

Applications of Bootstrap and Shrinkage in the GARCH Model

Jianzhi Wang, Jingxu Xie, Mingrui Zhang

May 10, 2024

1 Introduction

1.1 Background

Time series analysis is a fundamental statistical method for analyzing data collected over time. It provides crucial insights into underlying patterns, essential for forecasting and making informed decisions. A notable application of time series analysis is in the financial sector, particularly in studying stock market indices such as the SPDR S&P 500 ETF Trust (SPY). This index tracks the performance of the S&P 500 and serves as a barometer for the U.S. equities market. Techniques like calculating daily log returns and examining autocorrelations help analysts understand market dynamics and develop robust models to forecast future price movements.

This report primarily focuses on predicting the volatility of SPY data, a key aspect of financial time series that reflects the variability in trading prices. We employ the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, known for its effectiveness in modeling the time-varying nature of volatility and capturing volatility clustering common in financial time series.

1.2 Bootstrap and shrinkage methods

To enhance the predictive accuracy of the GARCH model, we explore bootstrap methods. These methods allow analysts to understand the variability and uncertainty in predictions by resampling the original dataset with replacement. This process generates numerous simulated samples, enabling the estimation of an estimator's distribution. Bootstrap techniques, particularly those developed by Efron (1983, 1986) and refined by Efron and Tibshirani (1997) through the .632+ rule, effectively estimate prediction error rates that traditional models often underestimate due to overfitting.

Shrinkage methods adjust traditional estimators to improve prediction accuracy by introducing some bias. As illustrated in Stein's paradox (Efron and Morris, 1977), shrinkage methods can outperform traditional unbiased estimators, especially in high-dimensional settings. We apply Stein-type predictors, discussed by Copas (1983), which under certain conditions offer lower prediction mean squared errors compared to traditional least squares estimates. This significantly enhances the accuracy and reliability of financial forecasts.

By integrating these advanced statistical techniques, this report aims to develop a more robust framework for understanding and forecasting market volatility, thereby aiding more informed investment decision-making.

2 Preliminary data analysis

2.1 SPY data and autoregressive model

We collect the historical closing prices, which have been adjusted for dividends, of the SPY index for the period from 2008 to 2024. From these closing prices, the daily log returns (the natural logarithm of the ratio of consecutive closing prices) was calculated. Figures 1 and 2 illustrate the time series graphs of the closing prices and the log returns respectively.

Figure 3 presents the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the daily log returns. The significant partial autocorrelation at lag 1 in the PACF plot supports the use of an autoregressive model of order 1 (AR(1)) to model the daily log returns. Also, extracting out the partial autocorrelation signal makes the residuals more suitable for the fitting of GARCH model later on. We denote the log return at time t as r_t and the corresponding residual from the AR(1) model fit as $\tilde{\epsilon}_t$. Throughout the rest of the report, variables marked with a tilde signify those derived from observed quantities. Figure 4 displays the fitted values as well as the residuals of the AR(1) model fit.

We calculate the daily realized volatility at time t , denoted by $\tilde{\sigma}_t$, by aggregating intraday 30-minute changes in price $(r_{t,i})_{i=1}^{13}$.

$$\tilde{\sigma}_t = \sqrt{\sum_{i=1}^{13} r_{t,i}^2}$$

The actual standardized residual is then defined as $\tilde{e}_t = \tilde{\epsilon}_t / \tilde{\sigma}_t$. The series $(\tilde{\sigma}_t)_t$ and $(\tilde{e}_t)_t$ are plotted in Figures 5 and 6, respectively. Figure 7 depicts the ACF and PACF of the actual standardized residuals $(\tilde{e}_t)_t$. The daily volatility graph shows periods of extreme volatility. In addition, the ACF and PACF plots show that there were only a few significant autocorrelation or partial autocorrelation coefficients. Hence, the GARCH model can be used at this stage to model $(\tilde{e}_t)_t$. The only caveat is that according to the Shapiro–Wilk Test, the actual standardized residual $(\tilde{e}_t)_t$ distribution gave a p -value of less than 10^{-16} , which deviates from the normality assumption of the GARCH model.

2.2 GARCH model

We allocate the first 80% of the data, corresponding to 3,218 days, as the training set and use the remaining 20% of the data, comprising 805 days, as the test set. Our analysis utilizes the following

GARCH(1,1) model, indicating one autoregressive lag and one moving average lag.

$$\begin{aligned}\epsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2\end{aligned}$$

In the model, $e_t \sim N(0, 1)$ represents white noise and is independently and identically distributed. The latent volatility process is denoted by σ_t^2 .

Table 1 provides a summary of the GARCH(1,1) model fit, including the estimated coefficients $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$. These coefficients allow for volatility predictions at time t via the following recursive formula.

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha} \tilde{\epsilon}_{t-1}^2 + \hat{\beta} \hat{\sigma}_{t-1}^2.$$

Figures 8–9 depict the comparison between predicted $\hat{\sigma}_t$ with the actual volatility $\tilde{\sigma}_t$, for horizons of 1 day and 30 days respectively. Comparing the two figures visually, a longer horizon leads to lower variance in predictions. However, it may suffer from higher prediction error, which is clear from the period between 2021 to 2022 where the predicted volatility is almost constantly higher than the realised volatility.

The Mean Squared Prediction Error (MSPE) is used to assess prediction quality on the test set.

$$\text{MSPE} = \frac{1}{|\text{test set}|} \sum_{t \in \text{test set}} (\tilde{\sigma}_t - \hat{\sigma}_t)^2$$

Figure 10 illustrates the MSPE variation across different horizons, indicating a decline in prediction accuracy as the horizon increases. This corroborates well with Figures 8–9.

2.3 Bootstrap and prediction error estimation

The fitted standardized residuals are calculated by $\hat{e}_t = \tilde{e}_t / \hat{\sigma}_t$. Since the variable e_t in the GARCH model is assumed to be independently and identically distributed, it is a natural variable to conduct bootstrap on. To construct a bootstrap sample, we first resample $\{\hat{e}_t^b\}$ from the pool of fitted standardized residuals $\{\hat{e}_t\}$ with replacement. Using these samples, we reconstruct the bootstrap residuals $\{\hat{e}_t^b\}$, where each $\hat{e}_t^b = \hat{\sigma}_t \hat{e}_t^b$. With the bootstrap samples in place, we then refit the GARCH(1,1) model to $(\hat{e}_t^b)_t$. Repeating this $N = 100$ times gives us the bootstrap standard errors for the GARCH model coefficients, shown in Table 2. Comparing to the nominal standard errors in Table 1, we see that the bootstrap standard errors were close to, but less than, the nominal standard errors. The closeness of these two sets of standard errors show that GARCH(1,1) is a reasonable model for fitting the volatility.

