Realized Volatility Timing for Short-Volatility Carry Strategies

Quantitative Research Project in Volatility Strategies

Théo Verdelhan — Léo Renault
Paris Dauphine University – PSL

May 4, 2026

Abstract

This paper studies whether model-based estimates of realized volatility can improve the execution of short-volatility carry strategies. Using daily option-chain data on SPY and AAPL, we build an at-the-money implied-volatility proxy, estimate latent variance with a Heston state-space model filtered through an Unscented Kalman Filter (UKF), and dynamically scale a short one-week strangle according to the spread between implied volatility and estimated realized volatility. The UKF-based timing rule is compared against both a static carry strategy and a simpler benchmark based on 21-day realized volatility. Two main results emerge. First, volatility timing is economically valuable: in both assets, dynamic timing improves on static carry. Second, the incremental value of the latent-state model is mixed: while the Heston-UKF signal is directionally useful and improves baseline carry, the simpler realized-volatility benchmark remains more competitive in the current implementation. The project is implemented as a fully reproducible Python research pipeline, from market-panel construction to reporting and backtesting.

1 Introduction

Short-volatility carry strategies monetize the tendency of implied volatility to trade above subsequently realized volatility. In practice, however, the profitability of these strategies depends heavily on position timing and exposure management. A static short-volatility allocation can perform reasonably well in benign market conditions, but it can suffer when realized volatility rises sharply or when the volatility risk premium compresses.

This project asks a focused quant-research question: can a model-based estimate of realized volatility improve the execution of an otherwise unchanged short-volatility carry strategy? More precisely, can the spread

$$s_t = \sigma_t^{IV} - \widehat{\sigma}_t$$

be used as a useful timing signal for dynamically scaling short-volatility exposure?

To answer this question, we estimate latent variance through a Heston state-space model combined with an Unscented Kalman Filter, then compare the resulting timing rule against two conservative baselines:

1. a static carry strategy with constant allocation,
2. a dynamic rule based on a much simpler 21-day realized-volatility proxy.

The design of the experiment is deliberately conservative. The option strategy itself is kept fixed throughout the analysis. Only the exposure multiplier changes over time. This isolates the value of the timing signal and makes the comparison closer to what a quant researcher would want in an ablation-style empirical study.

The main contributions of the project are threefold. First, it provides an end-to-end implementation of a volatility-timing pipeline on real option data. Second, it evaluates whether latent-state modeling adds value beyond strong empirical benchmarks. Third, it documents the practical trade-off between model sophistication, signal stability, and computational cost in a realistic research setting.

2 Research Design and Strategy Overview

The base strategy is a short one-week strangle. The timing model never flips the sign of the position into a long-volatility trade. Instead, it scales exposure between a flat allocation and a leveraged short-volatility allocation, depending on whether the implied-realized spread is favorable or unfavorable.

This distinction matters. The objective of the project is not to build a directional volatility-trading system. It is to improve the execution of a carry strategy while preserving its economic interpretation.

3 Data and Sample Construction

The empirical analysis uses option-market data available locally in the project repository. For each trading date and ticker, the pipeline constructs a market panel containing:

- the underlying spot price,
- daily log-returns,
- an at-the-money implied-volatility proxy,
- realized-volatility benchmarks computed from rolling returns.

The implied-volatility input is extracted from the option chain by selecting the option closest to at-the-money on both the call and put sides, around a target maturity of 30 calendar days, and averaging the two implied volatilities. This is intentionally simple and robust, even though it does not exploit the full information content of the implied-volatility surface.

Two assets are studied:

- **SPY**, used as the main reference asset for a volatility-carry study,
- **AAPL**, used as a robustness check on a more idiosyncratic single-name underlier.

The full-sample windows are:

- SPY: 2020-01-02 to 2022-12-30,
- AAPL: 2016-01-04 to 2023-03-31.

