

Time Series Analysis of QQQ Risk Premium

April 2026

1 Introduction

1.1 Motivation and Background: The Role of Variance Risk

Volatility is central to derivative pricing, risk management, and portfolio construction, yet it is latent: it must be inferred from return data under the physical measure \mathbb{P} or from option prices under the risk-neutral measure \mathbb{Q} . Classical time-series treatments emphasize both linear dependence structures and nonlinear state-space representations for latent dynamics [1], [2]. The wedge between \mathbb{Q} -implied variance and \mathbb{P} -expected realized variance is the **variance risk premium (VRP)**. A positive VRP is often interpreted as compensation demanded by sellers of volatility protection; its dynamics around stress episodes provide a useful empirical summary of how option markets price variance relative to return-based forecasts.

1.2 Research Question: POMP-Based VRP Estimation

This project asks whether a Heston-style stochastic volatility model, estimated as a partially observed Markov process (POMP) on QQQ returns under \mathbb{P} , yields a credible 21-day variance forecast that, when compared to the VIXN-implied \mathbb{Q} expectation mapped to the same units, produces an interpretable VRP time series for the Nasdaq-100 exposure carried by QQQ.

The empirical strategy is: (i) estimate a rich GARCH family as a benchmark; (ii) estimate the POMP specification by iterated filtering (IF2); (iii) filter latent variance on the full sample; (iv) construct $\mathbb{E}_t^{\mathbb{P}}[RV_{t+21}]$ in closed form; (v) contrast with $\mathbb{E}_t^{\mathbb{Q}}[RV_{t+21}]$ from VIXN; (vi) invert the affine \mathbb{Q} mapping to recover an implied price of variance risk λ_t ; and (vii) document persistence and predictive regression evidence for the VRP.

1.3 Scholarly Context and Novel Contribution

Course projects in this sequence have repeatedly found that GARCH-type models (often with Student- t innovations) are difficult to beat in-sample on likelihood. That pattern matches the broader message that standard volatility models often forecast realized variation surprisingly well [3]. We replicate it on QQQ. The contribution here is not a marginal likelihood gain, but the use of the POMP variance state as a **structural** \mathbb{P} input to a VRP calculation that is aligned with the affine stochastic volatility logic underlying standard \mathbb{Q} -side representations. GARCH conditional variance

is a deterministic function of past returns; the POMP treats variance as a latent Markov state with its own innovations, which is the object needed for a clean economic translation to λ [4].

2 Data Acquisition and Exploratory Analysis

2.1 Physical Measure (\mathbb{P}) Data: QQQ Returns and Realized Variance

Daily adjusted closing prices for the Invesco QQQ Trust ETF are downloaded from Yahoo Finance (`yfinance`) from April 2019 through April 2026. Log returns are $r_t = \log(P_t/P_{t-1})$. For GARCH model comparison and AIC selection we use the first 1,255 trading days as the **training** sample; the final 504 days are reserved for out-of-sample evaluation of the selected EGARCH specification. POMP estimation and VRP construction follow the notebooks on the full sample once parameters are fixed.

Realized variance is built from squared daily returns aggregated over 21 trading days, in the spirit of realized-volatility construction from return increments [5]. The **backward** sum $RV_t^{\text{back}} = \sum_{s=t-20}^t r_s^2$ matches the information set at t ; the **forward** sum $RV_t^{\text{fwd}} = \sum_{s=t+1}^{t+21} r_s^2$ aligns temporally with the horizon priced by VXN. When annualized as fractional variance, we apply the factor 252/21.

Summary moments and formal tests on the training window (heavy tails, volatility clustering, departures from Gaussianity) motivate both GARCH and stochastic volatility; stationarity of the log-price level is assessed with standard unit-root / stationarity tests [6], while departures from Gaussian returns use normality diagnostics [7]. The exploratory figures below visualize prices, returns, and second-moment persistence.

2.2 Risk-Neutral Measure (\mathbb{Q}) Data: The VXN Volatility Index

The CBOE Nasdaq-100 Volatility Index (VXN) is a model-free, risk-neutral forward-looking measure of annualized volatility in percentage points, analogous in construction and interpretation to better-known volatility benchmarks discussed in the index-options literature [8]. To express it as annualized **fractional variance** on the same scale as our \mathbb{P} forecasts, we use

$$\text{eq_v}_t = \frac{\text{VXN}_t^2}{100^2},$$

so that eq_v_t corresponds to $\mathbb{E}_t^{\mathbb{Q}}[\overline{\sigma}_{t \rightarrow t+21}^2]$ in annualized variance units.

Figure 1 plots the VXN series over the sample with stress intervals highlighted in the notebook. The figure situates subsequent VRP discussion relative to major volatility spikes.

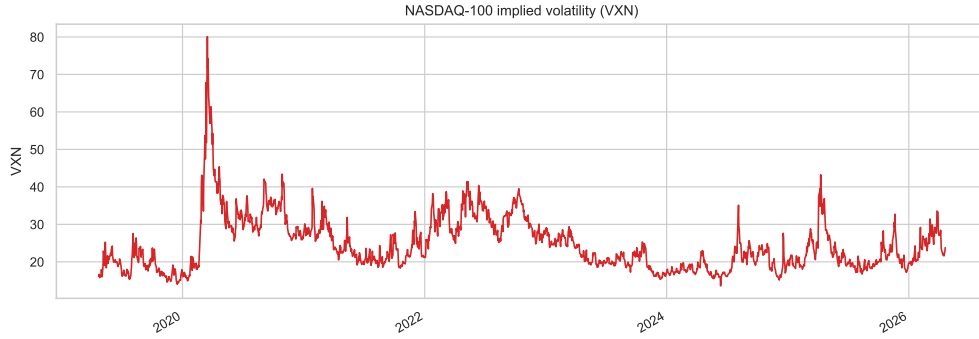


Figure 1: VXN index (annualized %, full sample) with stress episodes shaded.

