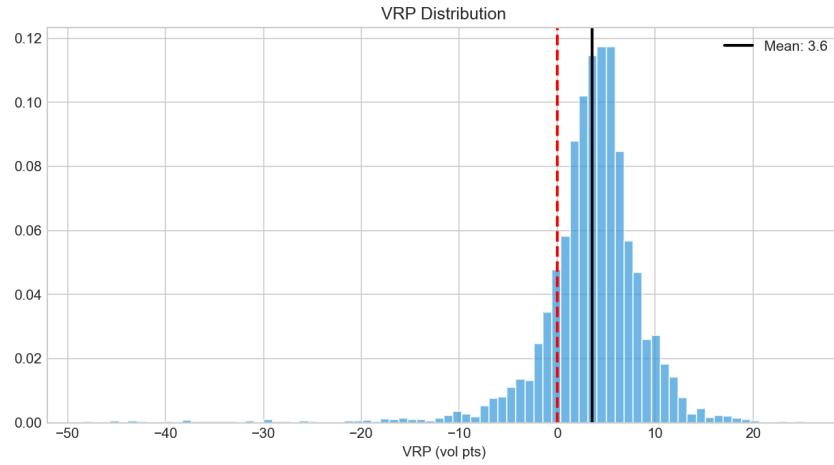


Figure 3: Distribution of daily VRP shows positive mean (3.6 vol points) but substantial negative skewness and fat left tail from crisis periods.

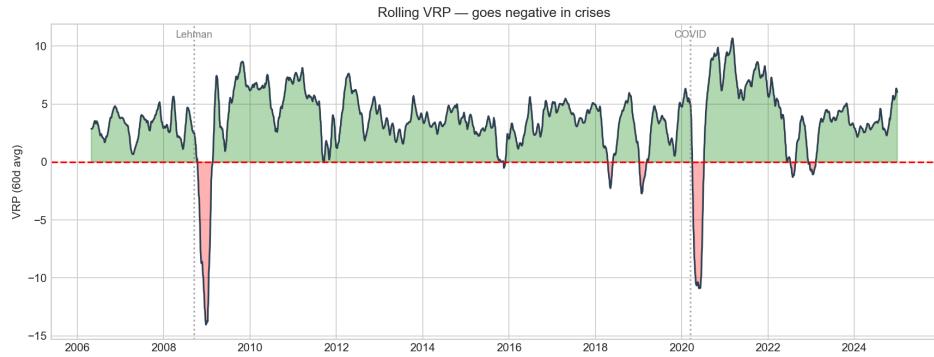


Figure 4: Rolling 60-day VRP showing regime dependence. Premium is persistently positive but turns sharply negative during Lehman (2008) and COVID (2020) crises.

