Stochastic Volatility Modeling of the NASDAQ Composite

Abstract:

This paper develops a stochastic-volatility model for daily NASDAQ Composite returns using two observable signals: close-to-close log returns and a high-low range-based volatility proxy. Estimation is carried out with particle filtering and iterated filtering, and model selection proceeds through a screening-and-refinement search over multiple random starts. Several richer alternatives were explored during development, including a dynamic return mean, a dynamic range-bias process, and a leverage state, but these additions typically weakened convergence or produced poorly identified parameter paths. The final specification therefore combines a parsimonious latent log-volatility process with a static range-bias correction and heavy-tailed measurement errors. It was selected because it delivered strong held-out likelihood, stable convergence behavior, and acceptable residual diagnostics. The results suggest that range information materially improves volatility inference for the NASDAQ, while additional latent structure is not well supported by the data in this setting.

1 Introduction

This project models daily NASDAQ Composite volatility with a partially observed stochastic-volatility model in the broad stochastic-volatility tradition (Taylor 1982). The main objective is to estimate a latent log-volatility process that explains both close-to-close returns and intraday volatility information from daily high-low ranges.

The final model is deliberately parsimonious. Richer alternatives, including dynamic mean terms, dynamic range bias, leverage effects, and more flexible volatility dynamics, were retained only if they improved held-out likelihood without weakening convergence. The selected specification uses one latent volatility state, a static range-bias correction, and heavy-tailed observation errors.

2 Data

The data Yahoo Finance (2026) are daily NASDAQ Composite open, high, low, close, and volume observations from January 2006 through January 2026. The model uses two observed series. The first is the daily log return,

$$r_t = \log(C_t) - \log(C_{t-1}),$$

where C_t is the closing price. The second is the log Parkinson high-low volatility proxy (Parkinson 1980),

$$q_t = \log\left(\frac{(\log H_t - \log L_t)^2}{4\log 2}\right),$$

where H_t and L_t are the daily high and low prices. The high-low range gives direct information about intraday volatility, which close-to-close returns alone do not provide.

Observations through December 31, 2020 are used for training. Later observations are held out for model evaluation.

	sample	n_obs	return_mean	return_sd	range_mean
0	training	3775	0.000463	0.013809	-9.994886
1	held-out	1255	0.000470	0.014269	-9.650873
2	total	5030	0.000465	0.013923	-9.909054

The return means are small relative to their standard deviations, which supports using a simple conditional mean and focusing the model on volatility. The range proxy is on a log-variance scale, so its level is not directly comparable to returns; instead, it serves as a noisy measurement of latent log volatility.

3 Exploratory Data Analysis

The first exploratory plot shows daily log returns. The series is centered near zero, but its variance is clearly time-varying, with calm periods alternating with clusters of large positive and negative returns. This is the main empirical motivation for a stochastic-volatility model: the conditional variance evolves over time while the conditional mean remains small.

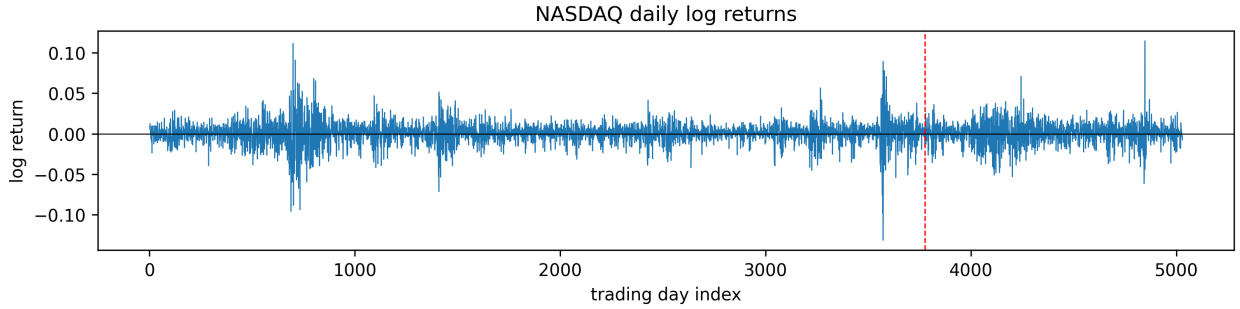

Figure 1: Daily NASDAQ log returns. The dashed red line marks the training/test split.

The return plot supports a volatility model with a simple conditional mean. Returns fluctuate around zero with no persistent trend, while their magnitude changes sharply over time. This clustering, together with occasional large moves, motivates a latent stochastic-volatility process with heavy-tailed measurement errors.

The next plot shows the log Parkinson range variance. This series is much more directly tied to daily volatility than the raw return series. It has persistent high-volatility episodes, which indicates that the range proxy contains serial dependence that the latent volatility process should explain.