2.3 Exploratory Data Analysis and Visualizing the VRP Motivation

Figure 2 shows QQQ levels and log returns. Figure 3 contrasts the empirical return density with a Gaussian benchmark and provides a Student- t QQ plot. Figure 4 reports autocorrelations for r_t , r_t^2 , and $|r_t|$; portmanteau tests on squares are the standard way to document volatility clustering [9]. Squared and absolute returns display the slow decay characteristic of long-memory features in financial volatility [10]. Together these panels document thick tails and persistent second moments—the stylized facts that motivate the specifications in Sections 3–4.

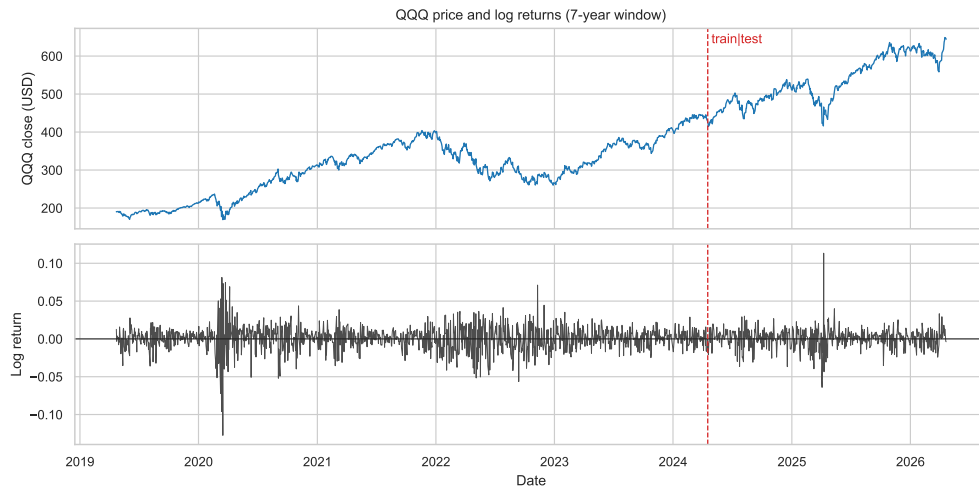


Figure 2: QQQ adjusted close and daily log returns.

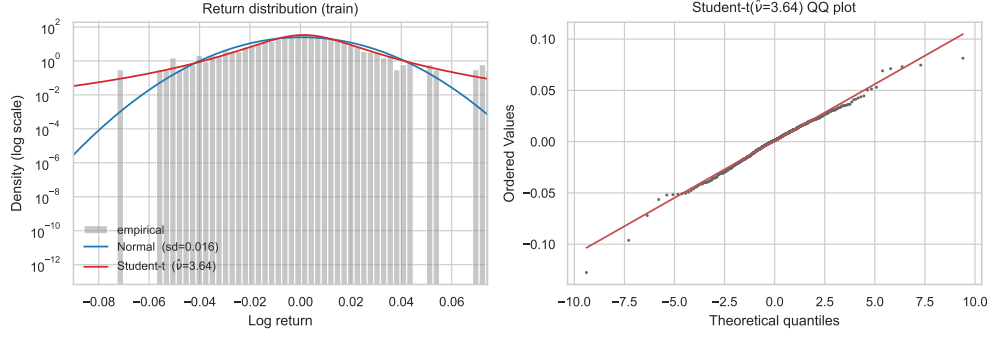


Figure 3: Kernel density of log returns (log scale for frequency) and QQ plot against Student- t .

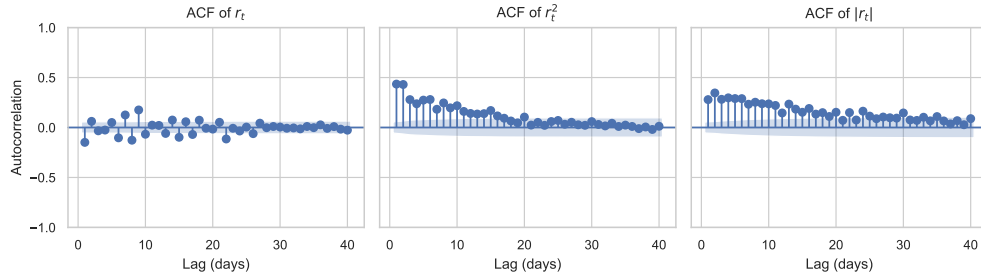


Figure 4: Sample ACFs (40 lags): r_t , r_t^2 , and $|r_t|$.

3 Benchmark Modeling under the Physical Measure

We estimate a grid of GARCH-family models by maximum likelihood (`arch` package) on the training sample. The mean equation is $r_t = \mu + \epsilon_t$ with $\epsilon_t = \sigma_t z_t$ and z_t i.i.d. standard Normal, Student- t , or skewed- t as indicated for each specification.