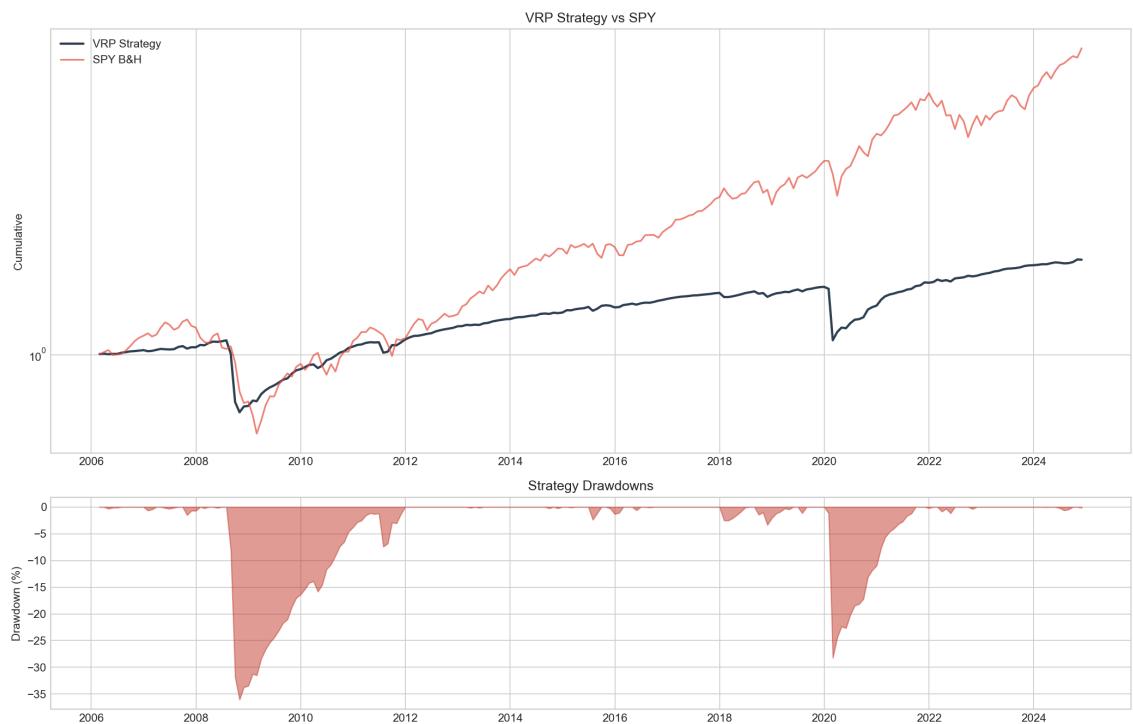


Figure 5: Unfiltered VRP strategy (10% vol target) shows modest growth with severe drawdowns during crises. Compare to SPY buy-and-hold for context.

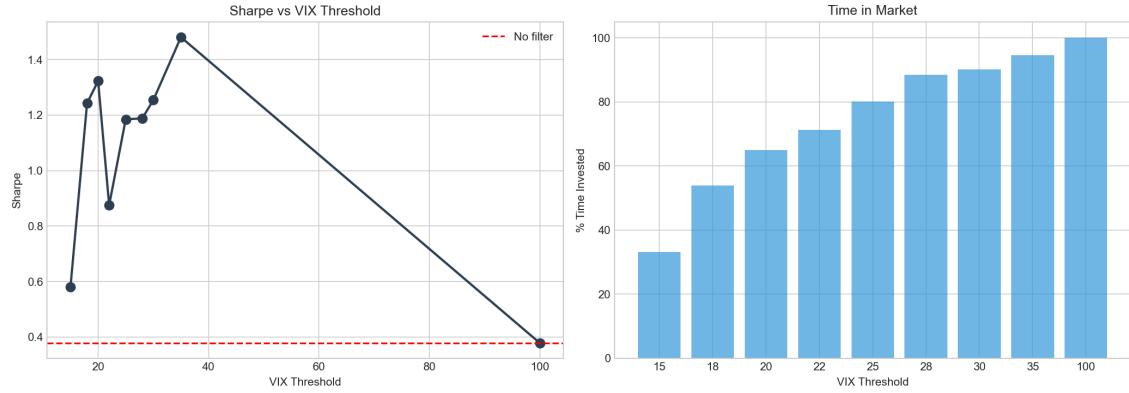


Figure 6: Sharpe ratio as a function of VIX threshold (left) and corresponding time in market (right). Performance peaks around  $VIX \leq 35$ , with trade-off between premium capture and tail risk avoidance.