An additional short sample is used only for diagnostics and interpretability. It helps visualize the signal mechanics, the warm-up period of the rolling calibration, and the behavior of the allocation rule during stressed conditions.

4 Heston State-Space Model and UKF Estimation

The latent variance process follows a Heston specification:

$$dS_t = \mu S_t \, dt + S_t \sqrt{v_t} \, dW_t^{(1)}, \tag{1}$$

$$dv_t = \kappa(\theta - v_t) \, dt + \xi \sqrt{v_t} \, dW_t^{(2)}, \tag{2}$$

$$d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \tag{3}$$

Here, S_t is the observed spot price and v_t is the latent instantaneous variance. The parameters μ, κ, θ, ξ, and ρ respectively govern drift, mean reversion, long-run variance, volatility of variance, and spot-volatility correlation.

Using daily log-returns

$$r_{t+1} = \log \left(\frac{S_{t+1}}{S_t} \right),$$

a first-order discretization yields the nonlinear state-space system

$$v_{t+1} \approx v_t + \kappa(\theta - v_t)\Delta t + \xi \sqrt{v_t \Delta t} \, \varepsilon_{t+1}^v, \tag{4}$$

$$r_{t+1} \approx \left(\mu - \frac{1}{2} v_t \right) \Delta t + \sqrt{v_t \Delta t} \, \varepsilon_{t+1}^S, \tag{5}$$

with correlated shocks:

$$\mathrm{Corr}(\varepsilon_{t+1}^S, \varepsilon_{t+1}^v) = \rho.$$

Because v_t is not observed directly, the latent state is filtered through an Unscented Kalman Filter rather than a linear Kalman filter. This preserves the nonlinear structure of the model while remaining computationally tractable for rolling estimation.

5 Rolling Calibration and Benchmark Construction

The Heston parameters are re-estimated over a rolling window. Denoting the parameter vector by

$$\Theta = (\mu, \kappa, \theta, \xi, \rho, v_0),$$

the rolling estimator at time t is obtained from

$$\widehat{\Theta}_t = \arg\max_{\Theta} \sum_{u=t-W+1}^{t} \log p(r_u \mid r_{t-W+1:u-1}; \Theta),$$

where the likelihood is evaluated through the UKF.

In the baseline implementation used for the full-sample experiments, the rolling calibration uses:

- a 42-business-day estimation window,
- parameter refits every 10 observations,
- an optimizer budget of 12 iterations per refit.

To judge whether the latent-state model is genuinely useful, we compare it against a simple historical benchmark:

$$RV_t^{21} = \sqrt{252}\,\mathrm{Std}(r_{t-20}, \ldots, r_t),$$

and also report a slower 63-day benchmark:

$$RV_t^{63} = \sqrt{252}\,\mathrm{Std}(r_{t-62}, \ldots, r_t).$$

The two spreads studied in the project are therefore:

$$s_t^{UKF} = \sigma_t^{IV} - \widehat{\sigma}_t^{UKF}, \tag{6}$$
$$s_t^{RV} = \sigma_t^{IV} - RV_t^{21}. \tag{7}$$

This benchmark comparison is central to the project. The real question is not whether the Heston-UKF signal beats an untimed carry strategy, but whether it beats a simpler and cheaper realized-volatility alternative.

6 Dynamic Allocation Rule

The spread is transformed into a position-sizing rule in four steps.