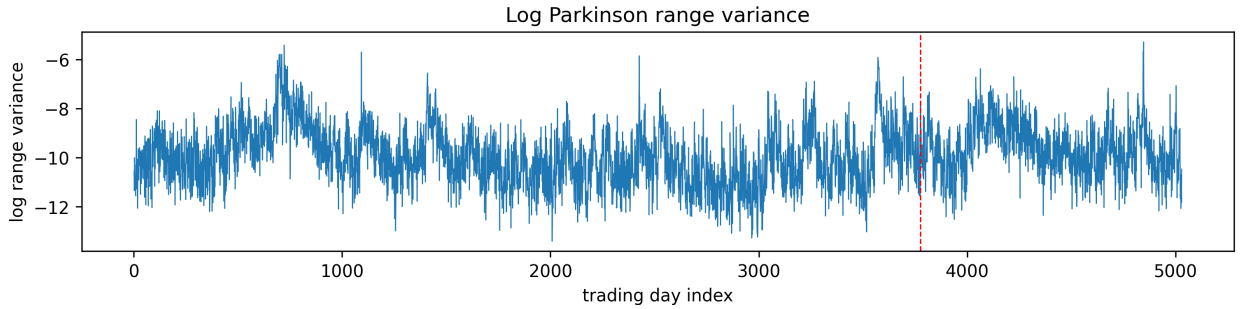

Figure 2: Log Parkinson range variance. The dashed red line marks the training/test split.

The range proxy plot provides a clearer view of volatility persistence than the return series. High values of log range variance cluster over time, indicating that volatility has memory, precisely the dependence the latent state is designed to capture. The plot also shows why the high-low range is useful: volatility can be elevated even when close-to-close returns are modest. Including the range proxy as a second measurement equation therefore adds meaningful information about the latent volatility path.

This persistence also defines a diagnostic target. After fitting, standardized range-measurement residuals should exhibit little serial correlation; otherwise, the model has failed to capture the observed dependence.

4 Model Specification

The latent state is daily log volatility, denoted h_t. The transition equation is

$$h_t = \mu_h + \phi(h_{t-1} - \mu_h) + \sigma_h \eta_t, \qquad \eta_t \sim N(0,1).$$

The initial state is fixed at the long-run mean,

$$h_0 = \mu_h,$$

which avoids estimating a weakly identified initial-state parameter.

The return equation is

$$r_t \mid h_t \sim t_7\left(\mu_0, \exp(h_t/2)\right),$$

and the range-volatility measurement equation is

$$q_t \mid h_t \sim t_5\left(h_t + b, \sigma_q\right).$$

The parameter b is a static range-bias correction. It accounts for the systematic difference between the Parkinson range proxy and the latent log volatility scale, an issue that became clear both in exploratory analysis and in the course case study on volatility modeling (E. Ionides 2026b).

Parameter	Description
μ_0	Constant location of daily log returns
μ_h	Long-term mean of latent log volatility
ϕ	Volatility persistence coefficient
σ_h	Innovation standard deviation of the latent log-volatility process
b	Static range-bias correction between the range proxy and latent log volatility
σ_q	Measurement-error scale for the log Parkinson range-volatility proxy
h_0	Initial latent log volatility, fixed at μ_h

The persistence and volatility innovation scale are regularized by smooth transformations:

$$\phi = 0.75 + 0.249\,\mathrm{logit}^{-1}(\phi_{raw}), \qquad \sigma_h = 0.01 + 0.49\,\mathrm{logit}^{-1}(\sigma_{h,raw}).$$

Thus $0.75 < \phi < 0.999$ and $0.01 < \sigma_h < 0.50$. These bounds keep the latent volatility process from becoming too jumpy and absorbing short-term measurement noise.

5 Estimation Procedure

The model is estimated as a partially observed Markov process using particle filtering and iterated filtering (Bréto et al. 2009; E. L. Ionides et al. 2015; E. Ionides 2026a). Because latent volatility is not observed directly, the likelihood is approximated numerically by propagating particles through the volatility transition equation and weighting them by the joint return-range measurement density.

For a fixed parameter vector, the particle filter produces a Monte Carlo likelihood estimate, while iterated filtering perturbs parameters across repeated filtering passes to search for high-likelihood regions. To reduce sensitivity to starting values, estimation proceeds in two stages: a broad screening step evaluates many random parameter draws with a relatively cheap particle filter, and a refinement step applies iterated filtering to the best screened candidates using larger particle counts. The final candidates are then reevaluated with more particles and replicated likelihood estimates, and the best training fit is assessed on the held-out sample.