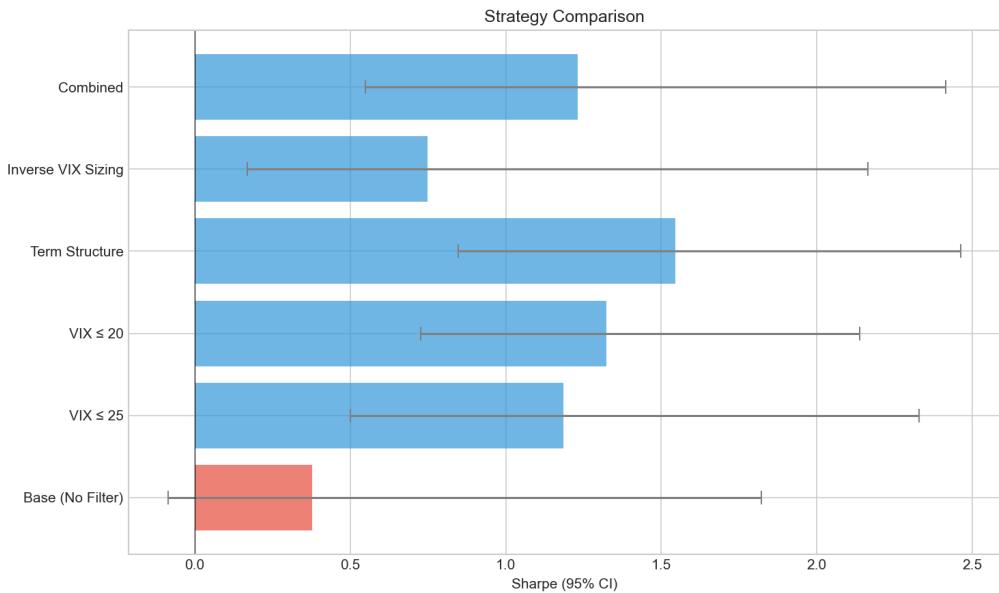


Figure 7: Bootstrap 95% confidence intervals on Sharpe ratios. Wide intervals indicate substantial uncertainty; most CIs overlap, preventing definitive ranking.

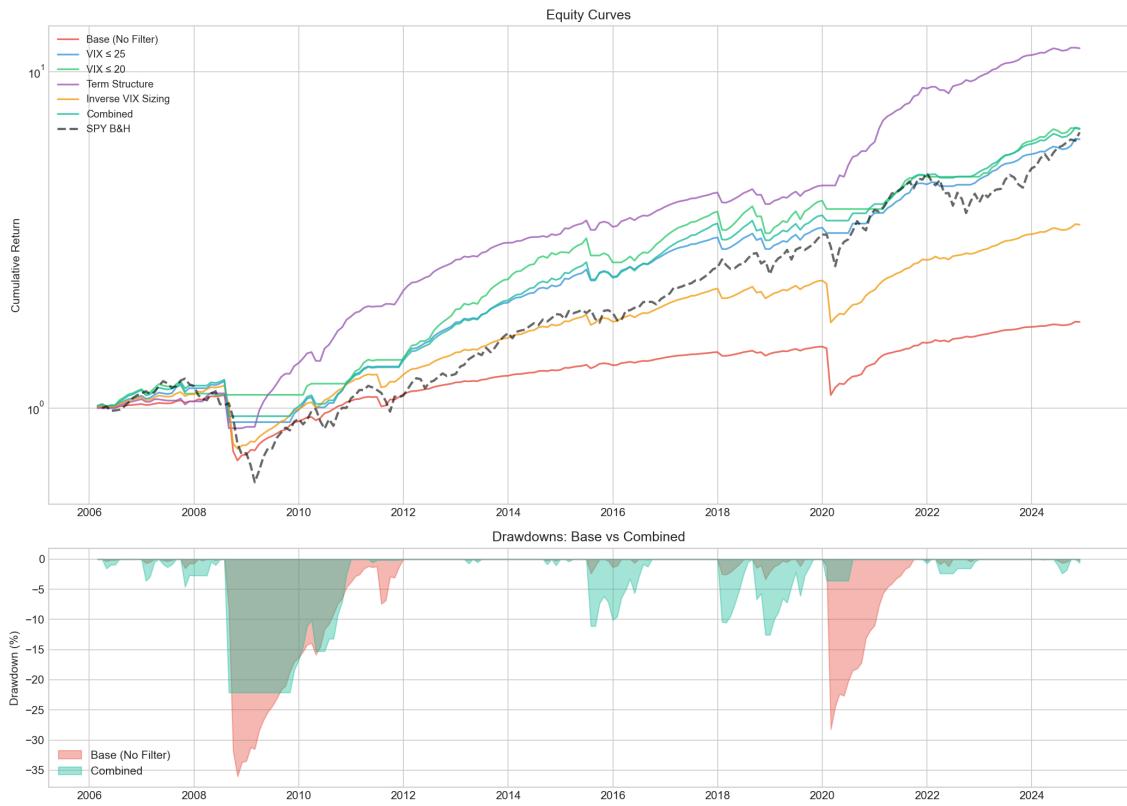


Figure 8: Equity curves for filtered strategies vs. unfiltered baseline and SPY. Filtered strategies show smoother paths with reduced crisis drawdowns.