First, the raw spread is smoothed with a short exponentially weighted moving average:

$$\widetilde{s}_t = \mathrm{EWMA}_5(s_t).$$

Second, the smoothed spread is standardized with a 21-day rolling window:

$$z_t = \frac{\widetilde{s}_t - \overline{\widetilde{s}}_{t,21}}{\mathrm{Std}_{t,21}(\widetilde{s})}.$$

Third, the z-score is clipped:

$$z_t^{clip} = \min(3, \max(-3, z_t)).$$

Fourth, the clipped signal is converted into a dynamic allocation multiplier:

$$a_t = \mathrm{clip}(a_0 + \lambda z_t^{clip}, a_{\min}, a_{\max}),$$

with baseline values

$$a_0 = 1.0, \qquad \lambda = 0.4, \qquad a_{\min} = 0, \qquad a_{\max} = 2.$$

The interpretation is intuitive. When implied volatility is high relative to model-estimated or historically realized volatility, the spread is positive and the short-volatility carry position is reinforced. When the spread weakens or turns negative, exposure is scaled down. The signal can switch the carry trade off entirely, but it never turns the portfolio into a long-volatility strategy.

7 Experimental Protocol

Each experiment compares three strategies:

1. **Base carry**: short one-week strangle with constant unit allocation;
2. **Timed carry UKF**: same base trades multiplied by the UKF-based allocation signal;
3. **Timed carry RV_21d**: same base trades multiplied by the benchmark allocation signal based on historical realized volatility.

The evaluation relies on four standard metrics:

- total return,
- Sharpe ratio,
- Calmar ratio,
- maximum drawdown.

This setup intentionally isolates the contribution of the signal. Since the option strategy itself is unchanged, any performance difference comes from volatility timing rather than from a different option structure.

Table 1: Baseline experiment configuration

Parameter	Value
Implied-volatility maturity target	30 calendar days
Rolling estimation window	42 business days
Refit frequency	Every 10 observations
Optimizer iterations	12
Signal smoothing	EWMA span of 5
Z-score window	21 business days
Base allocation	1.0
Allocation sensitivity	0.4
Allocation bounds	[0, 2]

8 Illustrative Short-Sample Diagnostics

Before turning to the full-sample results, it is useful to inspect a short window where the mechanics of the signal are visually transparent. Figure 1 shows an illustrative SPY diagnostic sample. It highlights four important features:

1. the UKF volatility estimate does not start on the first observation because of the warm-up required by rolling calibration;
2. the UKF spread and the realized-volatility spread can diverge meaningfully in stressed periods;
3. the dynamic allocation rule reacts mechanically to those spread differences;
4. the timing model mainly changes *position size*, not direction.