The main computational settings are summarized below.

	setting	value
0	NREPS_GLOBAL_SCREEN	75
1	NKEEP_GLOBAL	8
2	NP_SCREEN	400
3	NP_MIF_GLOBAL	1000
4	NP_EVAL_GLOBAL	2000
5	NFITR_GLOBAL	100
6	NREPS_EVAL	5

The global search boxes define the parameter region explored during screening while excluding clearly degenerate regions found in earlier experiments.

	parameter	lower	upper
0	mu0	-0.000537	0.001463
1	mu_h	-12.000000	-6.500000
2	phi_raw	-4.000000	5.500000
3	sigma_h_raw	-4.000000	4.000000
4	range_bias_mean	-3.429955	0.570045
5	log_sigma_range_meas	-1.897120	0.916291

The initial screening step samples `NREPS_GLOBAL_SCREEN` candidate parameter vectors from these boxes and ranks them using a low-cost particle filter. Only the best candidates are retained for refinement. Full screening results are reported in Appendix Section .

```
Global screening runtime (sec): 1.951
```

The second stage refines the retained candidates with iterated filtering and records parameter paths after each iteration for convergence assessment. After refinement, the candidates are reevaluated with larger particle counts and replicated likelihood estimates to obtain final training log likelihoods and Monte Carlo standard errors. The best training candidate is then evaluated on the held-out sample. Full refined results are reported in Appendix Section .

```
Global refinement runtime (sec): 82.692
```

The convergence figure should be interpreted cautiously because each line is based on particle-filter likelihood estimates. Some short-run variation is expected. What matters is whether the parameter paths stabilize into plausible regions and whether the likelihood traces avoid catastrophic collapses. The final model was selected partly because its convergence behavior was more stable than richer alternatives such as dynamic range bias or latent leverage models.

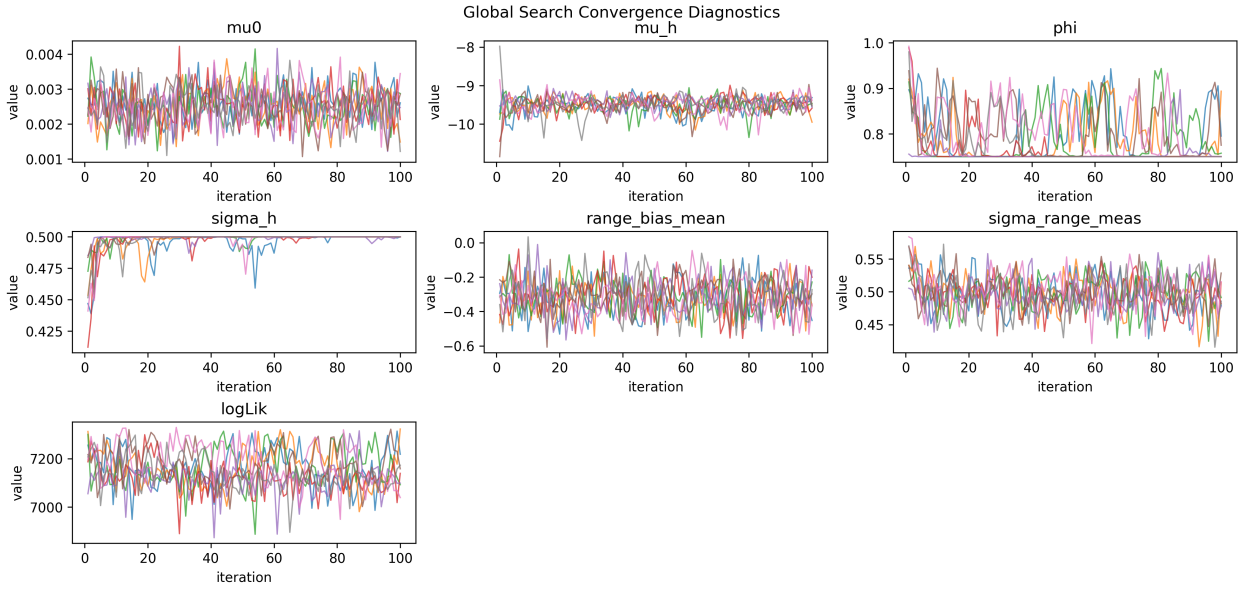

Figure 3: Global search convergence diagnostics for the final model.

6 Analysis and Model Comparison

The final model was chosen by combining three criteria: held-out log likelihood, convergence stability, and residual diagnostics. Held-out likelihood measures predictive performance, while convergence and diagnostics guard against selecting a model that fits well only because of weak identification or numerical degeneracy.

6.1 Final Model Performance

The tables below summarize the refined training candidates and the held-out evaluation for the selected model. Parameters are shown on transformed, interpretable scales with slightly shorter labels to keep the full summary on the page.

logLik	SE	mu0	mu_h	phi	sigma_h	bias	sigma_q
7326.39	0.946	0.0015	-9.951	0.894	0.5	-0.367	0.514
7225.2	1.889	0.0026	-9.311	0.795	0.5	-0.452	0.479
7184.57	0.692	0.0012	-9.529	0.75	0.5	-0.273	0.486
7160.47	2.067	0.0029	-9.528	0.775	0.5	-0.298	0.529
7157.76	0.469	0.0021	-9.586	0.75	0.5	-0.201	0.482

The training results are interpreted together with the final-model convergence figure from the estimation section. The leading candidates agree on the main structural features: a stable long-run volatility level, a static range-bias correction, and a plausible range-measurement scale.

logLik	SE	mu0	mu_h	phi	sigma_h	bias	sigma_q
2237.2	0.196	0.0015	-9.951	0.894	0.5	-0.367	0.514

Held-out likelihood evaluates the model on data not used for estimation and is the main safeguard against overfitting. Earlier variants added flexibility, but they often led to weaker convergence or poorer held-out performance. The final model preserves the gain from range-based volatility information without reintroducing weakly identified dynamic bias or leverage components.