We use the bootstrap samples again to estimate the standard error of the predicted volatility $\hat{\sigma}_t$ and construct the 95% confidence intervals for each actual volatility $\tilde{\sigma}_t$. Figure 11 displays the 95% confidence intervals with the actual volatility $\tilde{\sigma}_t$. The analysis reveals a coverage rate of approximately 91%, which is close to the intended coverage ratio. This shows that the bootstrap method is capable of producing a reasonable prediction interval for the volatility.

2.4 Shrinkage in the volatility prediction

We directly apply a shrinkage factor to the predicted volatility $\hat{\sigma}_t$.

$$\hat{\sigma}_t^{\text{shr}} = K \hat{\sigma}_t$$

Figure 12 displays the MSPE curves as the shrinkage factor K varies across different horizons. The results indicate that the optimal shrinkage factor K^* ranges between 0.9 and 1. Notably, the optimal K^* remains relatively stable across different prediction horizons, suggesting a consistent improvement effect of shrinkage across varying time frames.

Motivated by [Copas \(1983\)](#), we also experimented with applying shrinkage factors to the individual GARCH parameters. The results across various time horizons are shown in Figure 13. However, this way of shrinkage is not as effective as directly applying shrinkage to the whole prediction, as demonstrated by the sharp increase in MSPE away from the optimal shrinkage factor.

3 Data analysis with rolling windows

In this section, we replicate our analysis using rolling windows, prompted by the observation that the estimated coefficients $\hat{\omega}$, $\hat{\alpha}$, and $\hat{\beta}$ vary across different windows. This variation is depicted in Figure 14, which uses a window size of 500, and Figure 15, which uses a window size of 1,500.

When shrinkage is not applied, Figures 16–17 display the MSPE on the test set across various horizons, for window sizes of 500 and 1,500, respectively. Notably, for a one-day horizon, a smaller window size results in better prediction accuracy, whereas for a thirty-day horizon, a larger window size proves to be more accurate.

Next, we incorporate shrinkage into the analysis. Figures 18–19 illustrate the MSPE on the test set for the same respective window sizes. For a window size of 500, we observe a decreasing optimal shrinkage factor as the prediction horizon lengthens. Conversely, for a window size of 1,500, the optimal shrinkage factor tends to increase with longer prediction horizons. Strangely, for large enough horizon, the optimal shrinkage factor may even exceed 1, as shown in Figure 19. This suggests that the relationship between window size, shrinkage factor, and prediction horizon is complex and warrants further investigation to optimize forecasting performance.

Motivated by the trend in the time series of $\hat{\alpha}_t$ and $\hat{\beta}_t$, an experiment was done to fit a linear line to those series, obtaining the best linear estimates $\hat{\alpha}_{t,\text{ls}}$ and $\hat{\beta}_{t,\text{ls}}$. The originally fitted GARCH parameters $\hat{\alpha}_t, \hat{\beta}_t$ were shrunk towards the regression line to obtain new parameters $\hat{\alpha}'_t, \hat{\beta}'_t$. Subsequently, the new parameters $\hat{\alpha}'_t, \hat{\beta}'_t$ were used to perform the next prediction. This idea was motivated from [Efron and Morris \(1977\)](#), which suggested a reference point to shrink towards.

$$\hat{\alpha}'_t = \hat{\alpha}_{t,\text{ls}} + \lambda(\hat{\alpha}_t - \hat{\alpha}_{t,\text{ls}})$$

$$\hat{\beta}'_t = \hat{\beta}_{t,\text{ls}} + \lambda(\hat{\beta}_t - \hat{\beta}_{t,\text{ls}})$$

The result of the experiment, as shown in Table 3, is counterintuitive. It suggests that setting $\lambda > 1$ improves the MSPE, contrary to the spirit of shrinkage. Extremely large values of λ beyond 1 are not possible because they lead to negative $\hat{\alpha}'_t$ and $\hat{\beta}'_t$, which are not permissible as GARCH parameters. To explain this strange phenomenon, we suspect that pushing $\hat{\alpha}$ and $\hat{\beta}$ away from their respective rolling window regression line helped to tailor to the highly volatile $(\tilde{\sigma}_t)_t$ time series of the test set, resulting in a lower MSPE.

4 Simulation study

We simulate volatility data using a GARCH(1,1) model with true parameters fixed at $\omega = 0.02$, $\alpha = 0.1$, and $\beta = 0.9$. For the generation of $\tilde{\epsilon}_t$, it is necessary to simulate the series e_t within the GARCH framework. We employ two distinct methods to generate e_t and compare their results.

In our first method, we generate e_t independently from a standard normal distribution, $e_t \sim N(0, 1)$. The rationale for this simulation method is to be aligned with the original GARCH model assumption. A summary of the GARCH(1,1) model fit under this assumption is provided in Table 4. Here, we observe that the coefficient estimates are biased, even under this idealized scenario where the data are simulated from the GARCH process. Figure 20 compares the predicted volatility $\hat{\sigma}_t$ against the actual simulated volatility $\tilde{\sigma}_t$ for a one-day horizon.

Alternatively, we can generate e_t by sampling from the set of fitted standardized residuals \hat{e}_t (outlined in Section 2.3) with replacement. The summary of the GARCH(1,1) model fit for this method is displayed in Table 5. The coefficient estimates in this approach are similar to those found in Table 4, yet the standard errors are smaller. Figure 21 illustrates a comparison between the predicted $\hat{\sigma}_t$ and the actual simulated volatility $\tilde{\sigma}_t$ for a one-day horizon, analogous to the first method.

Finally, we compare the prediction interval coverage for both methods. Figure 22 shows the 95% prediction interval under the normal error simulation method, which has a coverage ratio of 87%. This is less than the coverage ratio (91%) from bootstrapping the residuals. The result is expected, since bootstrapping incorporates information from the residuals, whose distribution is not normal.

5 Conclusion

In summary, we explored the GARCH model in modeling the volatility of SPY. Within the GARCH model, we discovered a natural variable to bootstrap on. The bootstrap procedure then gave more optimistic confidence intervals for the GARCH parameters. The bootstrap methodology also gave an adequate, but slightly overconfident, prediction interval. Finally, we investigated various applications of shrinkage. Our results show that a modest shrinkage of the overall volatility prediction helps improve MSPE. Surprisingly, in some instances, the optimal shrinkage parameter has magnitude greater than 1, suggesting that enlarging the predictions or GARCH parameters

seem to improve the MSPE.