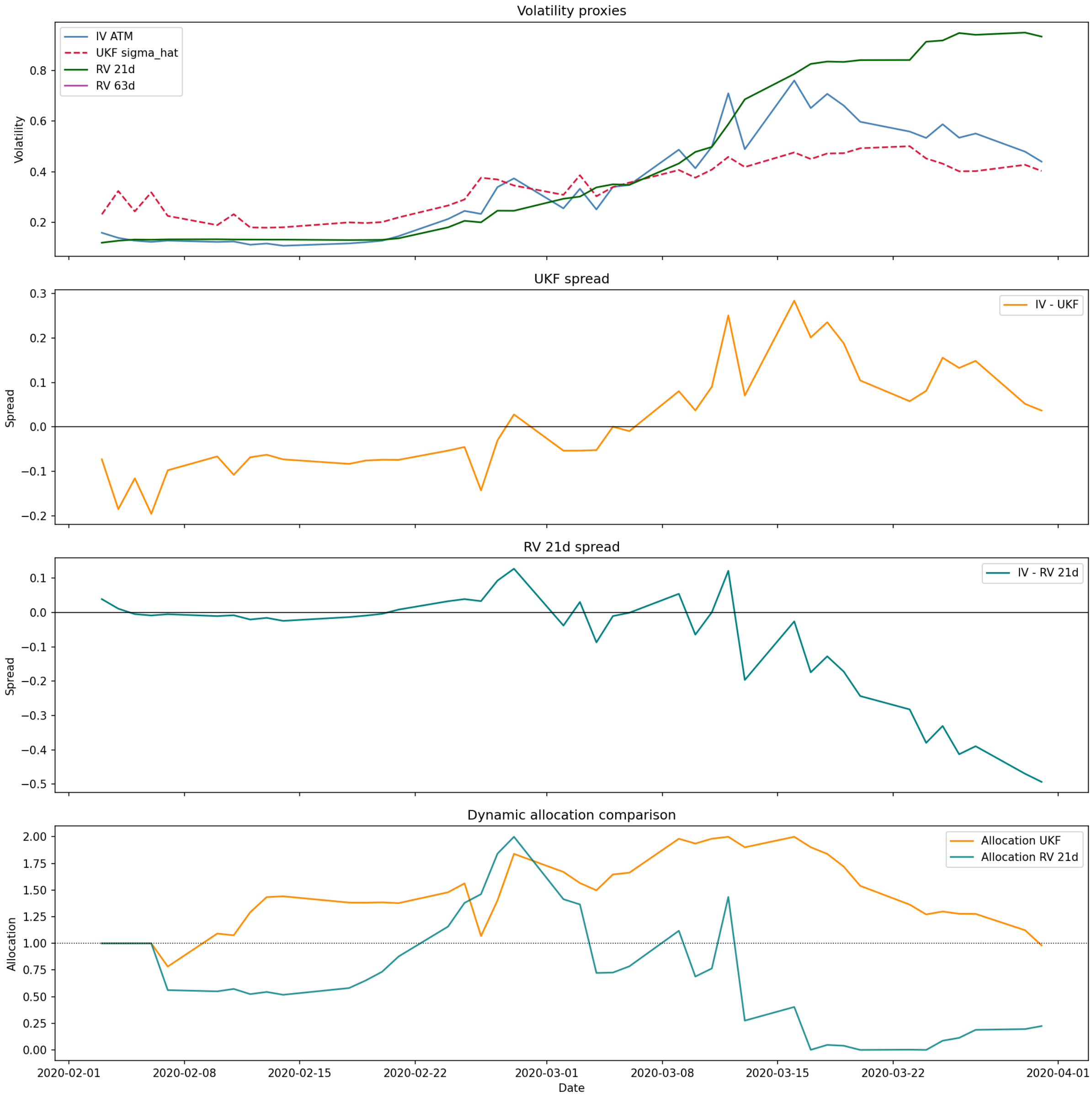

Figure 1: Illustrative SPY short-sample diagnostic view. The top panel compares implied volatility, UKF-estimated volatility, and historical realized-volatility proxies. The lower panels show the implied-realized spreads and the resulting allocation multipliers. This figure is especially useful for reading the warm-up effect and the difference between model-based and historical timing rules.

9 Full-Sample Results

9.1 Performance Comparison

Table 2 reports the full-sample performance of the three strategies on SPY and AAPL.

Table 2: Full-sample backtest comparison

Asset	Strategy	Total return	Sharpe	Calmar	Max drawdown
Panel A: SPY (2020-01-02 to 2022-12-30)					
SPY	Base carry	3.81%	0.486	0.418	-2.97%
SPY	Timed carry UKF	4.51%	0.497	0.495	-2.97%
SPY	Timed carry RV_21d	5.20%	0.855	0.933	-1.77%
Panel B: AAPL (2016-01-04 to 2023-03-31)					
AAPL	Base carry	18.66%	0.624	0.512	-4.59%
AAPL	Timed carry UKF	26.01%	0.693	0.579	-5.50%
AAPL	Timed carry RV_21d	28.99%	0.977	0.743	-4.65%

Two conclusions are immediate.

First, dynamic volatility timing is economically useful. On both assets, the timed strategies improve on the static carry benchmark in total return. On SPY, the UKF-timed strategy modestly improves return and Sharpe while leaving drawdown unchanged. On AAPL, the UKF timing rule produces a much larger uplift in total return, albeit with a somewhat larger drawdown than the base carry strategy.