The static range-bias term is a key result. The Parkinson proxy is systematically offset from the latent volatility scale, requiring a correction, but the data do not support a dynamic specification. This indicates that the high–low range is informative about volatility, with its relationship to latent volatility adequately captured by a constant offset.

The convergence plot should be interpreted parameter by parameter. Stable horizontal bands across starting values indicate convergence to a common region of the likelihood surface, while drift, boundary behavior, or likelihood collapses suggest weak identification. In the final model, key parameters exhibit greater stability than in earlier specifications. Although some Monte Carlo noise remains due to particle filtering, the traces avoid the severe instability observed in more complex alternatives.

6.2 Comparison With Alternatives

Several richer or more flexible specifications were tested and rejected. A latent return-mean process was weakly identified, consistent with the STL decomposition showing little persistent trend in daily returns. A dynamic range-bias process also failed to justify itself, with innovation variance collapsing toward zero and persistence behaving erratically. Allowing a free initial volatility state improved speed but worsened held-out likelihood, suggesting that it mainly fit the start of the training sample rather than improving generalization. The final model therefore fixes $h_0 = \mu_h$.

6.2.1 Leverage Candidate

A richer two-state leverage stochastic-volatility model was also tested, motivated both by the course case study on observation-driven latent volatility dynamics (E. Ionides 2026b) and by the related student project that implemented a similar idea for Apple stock data (**stats531project2024?**). In that specification, a latent leverage state allowed lagged returns to affect next-period volatility, making it possible to capture the common asymmetry in which negative returns are followed by higher volatility.

Despite that motivation, the leverage model did not converge convincingly. The global-search parameter paths often drifted toward boundaries rather than stabilizing in a common interior region, and some likelihood traces showed abrupt collapses. The leverage-related state summaries and innovation parameters behaved erratically, suggesting weak identification rather than a stable estimate of a leverage mechanism.