References

- John B Copas. Regression, prediction and shrinkage. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 45(3):311–335, 1983.
- Bradley Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American statistical association*, 78(382):316–331, 1983.
- Bradley Efron. How biased is the apparent error rate of a prediction rule? *Journal of the American statistical Association*, 81(394):461–470, 1986.
- Bradley Efron and Carl Morris. Stein’s paradox in statistics. *Scientific American*, 236(5):119–127, 1977.
- Bradley Efron and Robert Tibshirani. Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.

A Figures

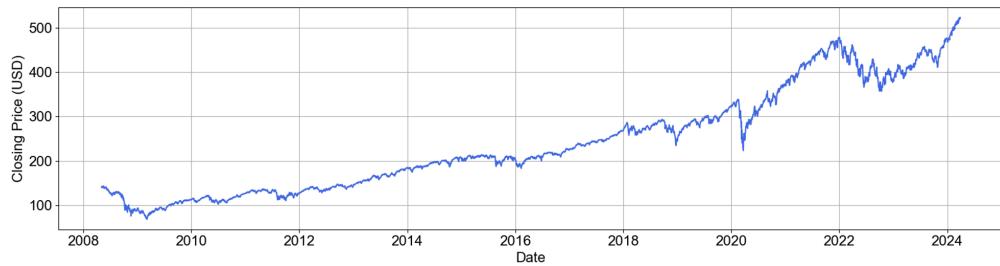


Figure 1: Closing prices of the SPY index from May 2008 to Apr 2024

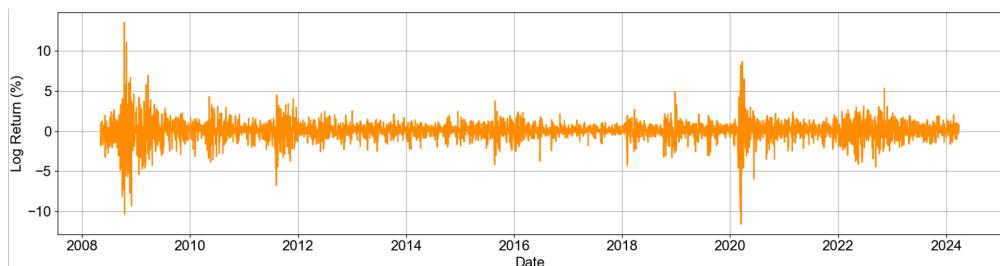


Figure 2: Log returns of the SPY index

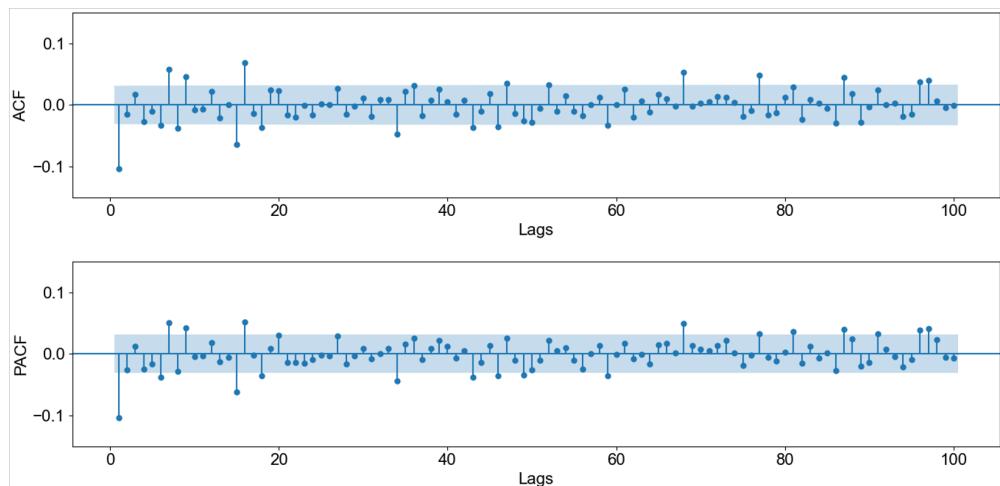


Figure 3: ACF and PACF of the log returns r_t

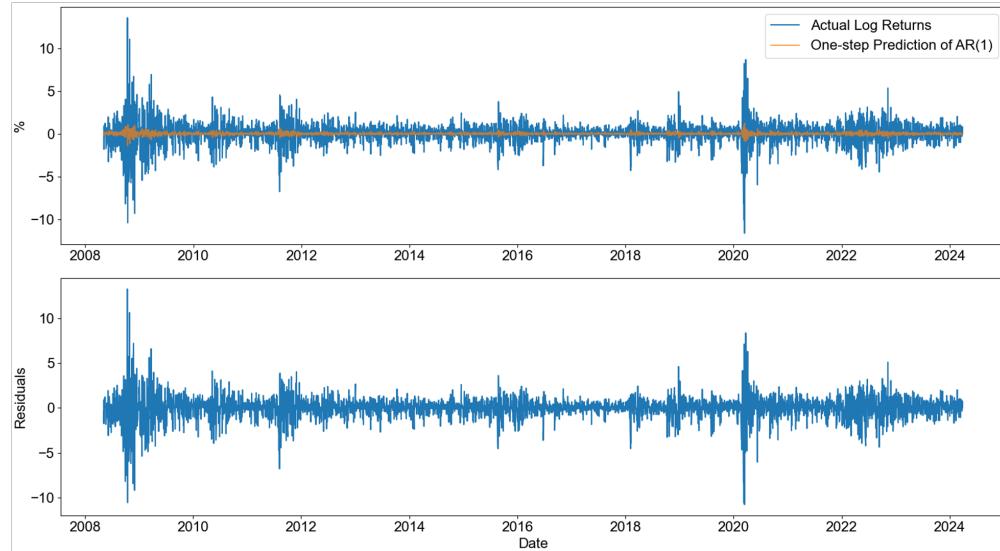


Figure 4: Fitted values (above) and residuals ϵ_t (below) of the AR(1) fit

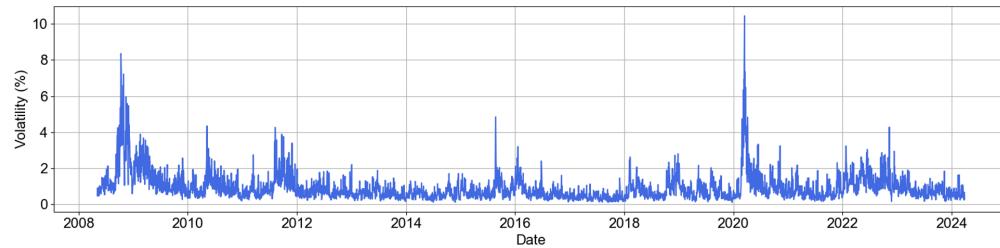


Figure 5: Daily volatility $\tilde{\sigma}_t$

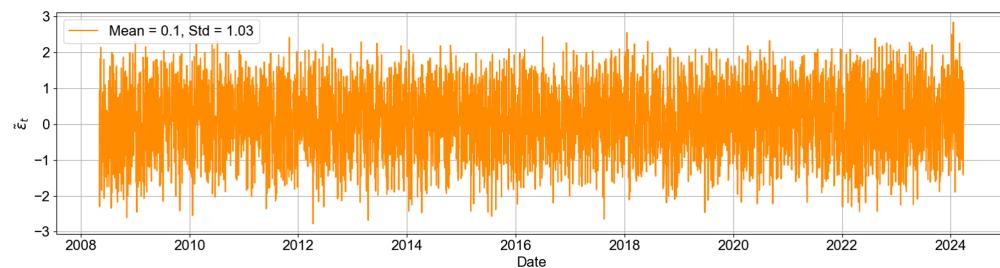