Second, the simpler realized-volatility timing rule remains the strongest specification in the current implementation. It dominates static carry and also outperforms the UKF-timed strategy on both assets, especially on risk-adjusted metrics. This is a valuable research result: the more complex model is not automatically the better trading rule. A more cautious interpretation is required for the UKF results. On SPY, the improvement of the UKF-timed strategy over the base carry strategy is economically small: total return increases from 3.81% to 4.51%, while the Sharpe ratio only moves from 0.486 to 0.497. This suggests that, although the UKF signal is directionally useful, its incremental contribution on the main reference asset is limited in the current specification. By contrast, the RV_21d rule generates a much larger improvement in risk-adjusted performance, with a Sharpe ratio of 0.855 and a lower maximum drawdown. Therefore, the evidence supports the usefulness of volatility timing more strongly than the superiority of the Heston-UKF model itself.

9.2 SPY: Main Reference Asset

Figure 2 displays the net asset value trajectories for the SPY full-sample experiment. The ordering in the plot confirms the numerical results reported in Table 2. Both dynamic strategies dominate the untimed carry benchmark by the end of the sample, while the RV_21d timing rule delivers the strongest terminal NAV and the shallowest drawdown profile.

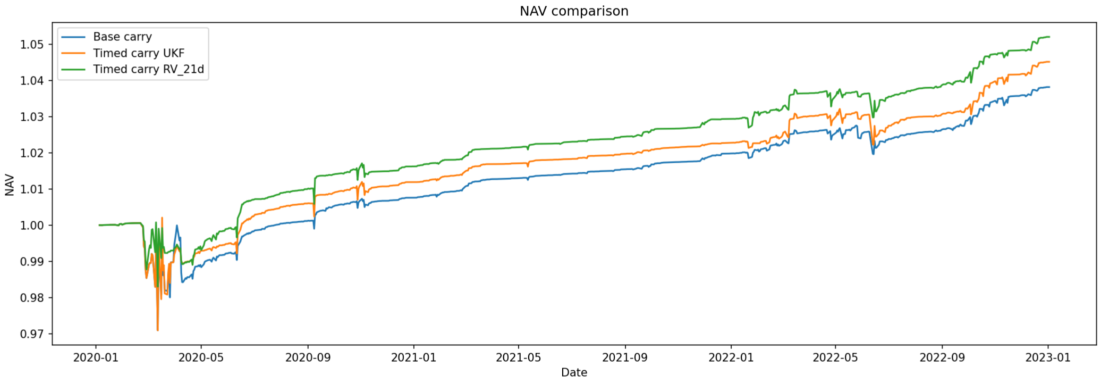

Figure 2: SPY full-sample NAV comparison. Dynamic timing improves on the static carry benchmark, but the simple RV_21d rule remains stronger than the Heston-UKF timing signal in the current implementation.

The practical interpretation is important. The UKF signal is not useless; it improves the strategy over static carry. But once a strong benchmark is introduced, the incremental value of model complexity becomes less obvious. For a recruiter or research reviewer, this is precisely the kind of honest empirical finding that matters: the project does not overclaim.

9.3 AAPL: Cross-Sectional Robustness

Figure 3 reports the same NAV comparison for AAPL. The broad message survives the change in underlier. Both timing rules outperform static carry, and the simpler RV_21d benchmark again remains strongest overall.