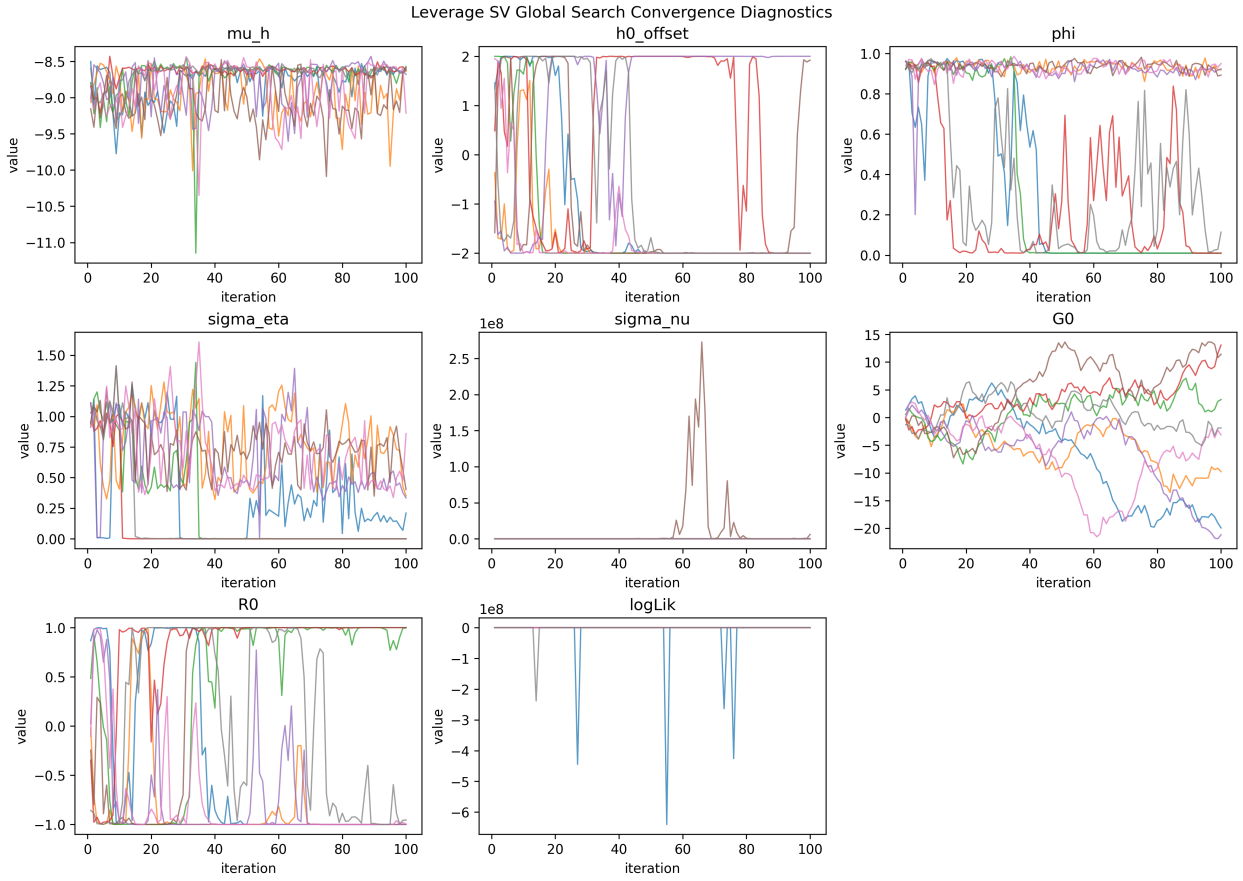

Figure 4: Convergence diagnostics for the candidate leverage stochastic-volatility model.

For the main paper, the leverage model mainly serves as a cautionary comparison: it could sometimes improve held-out likelihood, but the convergence behavior was poor enough that the extra flexibility was difficult to defend. Full leverage result tables are reported in Appendix Section and Appendix Section .

Final Selection Rationale

The model-selection results show that the final specification is not simply the most flexible. Although several extensions—especially the leverage model—were appealing, they introduced weak identification, boundary behavior, or unstable likelihood trajectories. The final model was chosen for its balance of predictive performance, convergence stability, residual behavior, and interpretability. It retains the most informative signal, the high–low range proxy, while excluding unsupported dynamic features.

The leverage comparison highlights the central tradeoff: increased flexibility can improve apparent fit but is not desirable when parameters are weakly identified. A simpler negative-return leverage term was also tested but did not improve held-out likelihood and added convergence noise. Accordingly, leverage effects were excluded. The final model instead uses Student-t measurement errors, a

static range-bias adjustment, bounded volatility persistence, and a range-based measurement equation. While less flexible, this specification is more robust, with clearer parameter interpretation, more stable convergence, and reduced dependence on unstable latent-state interactions.

7 Diagnostics

After fitting the model, the held-out filtered volatility path is estimated with a bootstrap particle filter. The standardized volatility measurement residual is

$$e_t = \frac{q_t - (\hat{h}_t + \hat{b})}{\hat{\sigma}_q}.$$

If the range measurement equation is adequate, these residuals should be centered near zero and show little serial correlation.