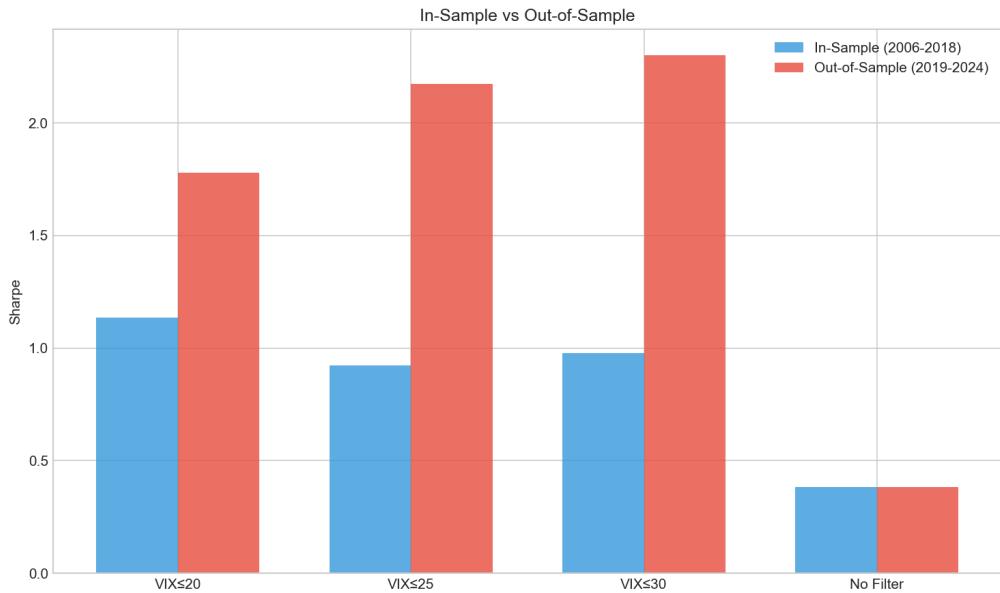


Figure 9: Out-of-sample Sharpe ratios exceed in-sample for filtered strategies – opposite of typical degradation pattern, suggesting regime luck rather than robust alpha.

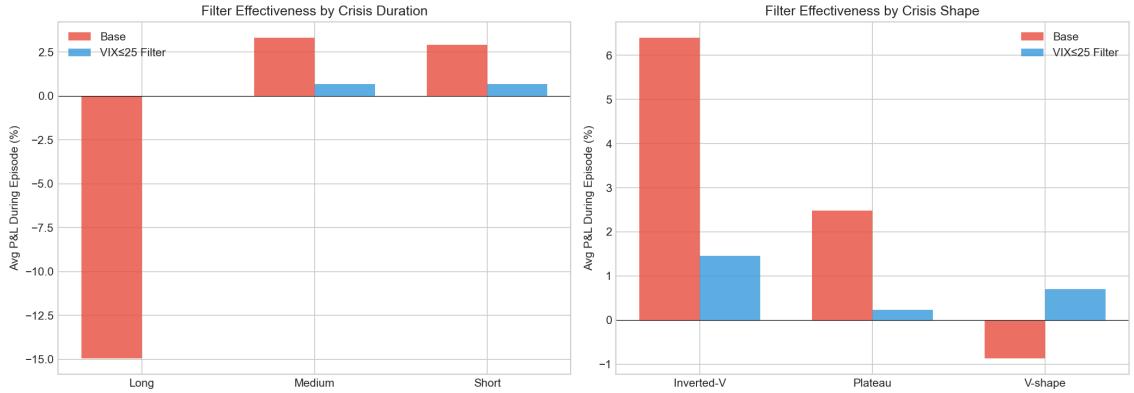


Figure 10: Filter effectiveness varies by crisis characteristics. Left: by duration (filter helps during long crises, costs premium during short/medium). Right: by shape.