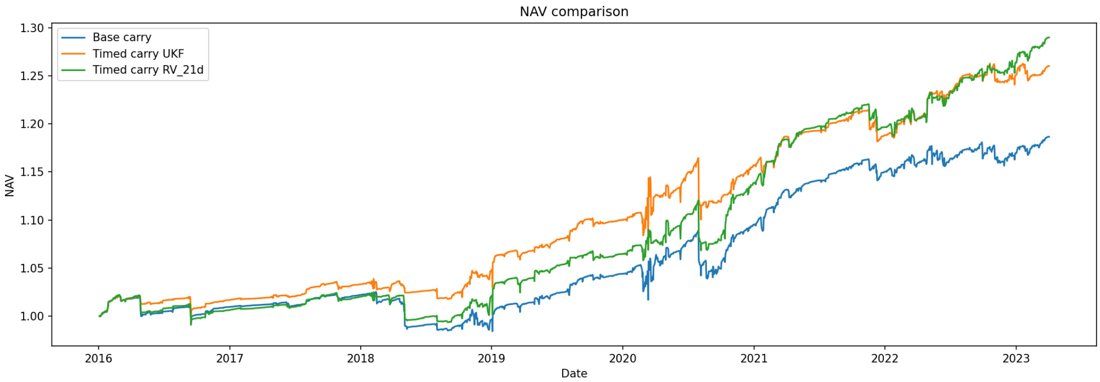

Figure 3: AAPL full-sample NAV comparison. The volatility-timing logic remains economically useful outside the index ETF case, but the realized-volatility benchmark still dominates the UKF-based timing rule in the present specification.

The AAPL result is especially interesting because it shows that the timing idea survives a move from a broad index proxy to a more idiosyncratic single-name equity. At the same time, the project remains disciplined in its interpretation: robustness of the timing idea does not imply that the latent-state filter is already the best implementation. The AAPL experiment strengthens the conclusion that timing short-volatility exposure is useful, but it also confirms the same model-selection issue. The UKF-timed strategy improves total return from 18.66% to 26.01%, which is economically meaningful. However, this comes with a larger maximum drawdown than the static carry benchmark. The RV_21d rule again delivers the best risk-adjusted performance. This suggests that the UKF signal may capture relevant volatility information, but that its estimation noise and parameter instability can reduce its practical advantage relative to a simpler historical-volatility rule.

9.4 Quality of the Volatility Proxies

To better understand the trading results, Table 3 compares the volatility proxies directly against realized-volatility benchmarks.

Table 3: Proxy comparison against realized-volatility benchmarks

Asset	Proxy	Corr. vs RV_21d	Corr. vs RV_63d	MAE vs RV_21d	MAE vs RV_63d
SPY	UKF sigma_hat	0.308	0.534	0.161	0.155
SPY	IV ATM	0.852	0.664	0.049	0.054
AAPL	UKF sigma_hat	0.380	0.623	0.171	0.158
AAPL	IV ATM	0.457	0.462	0.116	0.137

These numbers help explain why the RV_21d timing strategy is so competitive. On SPY, the ATM implied-volatility proxy is already strongly aligned with realized volatility, much more so than the UKF estimate. On AAPL, the picture is more mixed: the UKF estimate correlates better with the slower RV_63d benchmark than IV ATM does, but it still exhibits materially larger absolute errors. In other words, the UKF signal captures part of the volatility dynamics, but it is not yet a clearly superior volatility proxy. This proxy comparison is important because it links signal quality to trading performance. The UKF estimate does not fail because the Heston model is theoretically irrelevant; rather, it appears to be less aligned with the volatility measure that drives the short-horizon allocation rule. On SPY, the UKF estimate has a correlation of only 0.308 with RV_21d, compared with 0.852 for the ATM implied-volatility proxy. This weak short-horizon alignment can explain why the UKF-based spread produces a less effective timing signal. On AAPL, the UKF estimate performs better against the slower RV_63d benchmark, which suggests that the filter may be capturing lower-frequency volatility regimes rather than the short-term realized volatility needed for a one-week strangle strategy.

9.5 Parameter Stability and Model Behavior

Figure 4 shows the rolling Heston parameter estimates for SPY. The parameter paths are informative because they make the model's practical behavior visible. In particular, the rolling estimates are not fully stable across regimes, which is one reason the resulting signal can be noisier than a simple historical-volatility benchmark.