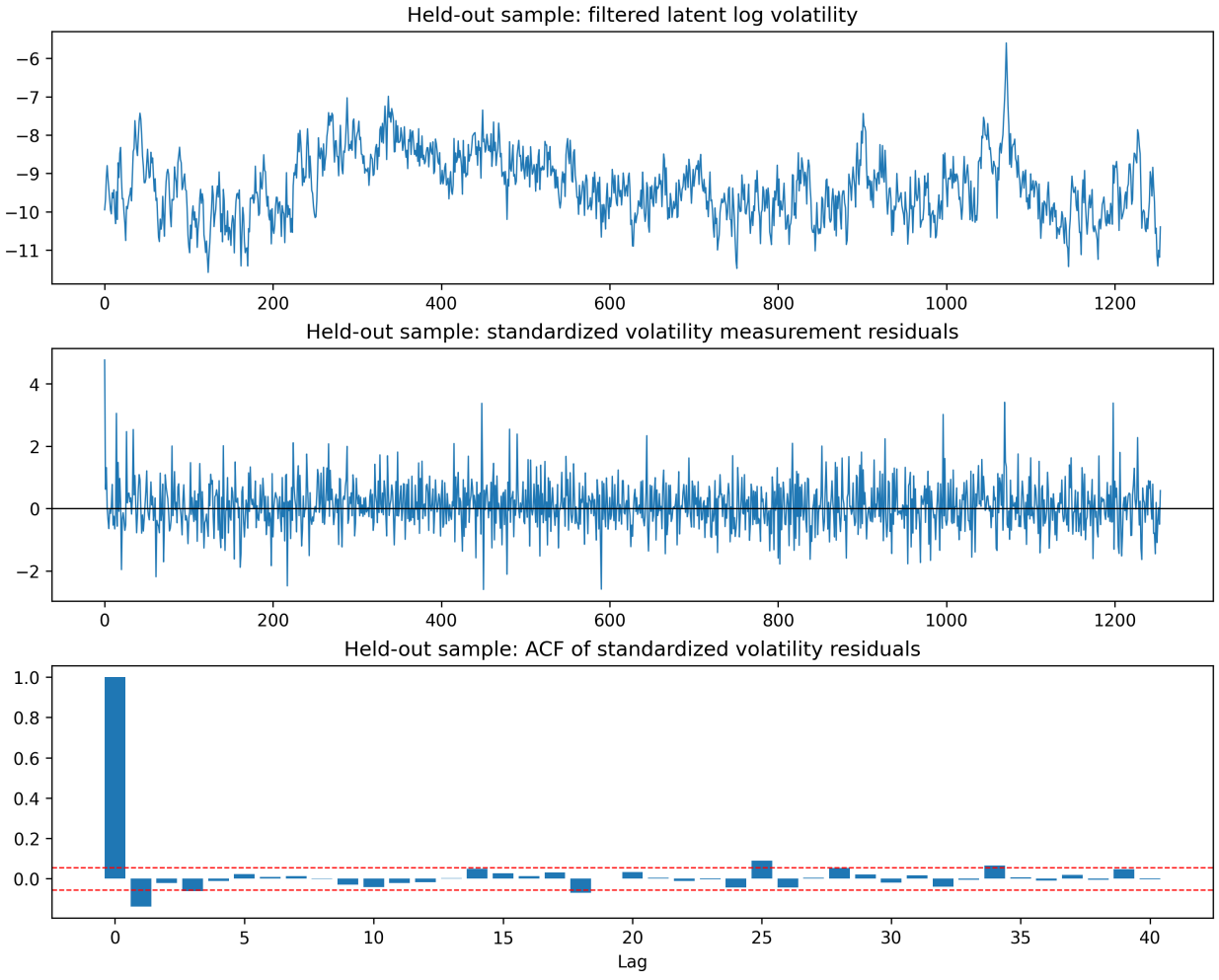

Figure 5: Held-out filtered latent volatility and standardized volatility measurement residual diagnostics.

The diagnostic figure has three main components. The filtered latent log-volatility path shows the model-implied evolution of volatility over time. A useful volatility filter should rise during turbulent market episodes and fall during calmer periods rather than remaining flat or moving erratically without connection to the data. In this model, the filtered path is allowed to change over time, but the regularized transition prevents it from becoming an arbitrary high-frequency fit to the range proxy.

The standardized volatility measurement residuals compare the observed log range variance with the fitted range measurement mean, after scaling by the estimated range measurement error. These residuals should be centered near zero. Persistent runs above or below zero would suggest that the range-bias term or volatility level is misspecified. Isolated large residuals are less concerning because the range equation uses a Student-t distribution, which explicitly allows occasional outlying range observations.

The residual ACF is the most direct check for remaining serial dependence in the volatility measurement equation. The lag-zero bar is always equal to one and is not interpreted. The important question is whether the bars at positive lags remain large or decay slowly. Large positive autocorrelations across many lags would imply that the latent volatility process failed to capture persistent structure in the range proxy. In the final model, the residual autocorrelations are much smaller than in earlier specifications, indicating that the fitted latent volatility state captures most of the serial dependence in the high-low volatility measure. This is one of the main pieces of evidence supporting the final static range-bias specification.

8 Conclusion

The final specification uses the information that was most consistently supported by the data: mean-reverting latent volatility, a static range-bias correction, and heavy-tailed observation errors. More complex components, including dynamic range bias and leverage terms, were considered but did not provide a stable improvement. The resulting model is simpler than some alternatives, but it is more interpretable and better supported by convergence and residual diagnostics.

9 Appendix

STL Decomposition of Returns

The STL decomposition is included in the appendix because it serves mainly as supporting exploratory evidence rather than a central result. It supports the decision to avoid a dynamic return-mean process: the trend component is small relative to the variability in the raw and residual series, and the seasonal component is also minor. What stands out most is the changing residual scale, which is consistent with volatility clustering rather than time variation in the conditional mean.

The STL decomposition supports focusing on volatility rather than a dynamic return mean. The trend component is negligible relative to the variation in the raw and residual series, indicating that time variation arises from changing dispersion rather than a drifting mean. The seasonal component is also small, suggesting that a deterministic weekly effect is not important. In contrast, the residuals show periods of varying variability, reinforcing the presence of stochastic volatility. Accordingly, the final model uses a constant return mean and assigns the latent state to log volatility.