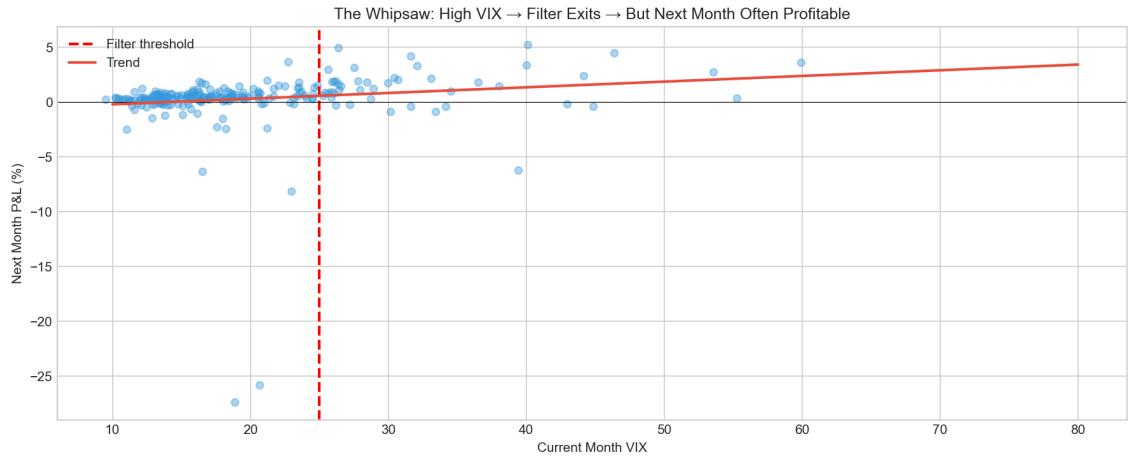


Figure 11: High VIX months (where filter exits) are often followed by positive next-month returns due to volatility mean reversion. The filter misses this recovery premium.

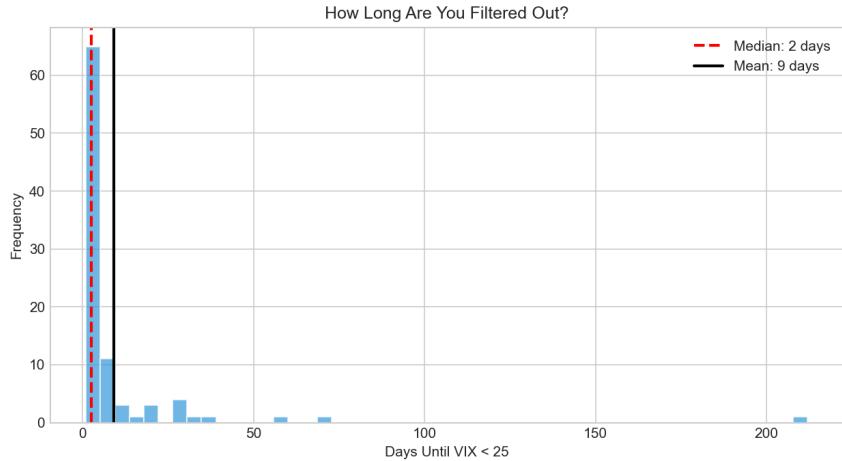


Figure 12: Distribution of filter-out duration. Median is only 2 days (quick spikes), but right tail extends to 212 days (GFC), explaining regime-dependent effectiveness.