3.1 EGARCH(2,1,1) with Normal Innovations

Among EGARCH, GJR, and plain GARCH variants with orders (p, o, q) as in the notebook grid, **EGARCH(2,1,1) with Normal innovations** minimizes AIC. The variance recursion is

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i (|z_{t-i}| - \mathbb{E}|z_{t-i}|) + \sum_{j=1}^o \gamma_j z_{t-j} + \sum_{k=1}^q \beta_k \log \sigma_{t-k}^2,$$

with $z_t = \epsilon_t / \sigma_t$. For the selected model, $(p, o, q) = (2, 1, 1)$. The leverage terms γ_j allow negative returns to impact future variance asymmetrically.

Figure 5 ranks training-set AIC across the top specifications. Table 1 reports maximum-likelihood estimates exported from the fitting notebook.

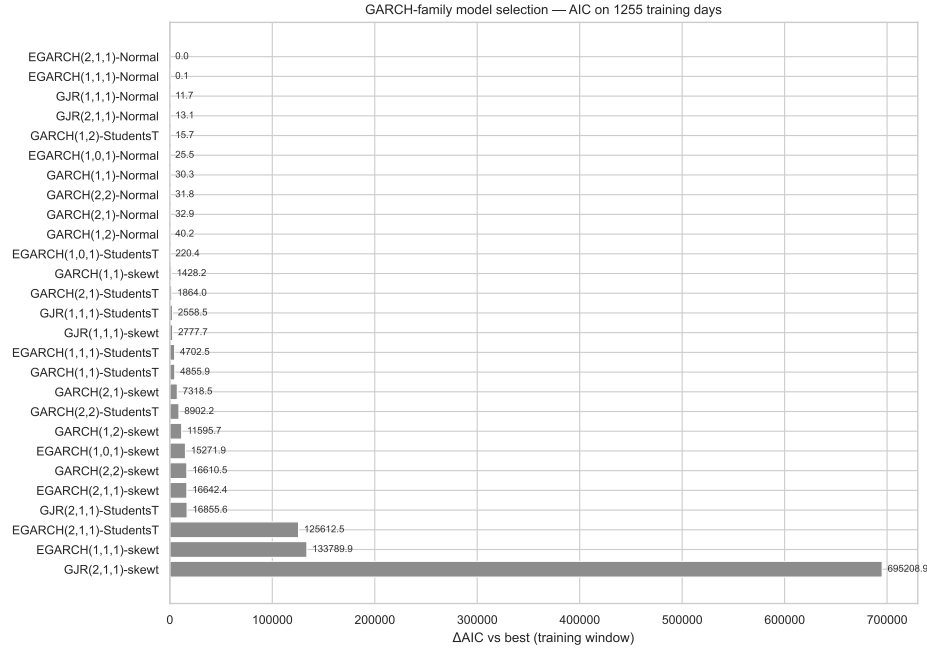


Figure 5: Training-set AIC for leading GARCH-family candidates (EGARCH(2,1,1)-Normal at the top).

Table 1: EGARCH(2,1,1)-Normal: maximum-likelihood estimates (training sample).

Table 1

| Parameter | Estimate | Std. error | p-value |
|-----------|-----------|------------|-----------|
| mu | 0.0005496 | 0.0001971 | 0.005292 |
| omega | -0.3111 | 0.0894 | 0.0005007 |
| alpha[1] | 0.1184 | 0.07651 | 0.1218 |
| alpha[2] | 0.1008 | 0.07697 | 0.1905 |
| gamma[1] | -0.1021 | 0.03197 | 0.001406 |
| beta[1] | 0.9632 | 0.01041 | < 1e-4 |

Figure 6 overlays the model-implied 21-day-ahead annualized volatility with trailing realized volatility for visual comparison. Figure 7 summarizes standardized-residual diagnostics.

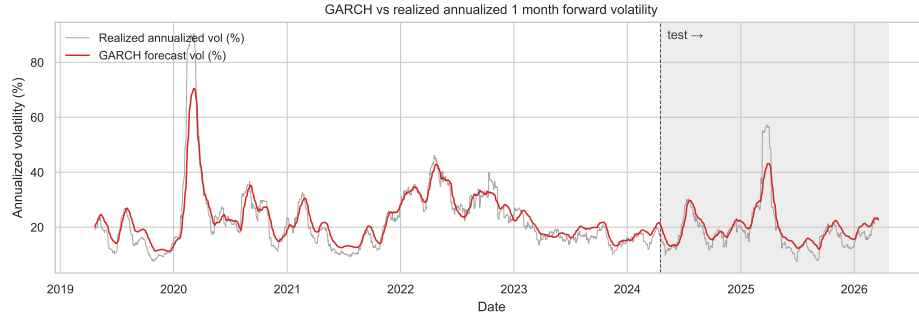


Figure 6: EGARCH(2,1,1)-Normal: 21-day-ahead annualized volatility vs trailing 21-day realized annualized volatility.

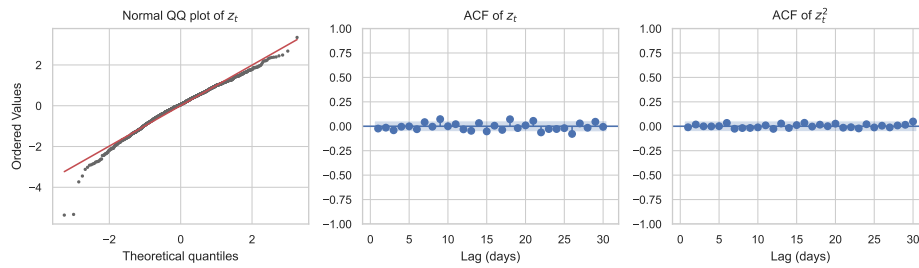


Figure 7: Standardized residual normal QQ plot and ACFs for z_t and z_t^2 .