Figure 6: Actual standardized residual \tilde{e}_t

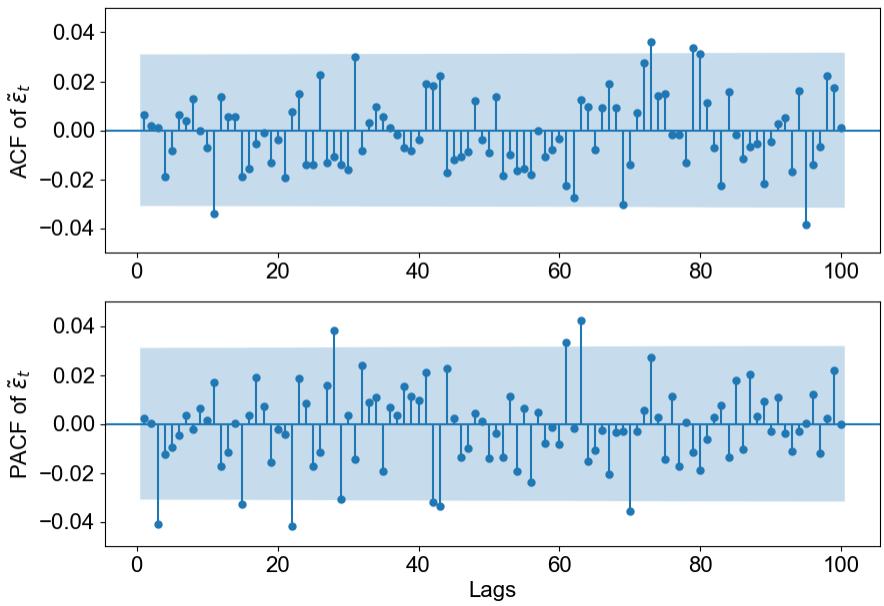


Figure 7: ACF and PACF of the actual standardized residuals \tilde{e}_t

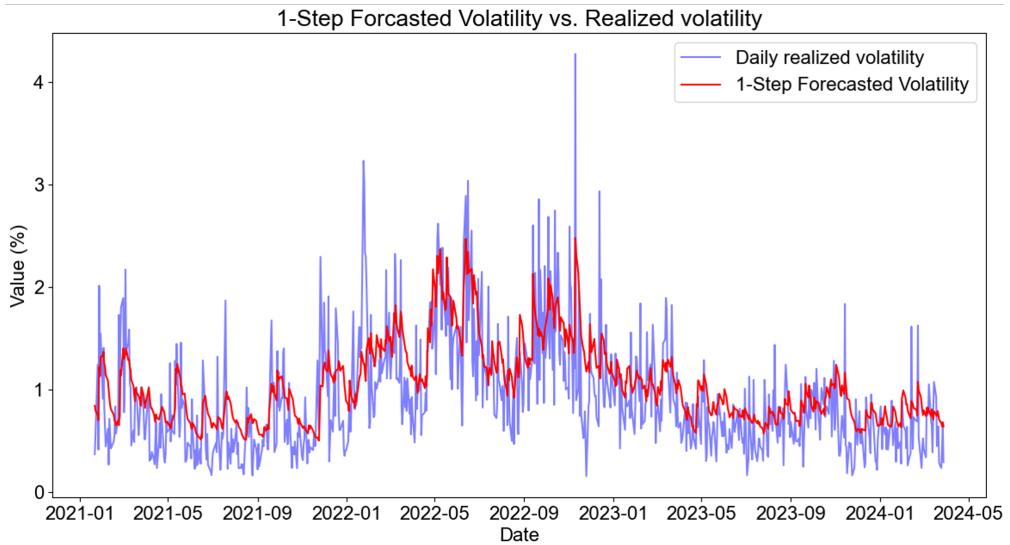


Figure 8: Prediction of $\hat{\sigma}_t$ on the test set with horizon = 1

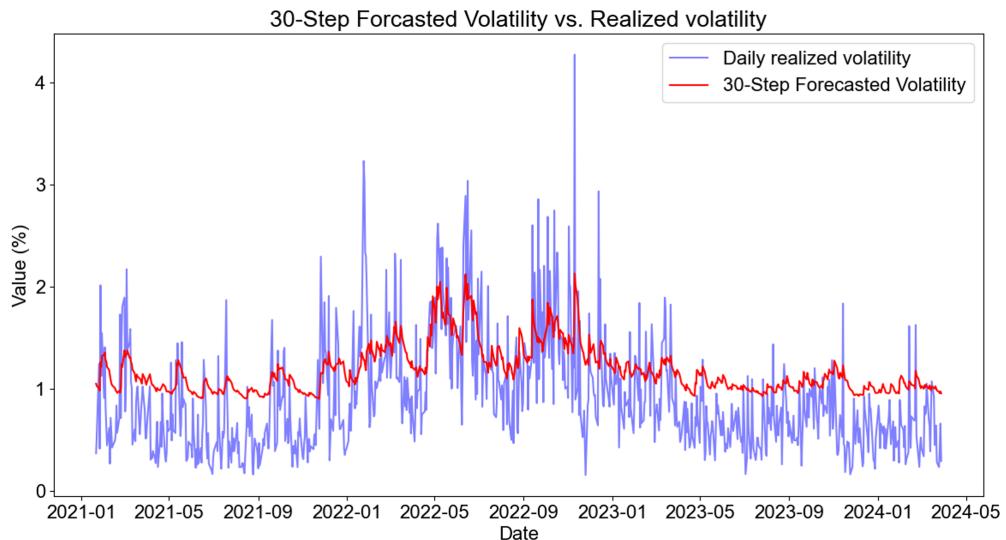


Figure 9: Prediction of $\hat{\sigma}_t$ on the test set with horizon = 30

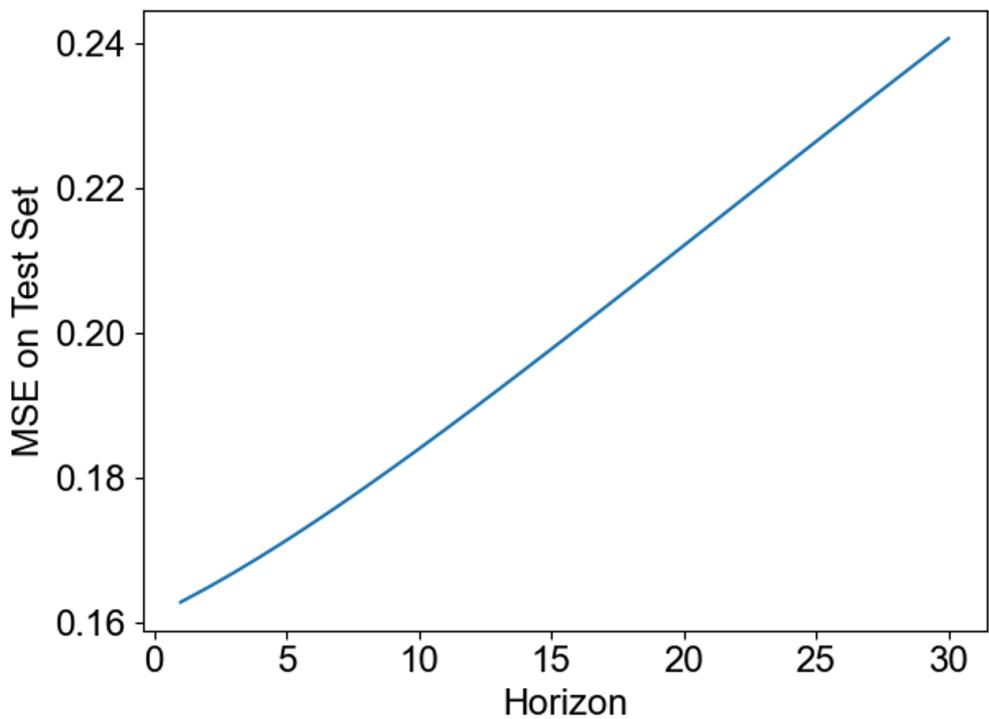


Figure 10: MSPE on the test set with increasing horizon

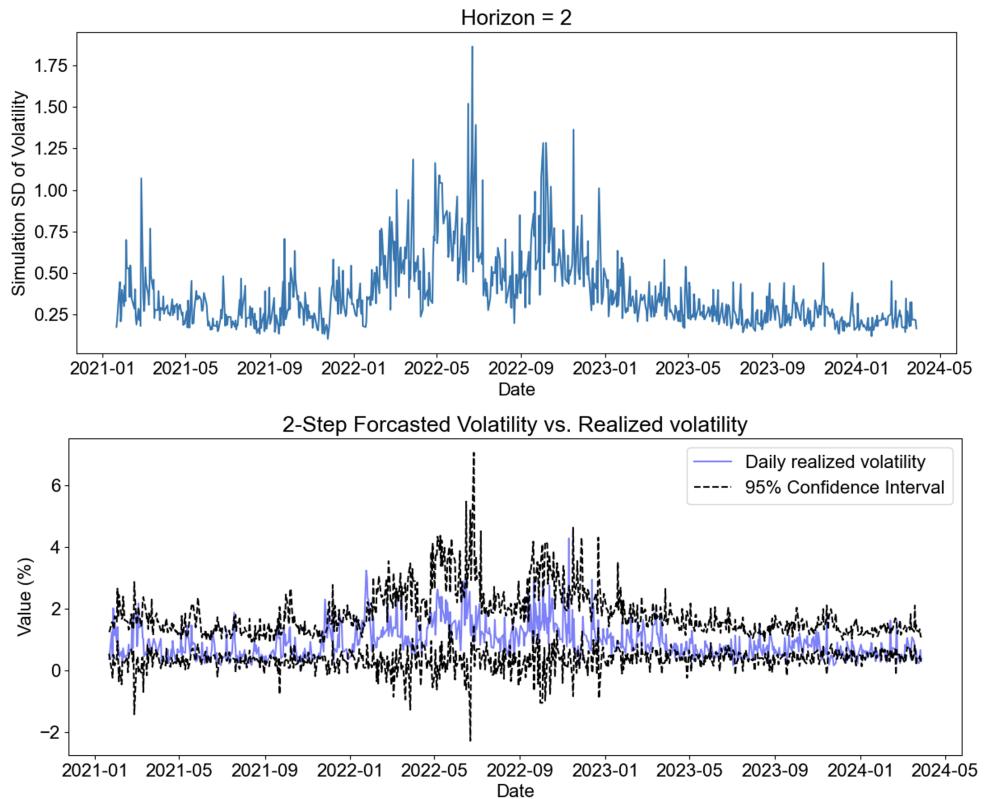