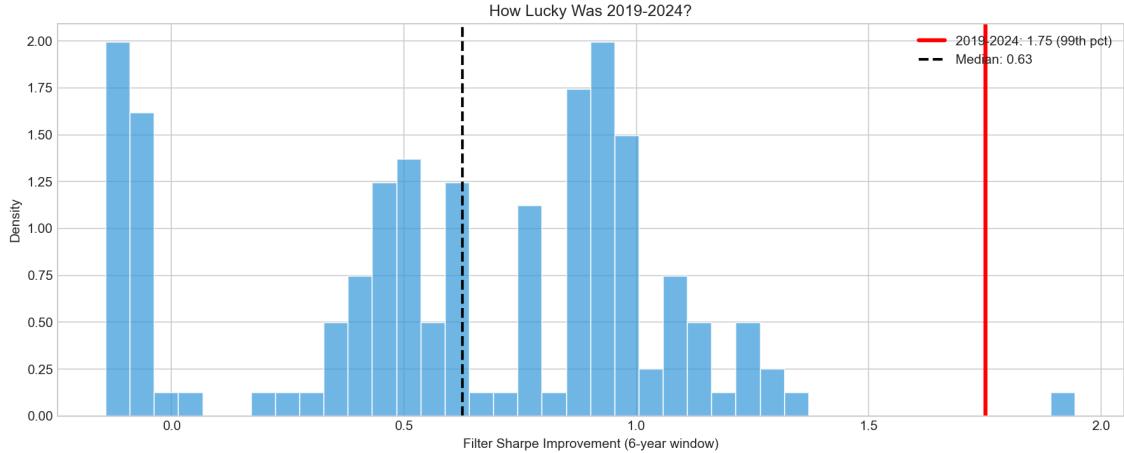


Figure 13: The 2019–2024 test period achieved filter improvement in the 99th percentile of all rolling 6-year windows, indicating exceptionally favorable regime conditions.

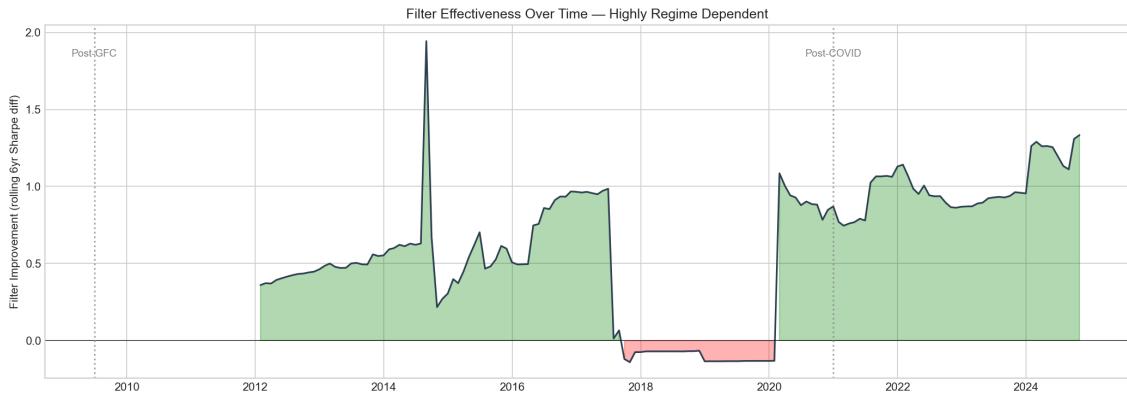


Figure 14: Rolling filter improvement shows high regime dependence. Effectiveness varies from negative (2015–2018, few crises) to strongly positive (post-COVID), with no stable pattern.