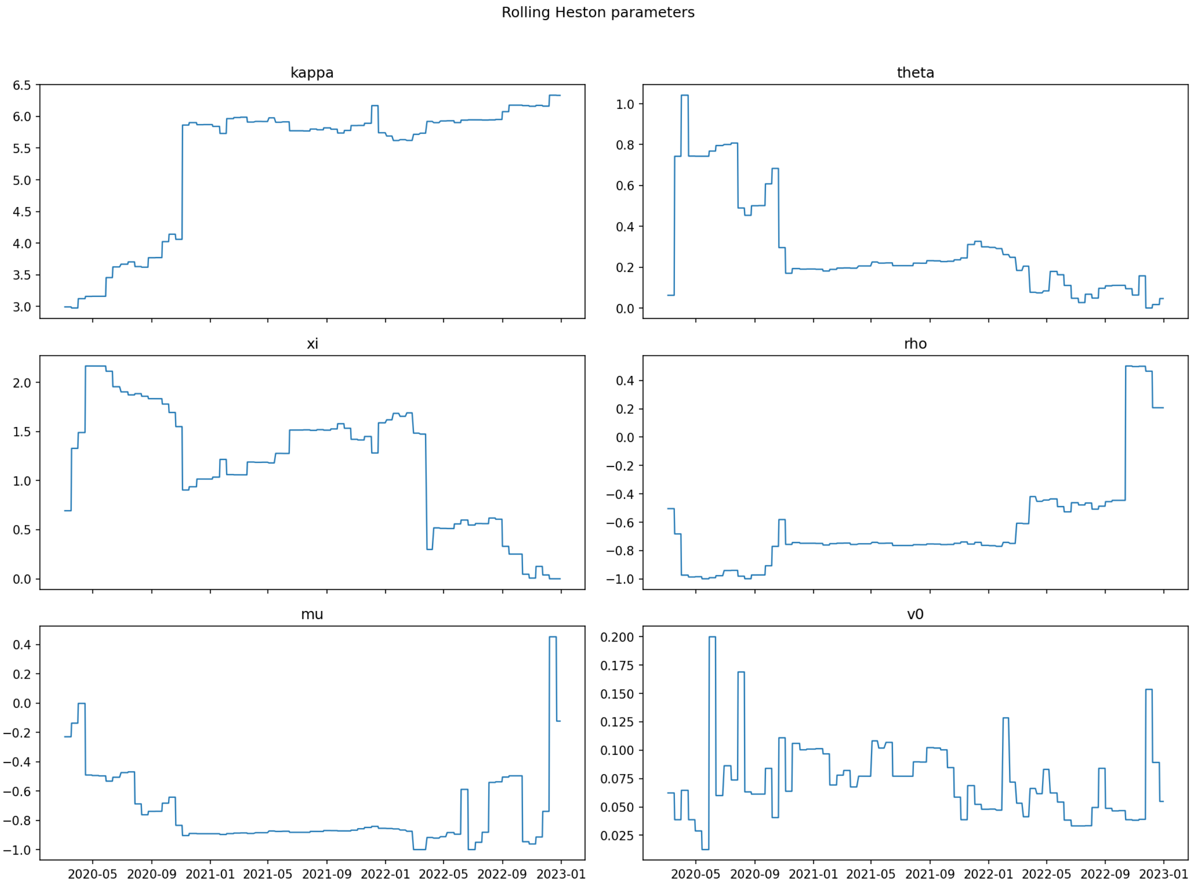

Figure 4: Rolling Heston parameter estimates on the SPY full sample. The figure illustrates the degree of parameter variation induced by repeated rolling maximum-likelihood refits.

This is not necessarily a flaw. Time variation in the estimated parameters may reflect genuine changes in volatility regimes. But it is a reminder that richer models introduce an additional layer of estimation risk, and that trading performance depends not only on model elegance but also on signal robustness. The parameter instability also raises an important practical issue. Since the allocation rule is built from the filtered volatility estimate, unstable Heston parameters can translate into unstable exposure decisions. This matters especially because the calibration window is relatively short, at 42 business days, and the optimizer budget is deliberately limited for computational reasons. The resulting estimates may therefore reflect a combination of genuine regime changes, sampling noise, and optimization error. This is one reason why the simpler RV_21d benchmark remains difficult to beat: it has fewer degrees of freedom and is less exposed to estimation risk.