Figure 11: Prediction intervals of the volatility σ_t

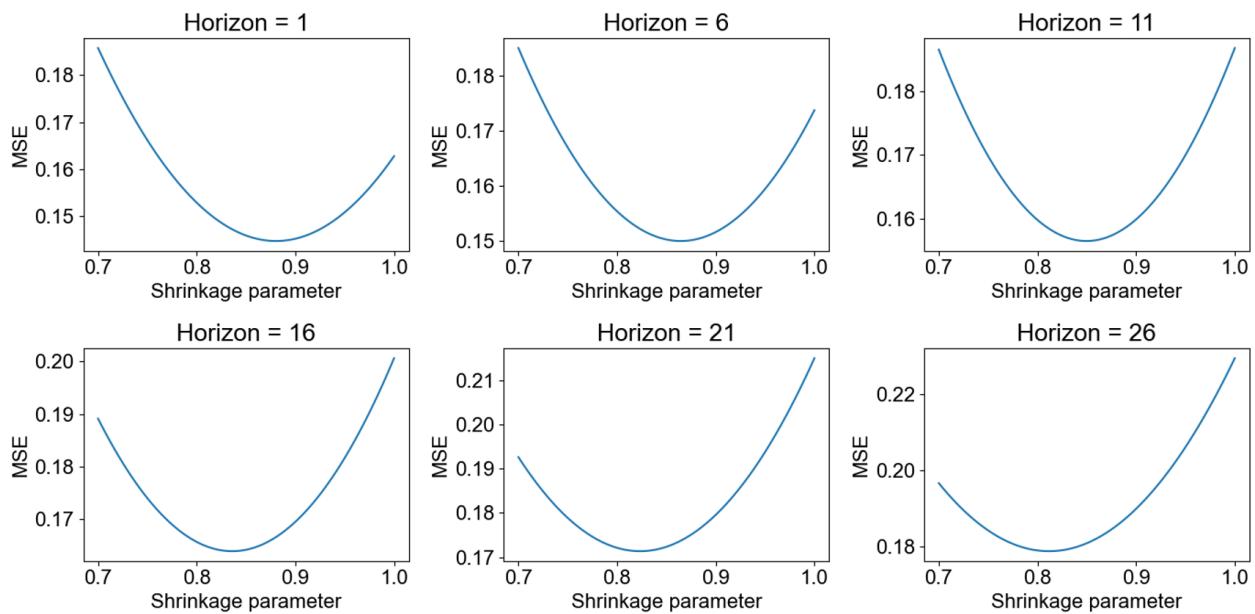


Figure 12: MSPE on the test set with shrinkage applied to volatility prediction

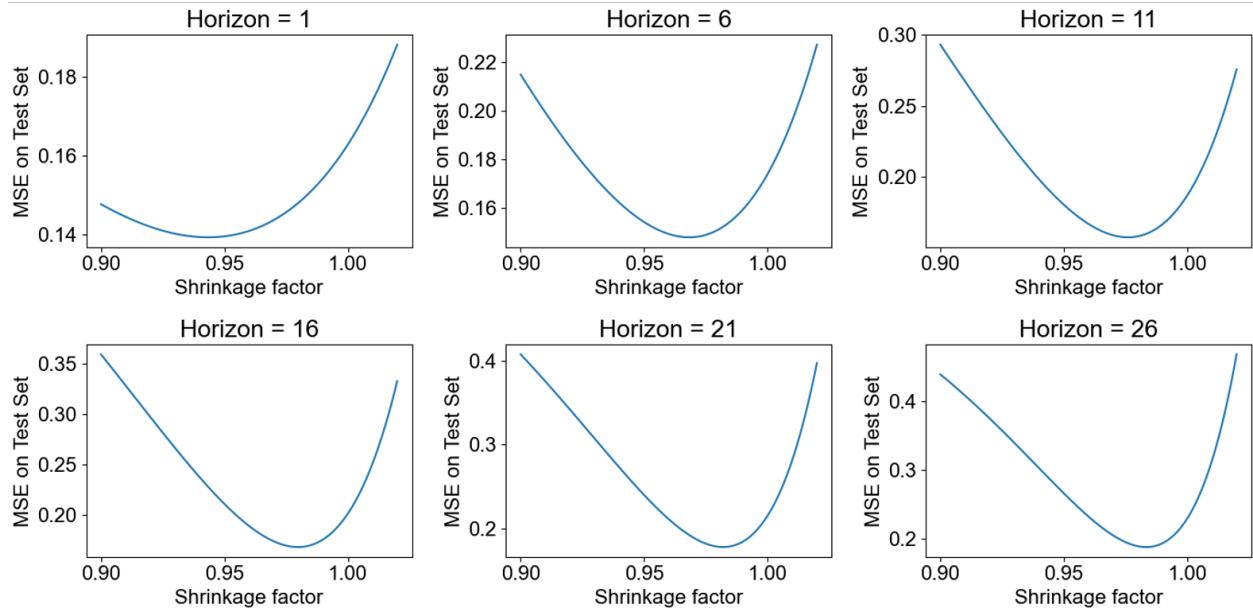


Figure 13: MSPE on the test set with shrinkage applied to parameter estimates

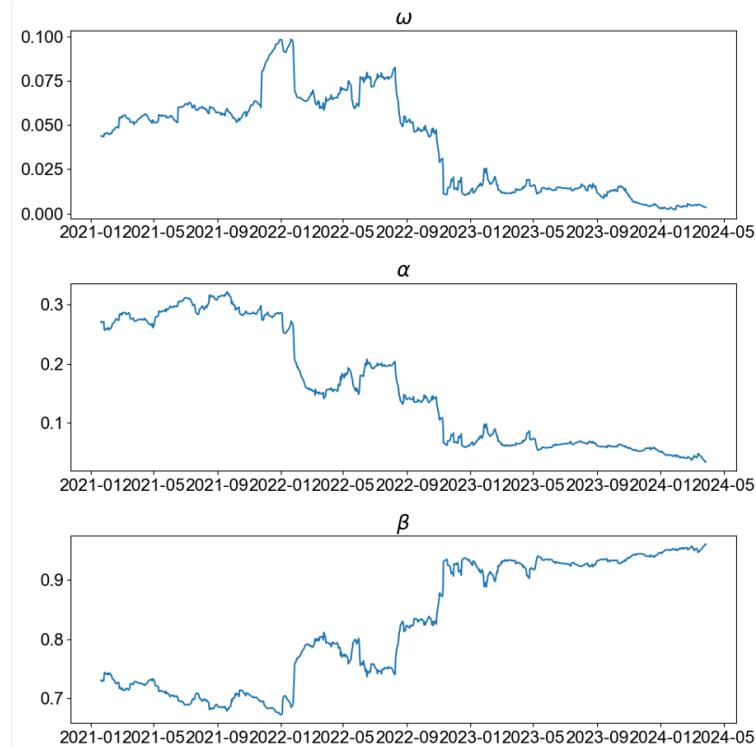


Figure 14: Graph of $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ when window size = 500

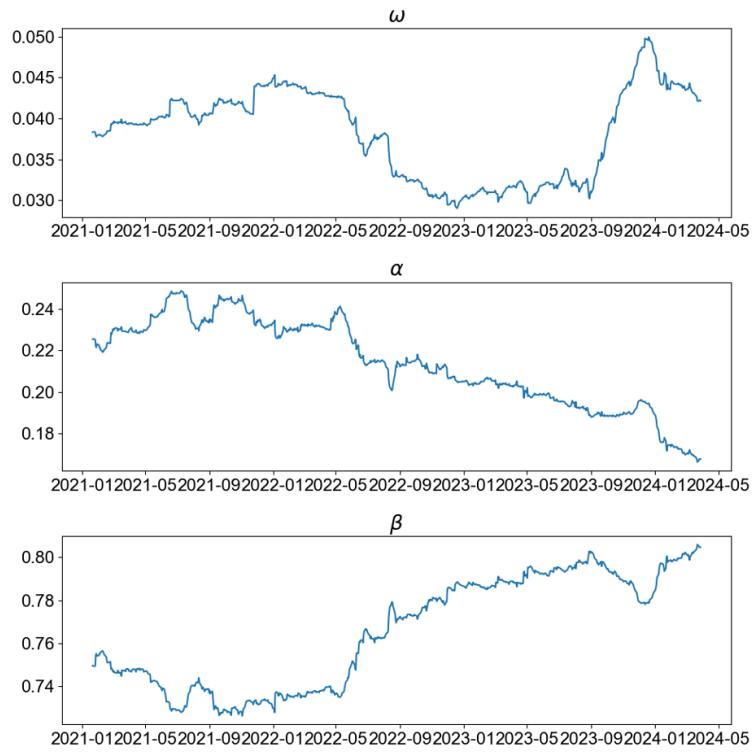


Figure 15: Graph of $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ when window size = 1,500