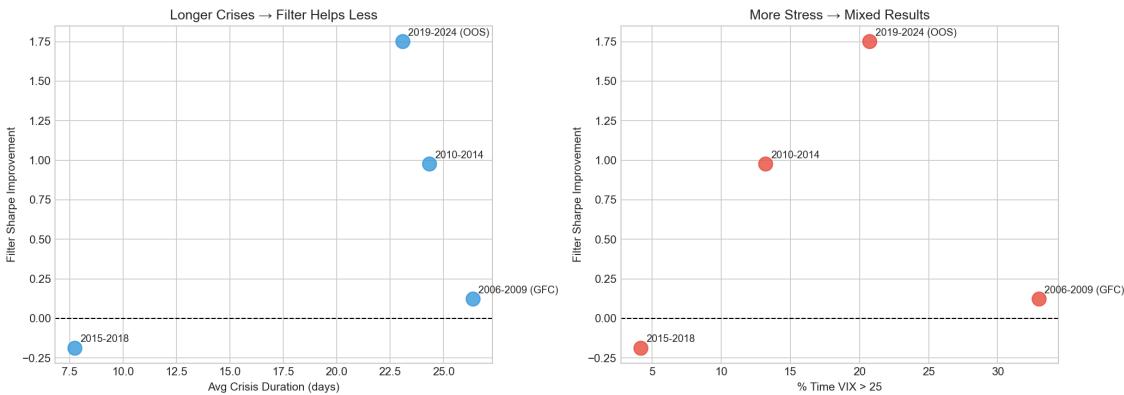


Figure 15: Filter improvement by period. Left: longer average crisis duration correlates with lower filter benefit (miss recovery). Right: relationship with stress frequency is mixed.

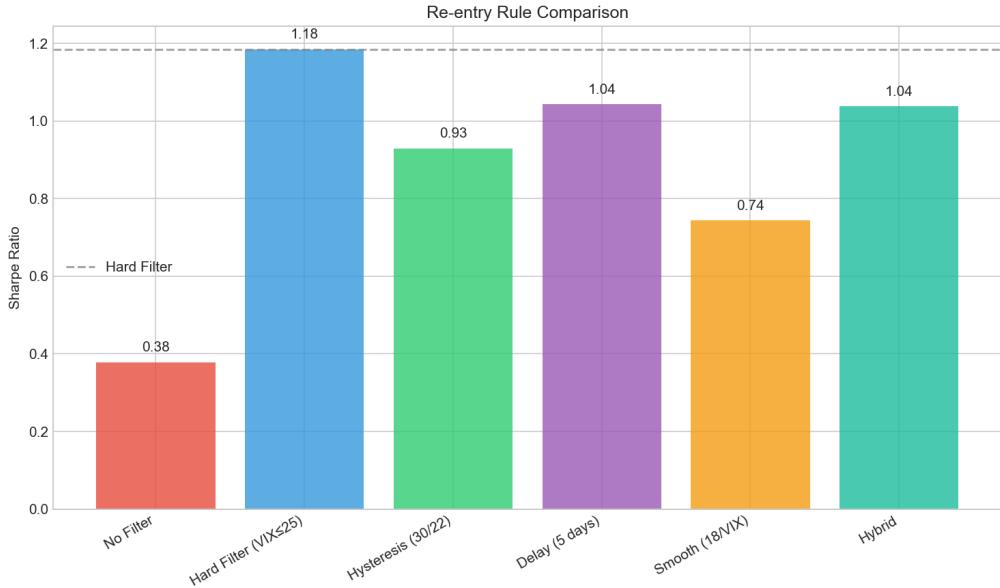


Figure 16: Comparison of re-entry mechanisms. The simple hard filter (1.18) outperforms "smarter" alternatives including hysteresis (0.93) and hybrid approaches (1.04).