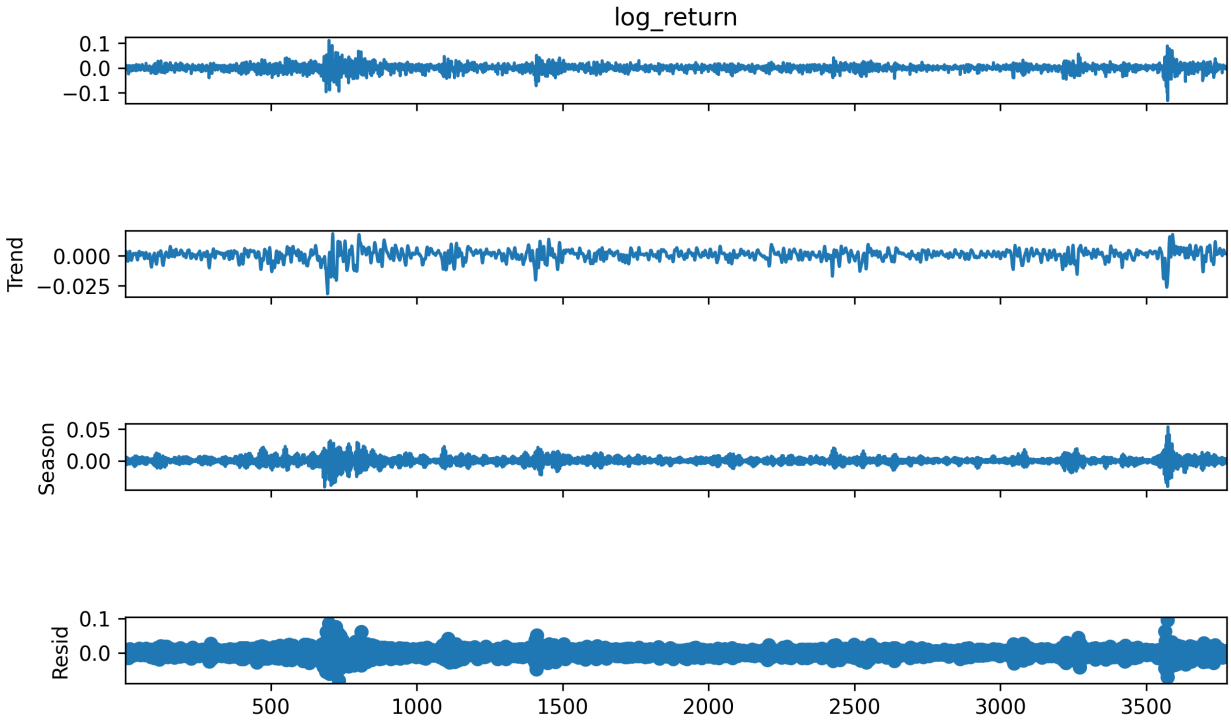

Figure 6: STL decomposition of training-sample log returns with a five-day seasonal period.

A.Screened Candidates

```
Top screened candidates:
    theta_idx       logLik  se       mu0       mu_h   phi_raw  sigma_h_raw  \
32         32  7194.336426 NaN -0.000096 -10.184870  1.366470     2.401022
22         22  7173.803711 NaN  0.001302  -9.587682  3.444559    -0.014417
28         28  7161.706543 NaN  0.000179  -9.080803  3.892637     1.810598
51         51  7047.165527 NaN  0.001109  -8.631680  5.003430    -0.498651
15         15  6927.145996 NaN  0.001379  -8.912409 -2.967624     3.655083
56         56  6875.271484 NaN  0.001248 -10.734420  2.356606     3.986435
8           8  6670.919922 NaN  0.000719  -8.997445  2.632609    -1.697045
19         19  6659.193848 NaN  0.000790  -6.758636  4.367156     3.515938
1           1  6515.011230 NaN  0.000724  -7.442936  1.100745    -0.073732
5           5  6288.291504 NaN  0.000416  -6.750123  0.880575     3.884778

    range_bias_mean  log_sigma_range_meas   sigma_h  sigma_range_meas  \
32        -0.568543             -0.769680  0.459284          0.463161
22        -0.519458             -0.306696  0.253234          0.735874
28        -0.231926             -0.561460  0.431123          0.570375
51        -0.035064             -0.349101  0.195150          0.705322
15        -0.289791             -0.817177  0.487648          0.441677
56        -0.394413             -1.233698  0.491069          0.291214
8         -0.246850             -0.617557  0.085877          0.539260
19         0.371101             -0.348938  0.485858          0.705437
1         -0.961865             -0.140386  0.245972          0.869023
5         -0.172155             -0.050332  0.490132          0.950913

         phi
32  0.948405
22  0.991298
28  0.994024
51  0.997339
15  0.762179
56  0.977451
8   0.982300
19  0.995880
1   0.936850
5   0.926028
```

B. Global Search Results

```
Final refined global-search results:
   theta_idx       logLik        se       mu0      mu_h    phi_raw  \
1          1  7326.389582  0.946445  0.001484 -9.951039   0.309273
0          0  7225.198643  1.889146  0.002605 -9.311466  -1.510286
```

```
7          7  7184.574925  0.692338  0.001214 -9.529055 -31.481552
5          5  7160.469527  2.066784  0.002852 -9.527737  -2.189986
3          3  7157.763623  0.468872  0.002128 -9.586003 -11.772535
4          4  7112.365383  1.679223  0.002469 -9.565683 -10.459702
2          2  7100.689543  0.435653  0.002872 -9.469073  -3.462732
6          6  7048.514018  0.443472  0.003445 -9.387525 -10.412394

   sigma_h_raw  range_bias_mean  log_sigma_range_meas   sigma_h  \
1    37.638260        -0.366602             -0.665399  0.500000
0     9.298450        -0.452161             -0.736114  0.499955
7    36.751968        -0.272701             -0.722354  0.500000
5    22.342865        -0.298027             -0.637669  0.500000
3    15.243361        -0.200872             -0.730142  0.500000
4     8.145663        -0.159821             -0.738626  0.499858
2    10.526359        -0.228582             -0.711091  0.499987
6    15.425654        -0.356928             -0.662674  0.500000

   sigma_range_meas       phi
1          0.514068  0.893600
0          0.478971  0.795043
7          0.485608  0.750000
5          0.528523  0.775063
3          0.481841  0.750002
4          0.477770  0.750007
2          0.491108  0.757567
6          0.515471  0.750007
Held-out test results from the best training fit:
   theta_idx       logLik       se       mu0      mu_h   phi_raw  sigma_h_raw  \
0          0  2237.203418  0.19559  0.001484 -9.951039  0.309273     37.63826

   range_bias_mean  log_sigma_range_meas  sigma_h  sigma_range_meas     phi
0        -0.366602             -0.665399      0.5          0.514068  0.8936
```