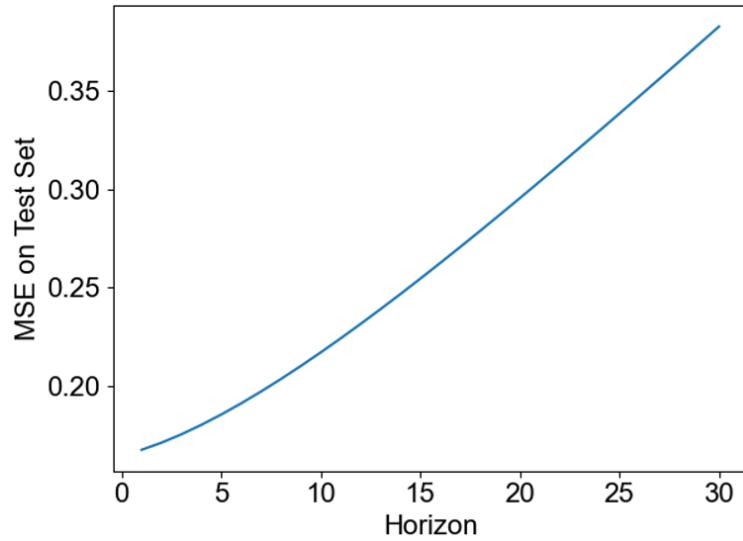


Figure 16: MSPE on the test set without shrinkage (window size = 500)

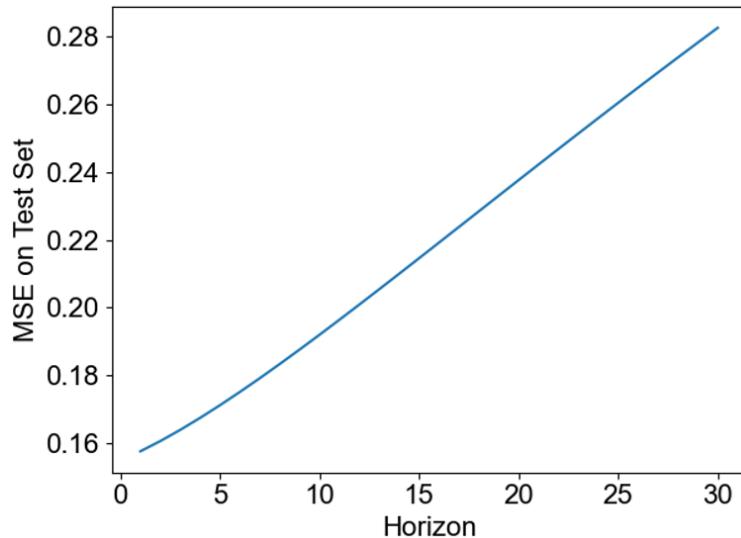


Figure 17: MSPE on the test set without shrinkage (window size = 1,500)

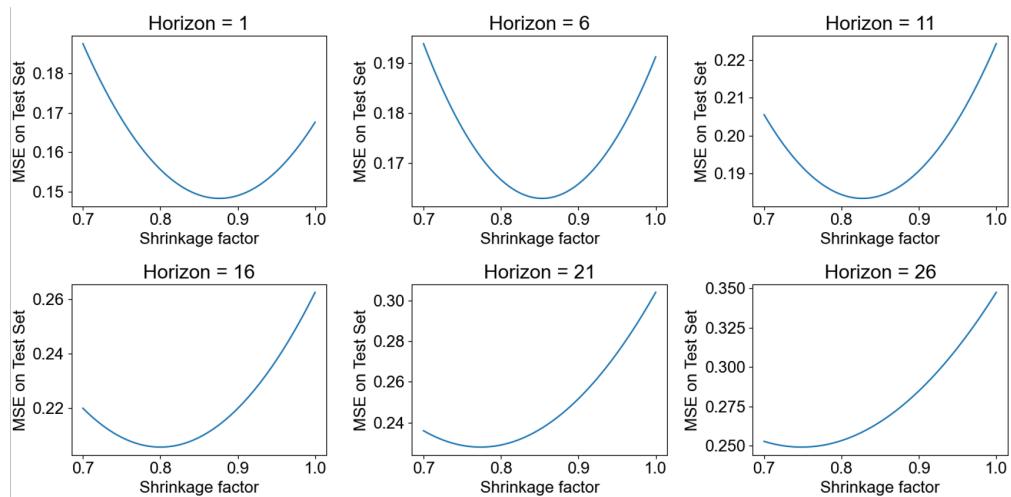


Figure 18: MSPE on the test set with shrinkage in volatility prediction (window size = 500)

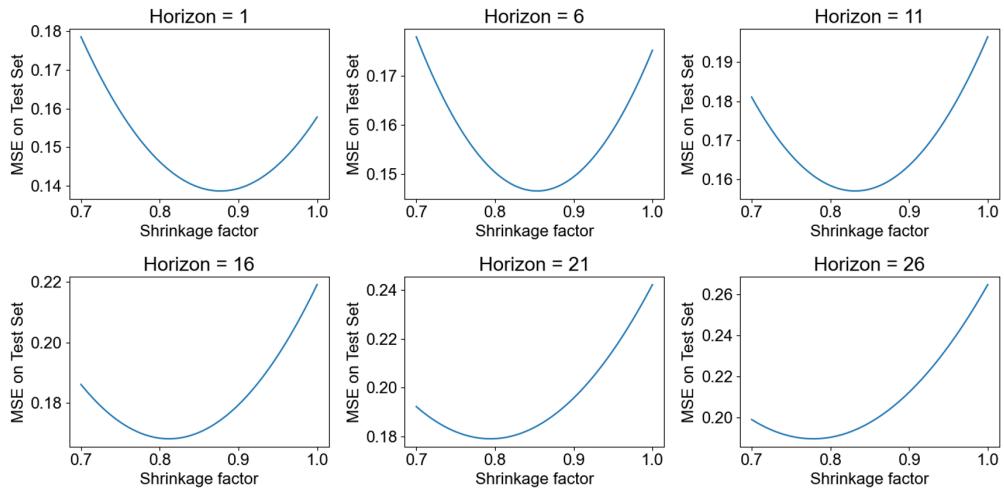


Figure 19: MSPE on the test set with shrinkage in volatility prediction (window size = 1,500)