3.2 Summary of Benchmark Model Performance and Fit

The held-out 504-day segment is used for out-of-sample comparison of 21-day variance forecasts to forward realized variance and for a Mincer–Zarnowitz style regression in the notebook, following the forecast-evaluation perspective emphasized in the volatility-forecasting literature [3]. Figure 8 reproduces the corresponding figure; Table 2 summarizes scalar OOS and MZ statistics in compact form.

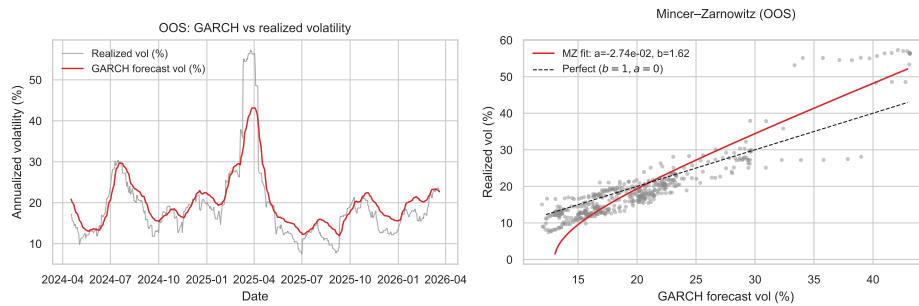


Figure 8: Out-of-sample EGARCH evaluation: forecast vs realized variance and MZ scatter.

Table 2: EGARCH(2,1,1)-Normal: out-of-sample likelihood, loss, and Mincer–Zarnowitz regression summary.

Table 2

| Quantity | Value |
|-------------------------------|-----------|
| Training days | 1255 |
| Test days | 504 |
| Log-likelihood (train, total) | 3636.82 |
| Log-likelihood (test, total) | 1527.82 |
| Mean log-likelihood (train) | 2.8979 |
| Mean log-likelihood (test) | 3.0314 |
| QLIKE (test) | -7.9006 |
| MSE (test, ann. variance) | 3.80e-07 |
| MZ intercept (a-hat) | -5.65e-05 |
| MZ slope (b-hat) | 1.315 |
| SE(b-hat) | 0.494 |
| t-statistic for H0: b = 1 | 0.637 |
| MZ R-squared | 0.105 |

4 Stochastic Volatility via Partially Observed Markov Processes (POMP)

4.1 POMP Model Specification and Latent State Dynamics

Let Y_t denote daily log returns. Latent variance V_t follows a discrete-time square-root (Heston-type) evolution with mean reversion κ , long-run level θ , volatility-of-volatility ξ , and correlation ρ between return and variance innovations. With Gaussian shocks (dW_t^s, dW_t^v) i.i.d. $\mathcal{N}(0, 1)$, the transition used in `03_heston_pomp.ipynb` is

$$\begin{aligned}
 V_{t+1} &= V_t + \kappa(\theta - V_t) + \xi\sqrt{V_t}(\rho dW_t^s + \sqrt{1 - \rho^2} dW_t^v), \\
 Y_{t+1} &= \mu - \frac{1}{2}V_t + \sqrt{V_t} dW_t^s.
 \end{aligned}$$

The measurement density is Gaussian with mean $\mu - V_t/2$ and variance V_t . The parameter vector is $(\mu, \kappa, \theta, \xi, \rho, V_0)$. The innovation dW_t^s is constructed from the previous observation’s return residual as in the notebook and passed through `covars` for the variance update. This specification nests leverage (typically $\rho < 0$ for equity index data) and delivers a filtered state $\hat{V}_t = \mathbb{E}[V_t | y_{1:t}]$ under \mathbb{P} for use in Section 5.

4.2 Implementation Framework: The `pypomp` Environment

Estimation uses `pypomp`: the usual POMP tuple (`rinit`, `rproc`, `dmeas`, `ParTrans`, `data`) with JAX compilation for `pfilter` and IF2 (`mif`). Parameters are optimized on a transformed scale (log/atanh) to enforce positivity and $\rho \in (-1, 1)$. The iterated filtering and particle-filter likelihood framework follows the partially observed Markov process treatment in the course reference [4].

4.3 Local Search and Convergence Diagnostics using IF2

A local IF2 run from an informative starting value illustrates convergence of the log-likelihood and parameter traces at moderate particle count. Figure 9 displays the trajectories used in the project.