C. Top Screened Leverage Candidates

```
Top screened leverage candidates:
    theta_idx        logLik  se       mu_h  h0_offset_raw   phi_raw  \
47         47  11612.813477 NaN  -9.519747       1.118810  1.093126
63         63  11519.716797 NaN  -9.919420       0.709298  2.393186
11         11  11447.878906 NaN -10.946070       1.584620  3.373212
52         52  11359.287109 NaN  -9.120897      -0.085326  4.396808
62         62  11184.492188 NaN  -6.881178      -1.484296  4.087827
4           4  11183.820312 NaN  -8.430325      -0.387958  1.002699
27         27  11119.266602 NaN  -9.545393       1.621243 -3.320145
```

```
30         30  11115.778320 NaN  -7.704338      -0.128964  4.096878
8           8  11109.786133 NaN  -8.749056       0.186509  5.471469
29         29  11042.395508 NaN  -8.196203      -0.541829 -0.065912

    log_sigma_eta  log_sigma_nu        G0       phi  sigma_eta  sigma_nu  \
47      -0.196013     -3.843147 -0.575171  0.747735   0.822001  0.021426
63      -0.607649     -2.619969 -0.203014  0.912562   0.544630  0.072805
11      -0.416630     -3.715824  0.933001  0.962354   0.659265  0.024335
52      -1.545662     -1.658590  0.439740  0.983016   0.213171  0.190407
62      -0.761696     -3.224063  0.696177  0.978749   0.466874  0.039793
4       -0.673484     -0.758955  1.017113  0.730615   0.509929  0.468156
27      -0.380855     -0.864710  0.887086  0.044363   0.683277  0.421174
30      -1.224423     -6.201994 -0.642389  0.978893   0.293927  0.002025
8       -2.194770     -2.636293 -1.339704  0.990875   0.111384  0.071626
29      -0.395585     -4.443036 -0.514602  0.486275   0.673286  0.011760

    h0_offset        R0
47   1.614309 -0.519147
63   1.220473 -0.200271
11   1.838641  0.731990
52  -0.170239  0.413429
62  -1.804539  0.601936
4   -0.739197  0.768688
27   1.849609  0.709951
30  -0.256508 -0.566524
8    0.368751 -0.871601
29  -0.988743 -0.473523
```

D. Final Refined Leverage Global-search Results

```
Held-out test results from the best leverage training fit:
   theta_idx       logLik        se      mu_h  h0_offset_raw   phi_raw  \
0          0  3632.127196  0.402253 -9.213733      -10.97839  3.037332

   log_sigma_eta  log_sigma_nu        G0       phi  sigma_eta  sigma_nu  \
0      -0.150835      -1.84285 -3.130135  0.949919    0.85999  0.158365

   h0_offset        R0
0       -2.0 -0.996186
```

References

Bréto, Carles, Daihai He, Edward L. Ionides, and Aaron A. King. 2009. "Time Series Analysis via Mechanistic Models." *The Annals of Applied Statistics* 3 (1): 319–48.

Ionides, Edward. 2026a. "Lecture 15: Iterated Filtering — Principles and Practice." Lecture slides, University of Michigan.

———. 2026b. "Lecture 17: A Case Study of Financial Volatility and a POMP Model with Observations Driving Latent Dynamics." Lecture slides, University of Michigan.

Ionides, Edward L., Dao Nguyen, Yves Atchadé, Stilian Stoev, and Aaron A. King. 2015. "Inference for Dynamic and Latent Variable Models via Iterated, Perturbed Bayes Maps." *Proceedings of the National Academy of Sciences* 112 (3): 719–24.

Parkinson, Michael. 1980. "The Extreme Value Method for Estimating the Variance of the Rate of Return." *Journal of Business* 53 (1): 61–65.

Taylor, Stephen J. 1982. "Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices 1961–75." *Time Series Analysis: Theory and Practice* 1: 203–26.

Yahoo Finance. 2026. "Historical Market Data for NASDAQ Composite Index (^IXIC)." https://finance.yahoo.com/.