10 Implementation and Reproducibility

From an engineering perspective, the project is organized as an end-to-end research pipeline rather than a single notebook prototype. The main modules are:

- `realized_vol_timing/data.py`: loads option data and constructs the market panel;
- `realized_vol_timing/heston_ukf.py`: latent-variance model, UKF, and rolling calibration;
- `realized_vol_timing/signals.py`: spread construction and allocation rule;
- `realized_vol_timing/strategy.py`: trade allocation and backtest summaries;
- `realized_vol_timing/experiments.py`: orchestration of SPY and AAPL experiments;
- `realized_vol_timing/reporting.py`: export of tables and figures;
- `scripts/run_dynamic_carry.py`: command-line entry point for reproducible runs.

This structure is worth emphasizing in a recruiter-facing document because it shows that the project is not only a research idea but also a reusable implementation. The notebook `Project_Report.ipynb` serves as the narrative layer, while the Python package contains the reusable research components.

A representative command-line run is:

```
python scripts/run_dynamic_carry.py --ticker SPY --full-sample
python scripts/run_dynamic_carry.py --ticker AAPL --full-sample
```

11 Limitations and Extensions

The project is fully operational, but several limitations remain.

First, the implied-volatility input is built from a simple ATM proxy. A richer construction based on the actually traded option legs, or on an estimated implied-volatility surface such as SSVI or SABR, would provide a more faithful representation of market-implied volatility.

Second, the rolling Heston calibration is computationally expensive. The repeated likelihood maximization makes long runs materially slower than the historical-volatility benchmark and complicates broad hyperparameter searches.

Third, the benchmark is strong. The 21-day realized-volatility proxy is simple, interpretable, cheap to compute, and difficult to beat. This is not a disappointing outcome; it is exactly the type of benchmark discipline that good quant research requires.

Fourth, the current allocation rule does not allow sign reversal. It can switch off short-volatility exposure, but it never becomes long volatility. This keeps the economic interpretation clean, but it may also leave money on the table in crisis-like environments.

Fifth, the current results are evaluated mainly through economic performance metrics. A more complete empirical study would test the statistical significance of the performance differences, for example through bootstrap confidence intervals for Sharpe ratios or paired tests on strategy returns. This is particularly important for SPY, where the UKF improvement over static carry is directionally positive but numerically small.

Sixth, the current evaluation does not fully characterize tail risk. Since short-volatility strategies are exposed to rare but severe losses, future work should report additional downside metrics such as worst daily return, conditional value-at-risk, skewness, kurtosis, and stress-period attribution. This would make the risk analysis more consistent with the economic nature of short-volatility carry.

Natural extensions include:

- richer observation equations using additional option information,
- surface-based implied-volatility inputs,
- explicit transaction-cost sensitivity analysis,
- regime-aware or Bayesian parameter updating,
- broader cross-sectional tests across underliers.

12 Conclusion

This project shows that realized-volatility timing can improve the execution of short-volatility carry strategies, but also that model sophistication must be judged against strong empirical baselines. The Heston-UKF framework is economically meaningful and directionally useful: it improves on static carry on both SPY and AAPL. However, in the current implementation, a simpler timing rule based on 21-day realized volatility remains more competitive.

For a quant research workflow, this is a meaningful result rather than a failure. The project demonstrates the full research loop: question formulation, state-space modeling, benchmark design, backtesting, diagnostics, and honest interpretation of mixed evidence. It also leaves a clear roadmap for future iterations, especially around implied-volatility construction, parameter stability, and computational efficiency.