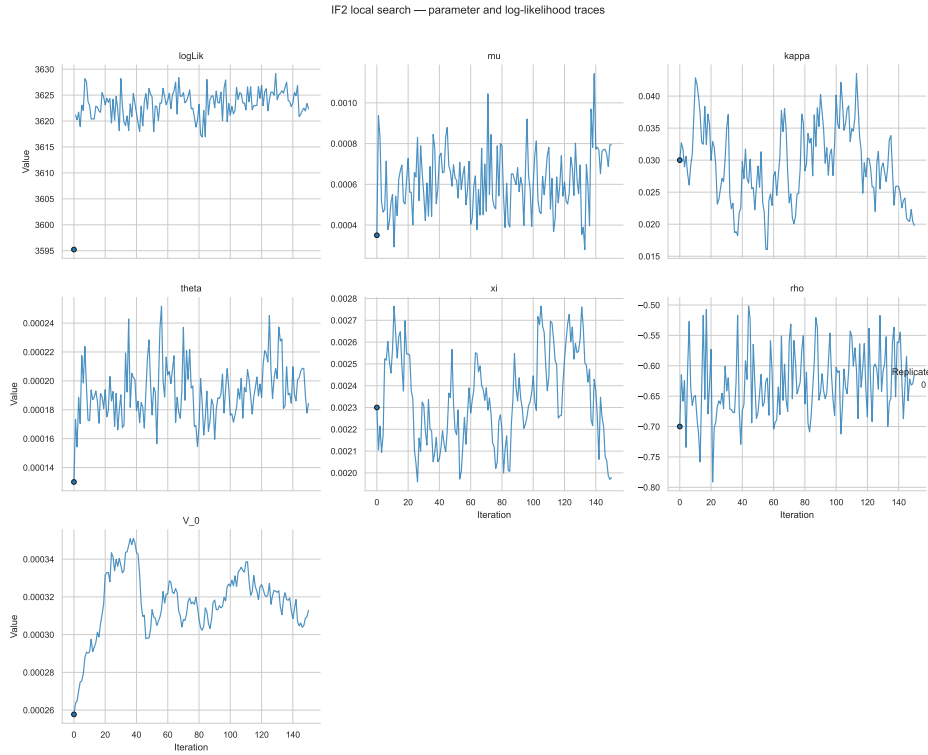


Figure 9: Local IF2: log-likelihood and parameter traces vs iteration.

4.4 Global Search and Parameter Identifiability Analysis

Twelve independent IF2 runs from randomized starting points support the claim that the best-found mode is not an artifact of a single initialization. Figures 10–12 visualize multi-start trajectories, endpoint scatter, and endpoint log-likelihood against each parameter. Table 3 lists terminal values sorted by likelihood (selected scale for μ , θ , ξ keeps the table within the text block).

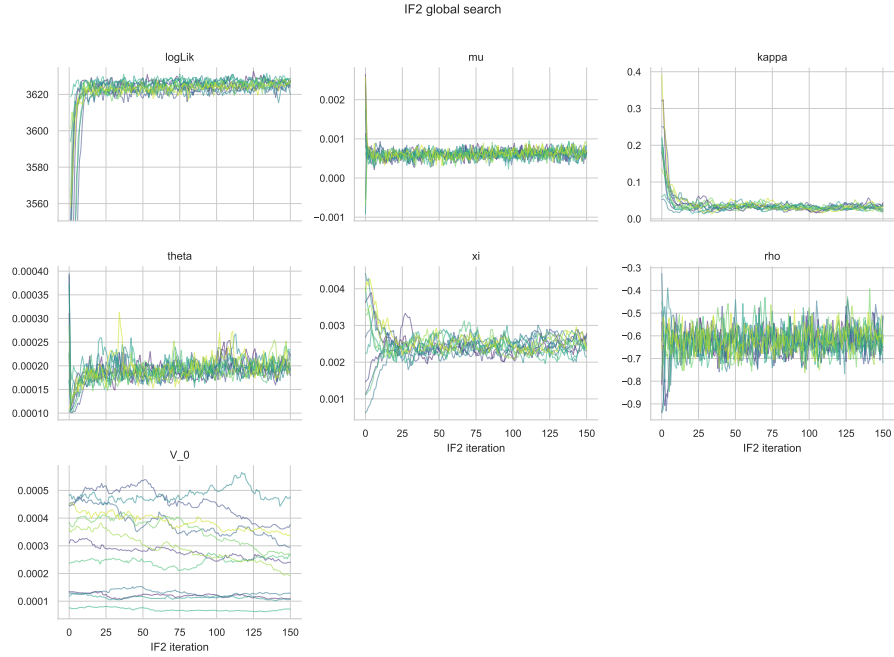


Figure 10: Global IF2 trajectories from twelve random starts (best replicate emphasized in notebook styling).

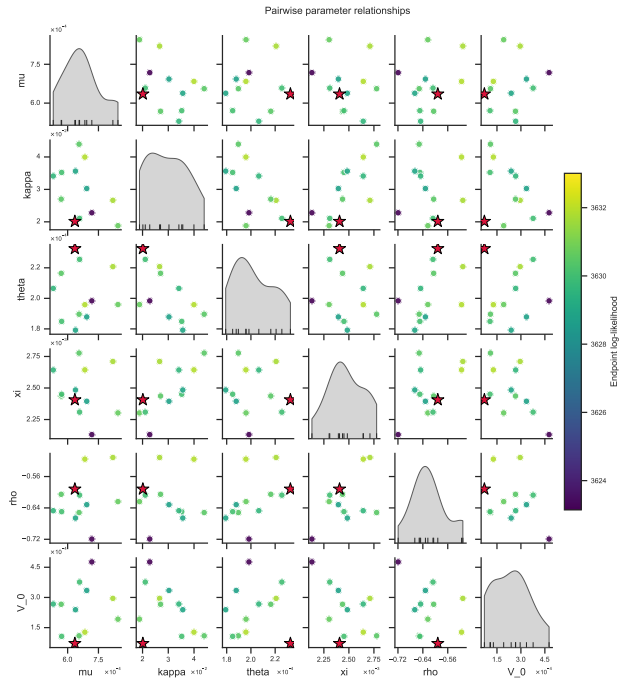


Figure 11: Pairwise scatter of global endpoints; diagonal KDEs; star marks the MLE replicate.