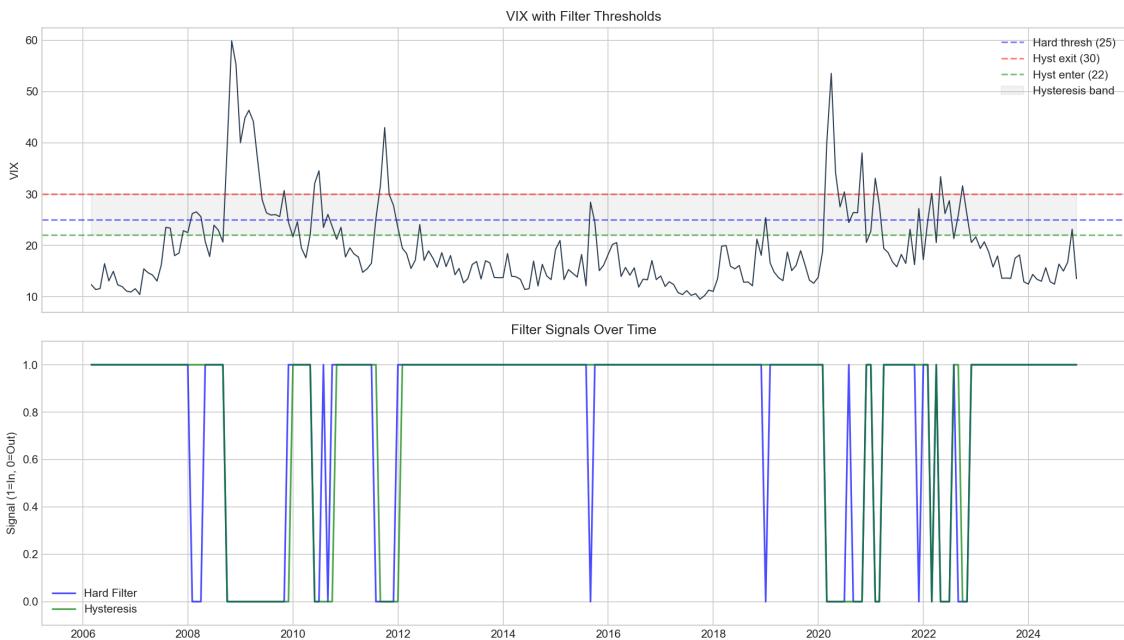


Figure 17: Signal comparison between hard filter ( $VIX \leq 25$ ) and hysteresis (exit  $>30$ , enter  $<22$ ). Hysteresis stays out longer during recovery, missing premium in the 22–25 VIX range.

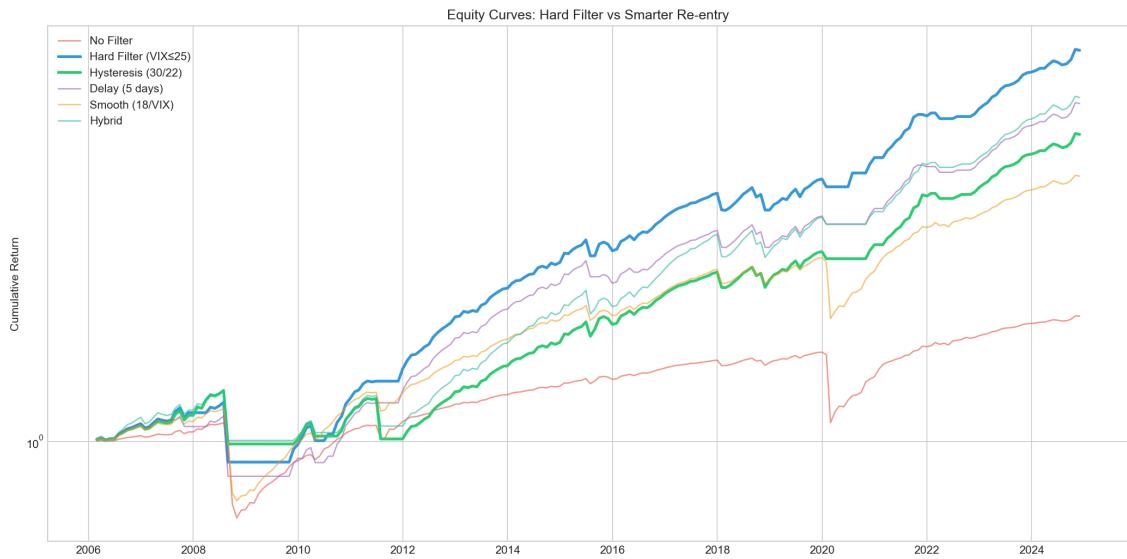


Figure 18: Equity curves for re-entry rule variants. Hard filter (blue) dominates throughout, with hysteresis (green) lagging due to slower re-entry during recovery periods.