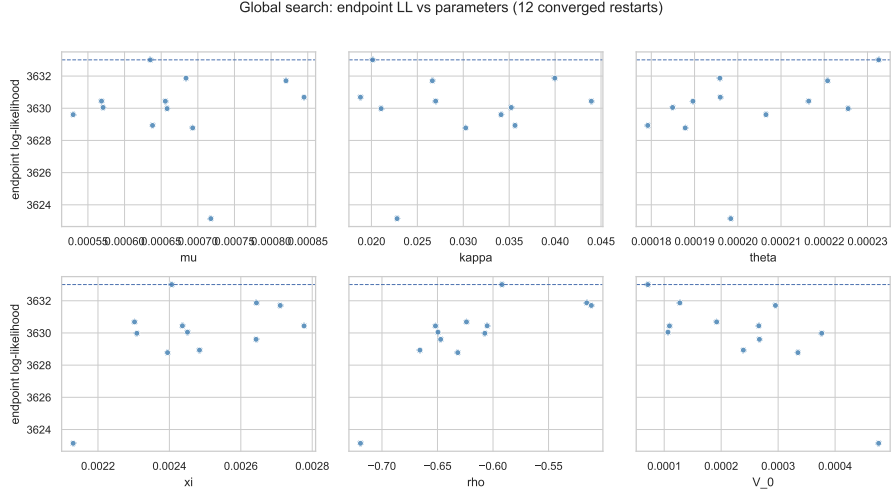


Figure 12: Endpoint log-likelihood vs each parameter across starts (identifiability diagnostic).

Table 3: Global IF2 terminal log-likelihoods and parameters (sorted by LogLik).

Table 3

| start | LogLik | $\times 10$ | $\times 10$ | $\times 10$ | $\times 10^3$ | $\times 10^3$ |
|-------|-------------|-------------|-------------|-------------|---------------|---------------|
| 7 | 3633.010000 | 6.352000 | 0.020100 | 2.326000 | 2.406000 | -0.592000 |
| 4 | 3631.870000 | 6.839000 | 0.040000 | 1.959000 | 2.644000 | -0.516000 |
| 3 | 3631.710000 | 8.203000 | 0.026600 | 2.208000 | 2.710000 | -0.511000 |
| 10 | 3630.690000 | 8.450000 | 0.018800 | 1.960000 | 2.302000 | -0.624000 |
| 8 | 3630.440000 | 5.686000 | 0.027000 | 2.164000 | 2.436000 | -0.605000 |
| 0 | 3630.440000 | 6.556000 | 0.043900 | 1.896000 | 2.777000 | -0.652000 |
| 6 | 3630.050000 | 5.707000 | 0.035200 | 1.849000 | 2.451000 | -0.650000 |
| 2 | 3629.980000 | 6.581000 | 0.021100 | 2.255000 | 2.309000 | -0.607000 |
| 9 | 3629.600000 | 5.299000 | 0.034100 | 2.065000 | 2.643000 | -0.647000 |
| 1 | 3628.930000 | 6.382000 | 0.035600 | 1.792000 | 2.484000 | -0.666000 |
| 11 | 3628.780000 | 6.930000 | 0.030300 | 1.878000 | 2.395000 | -0.632000 |
| 5 | 3623.140000 | 7.177000 | 0.022800 | 1.984000 | 2.131000 | -0.719000 |

Figure 13 compares the particle-filter mean of V_t at the MLE (converted to annualized volatility) to trailing realized volatility. Agreement indicates that the latent state tracks return-based volatility information.

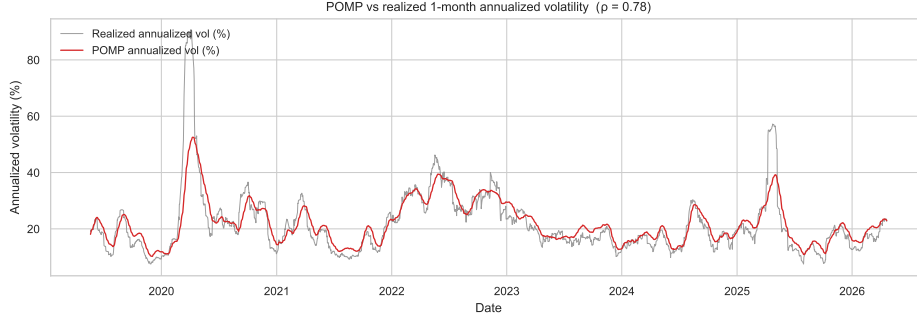


Figure 13: POMP filter mean at MLE vs trailing 21-day realized annualized volatility (full sample).

4.5 Comparative Likelihood Analysis: POMP vs. Benchmarks

Table 4 contrasts training-sample log-likelihood and information criteria for leading GARCH specifications and the POMP estimate (Monte Carlo standard error for the POMP likelihood from repeated `pfilter` runs is recorded in `ll_comparison.csv`).

Table 4: Training-sample likelihood and information criteria: leading GARCH models and POMP Heston.

Table 4

| Model | Log-lik. | AIC | BIC | k |
|---------------------------|-------------|--------------|--------------|---|
| EG(2,1,1)-N | 3636.820000 | -7261.600000 | -7230.800000 | 6 |
| EG(1,1,1)-N | 3635.750000 | -7261.500000 | -7235.800000 | 5 |
| GJR(1,1,1)-N | 3629.970000 | -7249.900000 | -7224.300000 | 5 |
| GJR(2,1,1)-N | 3630.280000 | -7248.600000 | -7217.700000 | 6 |
| GARCH(1,2)-StudentsT | 3628.970000 | -7245.900000 | -7215.100000 | 6 |
| POMP Heston (MLE via IF2) | 3632.850000 | -7253.700000 | nan | 6 |

The POMP likelihood is competitive but does not dominate the best EGARCH specification on AIC; Section 6 discusses why the POMP remains the appropriate object for the VRP exercise despite this ranking.

5 Extracting the Variance Risk Premium

5.1 Conditional Variance Forecasting under the Physical Measure (\mathbb{P})

Given filtered expectations \hat{V}_t and estimated (κ, θ) under \mathbb{P} , the conditional mean of integrated variance over the next $h = 21$ trading days follows from the affine structure. The annualized variance forecast used in the notebooks is

$$\mathbb{E}_t^{\mathbb{P}}[RV_{t+21}] = \frac{252}{21} \left[21\theta + (\hat{V}_t - \theta) \frac{1 - e^{-21\kappa}}{\kappa} \right].$$

5.2 Deriving Market Expectations from the Risk-Neutral Measure (Q)

Risk-neutral expected variance over the forward window is identified with eq_v_t from VXN as in Section 2.2. This construction is model-free on the \mathbb{Q} side and matches the units of the \mathbb{P} forecast above.

5.3 The Extracted VRP Time Series and Macroeconomic Correlates

Define the variance risk premium as

$$VRP_t = eq_v_t - \mathbb{E}_t^{\mathbb{P}}[RV_{t+21}].$$

Figure 14 plots \mathbb{Q} and \mathbb{P} forecasts in annualized volatility units ($100 \times \sqrt{\text{variance}}$), overlays trailing realized volatility, and shades regions where VRP_t is positive versus negative. Table 5 summarizes mean VRP and the fraction of negative-VRP days over coarse macroeconomic windows (values exported from `vrp.parquet` into `artifacts/paper_tables/regime_summary.csv` by `scripts/export_paper_tables.py`).

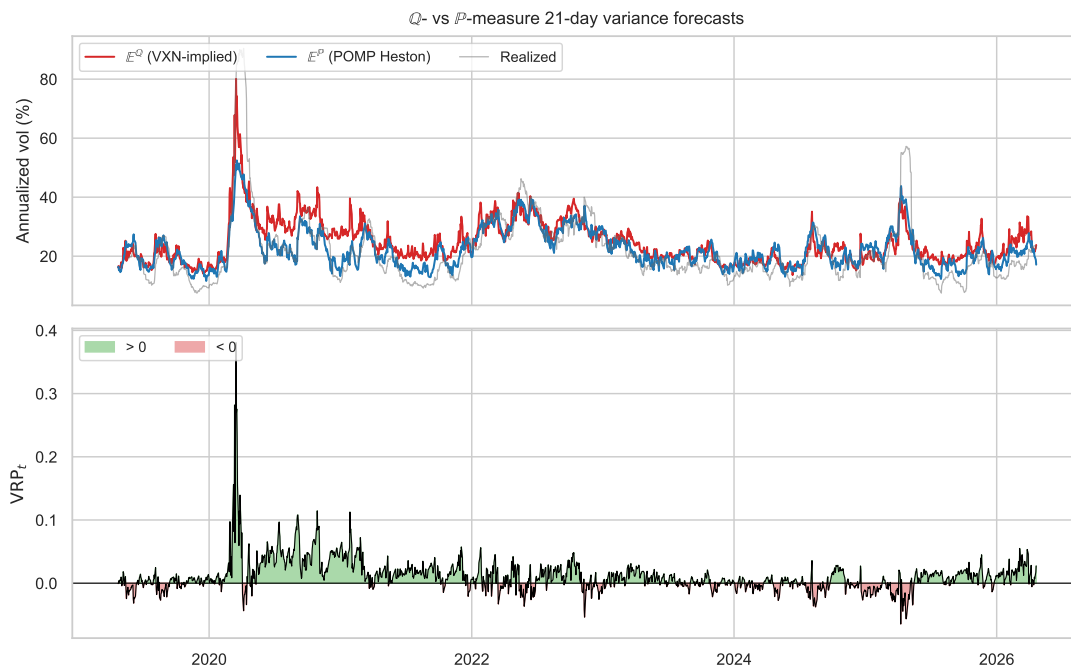


Figure 14: VXN-implied vs POMP \mathbb{P} forecast (top); VRP with sign shading (bottom).

Table 5: VRP summary statistics by regime.

Table 5

| Regime | n | Mean VRP | Pct. days VRP < 0 |
|-----------------------------|------|-----------|-------------------|
| COVID (Feb–Apr 2020) | 50 | 0.064300 | 26.000000 |
| Fed tightening (2022) | 251 | 0.006600 | 37.100000 |
| SVB (Mar 2023) | 23 | 0.015400 | 4.300000 |
| Tariff shock (Apr–May 2025) | 42 | -0.012700 | 57.100000 |
| Other days | 1393 | 0.011100 | 28.100000 |

5.4 Mapping VRP to the Economic Price of Risk (λ)

Under the affine change-of-measure used in the empirical VRP literature (see `04_vrp.ipynb` for the exact sign convention relative to the textbook Heston drift), the variance risk price λ shifts the drift of V so that $\kappa^{\mathbb{Q}} = \kappa^{\mathbb{P}} - \lambda$ and long-run \mathbb{Q} -variance is re-scaled accordingly. Holding $(\kappa^{\mathbb{P}}, \theta^{\mathbb{P}}, \hat{V}_t)$ fixed, we solve numerically for λ_t such that the closed-form 21-day \mathbb{Q} variance forecast equals `eq_v_t`. After running `04_vrp.ipynb`, include `artifacts/fig/fig_lambda.pdf` here for the implied λ_t path and scatter against \hat{V}_t .

5.5 Statistical Persistence and Predictive Regressions

Autocorrelation structure for VRP_t is reported in Figure 15; formal dependence diagnostics again follow the portmanteau principle [9]. Predictive regressions for forward realized variance use Newey–West standard errors with 21 lags in the notebook to accommodate overlapping windows [2]. Table 6 shows fit statistics.

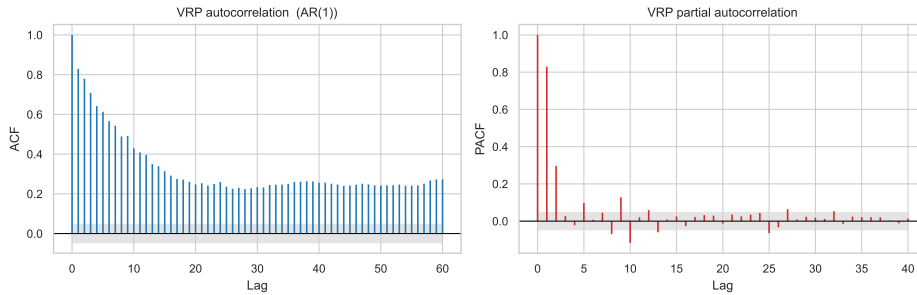


Figure 15: ACF and PACF of the VRP series (lags to 60).

Table 6: Predictive regressions: goodness of fit by specification.

Table 6

| Specification | R ² | Adj. R ² | n |
|------------------------------|----------------|---------------------|------|
| Q only (eq_v) | 0.195000 | 0.195000 | 1738 |
| P only (V21_P) | 0.156000 | 0.156000 | 1738 |
| VRP only | 0.068000 | 0.067000 | 1738 |
| MZ + VRP (eq_v + VRP) | 0.198000 | 0.197000 | 1738 |
| P + Q jointly (eq_v + V21_P) | 0.198000 | 0.197000 | 1738 |

The regression evidence indicates that VXN already encodes much of the predictable variation in forward realized variance at this horizon; incremental explanatory power from the raw VRP or from jointly including \mathbb{P} and \mathbb{Q} forecasts is limited once robust standard errors are used—consistent with overlapping windows and slow-moving implied volatility.

6 Discussion and Comparative Synthesis

6.1 Structural Advantages of POMP for VRP Estimation

Even when EGARCH achieves a slightly better in-sample AIC, the POMP supplies a latent variance process with stochastic innovations that maps directly into the affine \mathbb{Q} -variance algebra used to define and invert λ_t . That mapping is conceptually awkward under deterministic GARCH variance. For the economic question posed in the introduction, the POMP is therefore the more coherent statistical object, notwithstanding the likelihood ranking.

6.2 Comparison with Prior STATS 531 Projects

Earlier final projects, such as the Bitcoin (W25) and the gold (W25) projects, comparing POMP Heston to GARCH on individual names or indices typically conclude that GARCH is hard to beat descriptively. Our results align with that pattern on QQQ. The present analysis extends the comparison by using the POMP filter output as an input to a VRP decomposition rather than stopping at the likelihood table.

6.3 Model Limitations and Future Extensions

The single-factor Heston specification does not capture jump risk in returns or volatility that the VXN may embed; richer \mathbb{Q} -side models, Student- t measurement errors, or fully Bayesian treatment of filtering uncertainty for λ_t would be natural extensions.

6.4 AI Usage

We used AI in synthesizing our findings into this paper in a formalized format. The findings and figure generation were done on our own but the placements and wordings in the paper were edited and reviewed by AI.

7 Conclusions

We document that a POMP-based Heston stochastic volatility model estimated on QQQ returns yields a \mathbb{P} variance forecast that is statistically close to a strong EGARCH benchmark on likelihood, yet remains the appropriate building block for a VRP series and implied λ_t constructed relative to VXN. Exploratory and formal time-series evidence characterize the VRP's persistence and its limited incremental forecasting content for forward realized variance once VXN is included, which underscores interpreting the VRP as a **decomposition** of market-implied variance rather than as a standalone forecasting factor [5], [8].

References

- [1] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications: With R examples*, 4th ed. Springer, 2017.
- [2] J. D. Hamilton, *Time series analysis*. Princeton University Press, 1994.
- [3] T. G. Andersen and T. Bollerslev, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 1998.
- [4] E. L. Ionides, "STATS 531 (winter 2026) course notes: Analysis of time series." <https://ionides.github.io/531w26/>, 2026.
- [5] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, "Modeling and forecasting realized volatility," *Econometrica*, 2003.
- [6] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 1992.
- [7] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review*, 1987.
- [8] R. E. Whaley, "The investor fear gauge," *The Journal of Portfolio Management*, 2000.
- [9] G. M. Ljung and G. E. Box, "On a measure of lack of fit in time series models," *Biometrika*, 1978.
- [10] C. W. Granger and Z. Ding, "Varieties of long memory models," *Journal of Econometrics*, 1996.