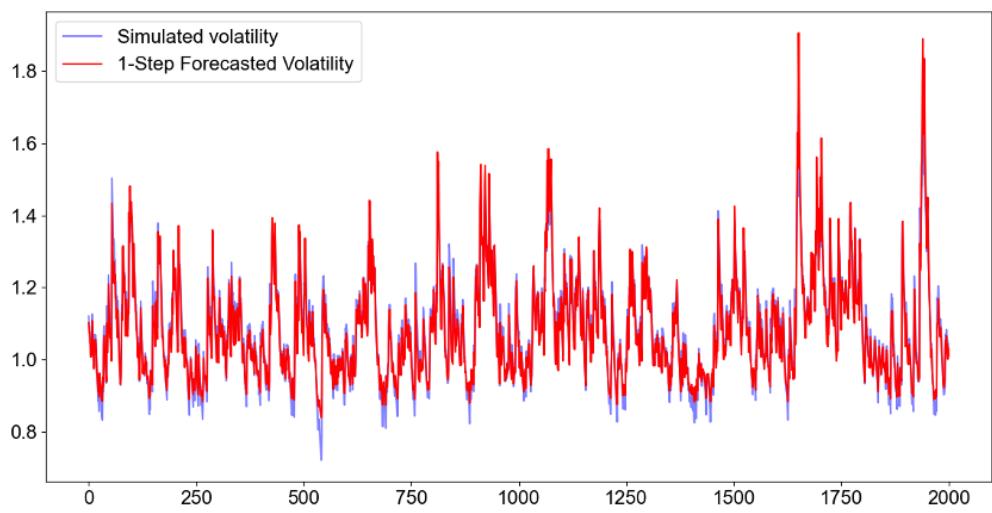


Figure 20: Predicted volatility with simulated data $e_t \sim N(0, 1)$

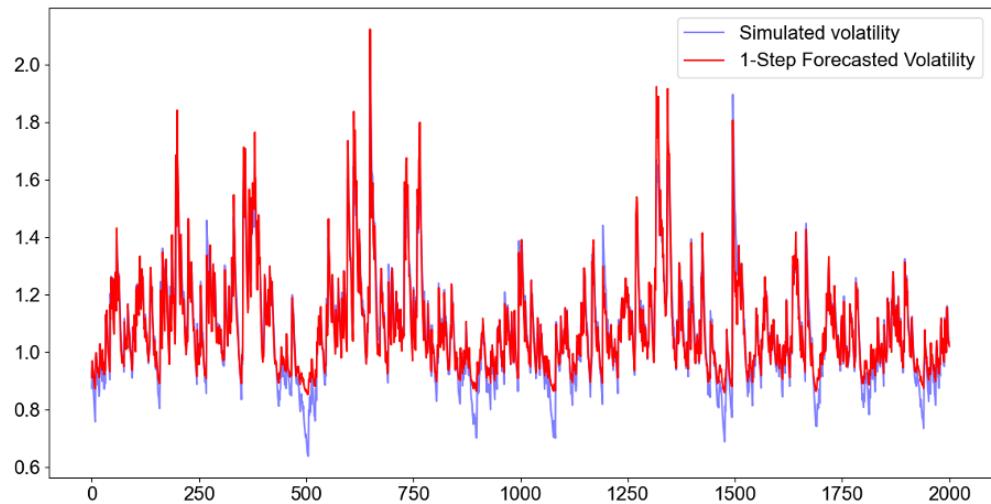


Figure 21: Predicted volatility with simulated data e_t sampled from $\{\hat{e}_t\}$

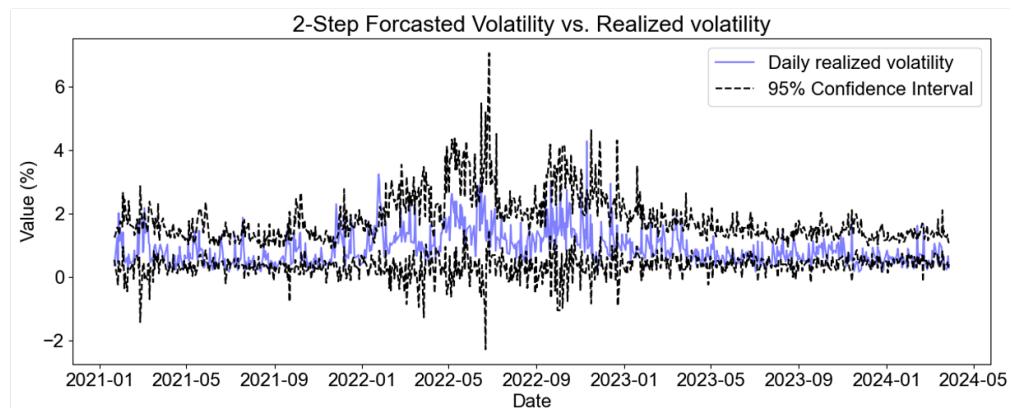


Figure 22: Prediction intervals of σ_t with normal errors

B Tables

	Estimate	SE	t-statistic	p-value
ω	0.0196	4.039e-03	4.845	1.264e-06
α	0.1090	1.141e-02	9.557	1.217e-21
β	0.8765	1.204e-02	72.817	0.000

Table 1: Summary of the GARCH model

	Bootstrap SE
ω	3.138e-03
α	9.494e-03
β	1.137e-02

Table 2: Bootstrap standard errors of coefficients in the fitted GARCH model

λ	MSPE ($\times 10^{-2}$)
0.8	5.75
0.9	5.20
1.0	4.75
1.1	4.45
1.2	4.35

Table 3: MSPE corresponding to shrinkage towards regression line parameter λ

	Estimate	SE	t-statistic	p-value
$\hat{\omega}$	0.1170	1.814-02	6.451	1.109e-10
$\hat{\alpha}$	0.0835	9.015-03	9.260	2.037e-20
$\hat{\beta}$	0.8169	2.113-02	38.657	0.000

Table 4: Summary of the GARCH model with simulated data $\omega = 0.02, \alpha = 0.1, \beta = 0.9, e_t \sim N(0, 1)$

	Estimate	SE	t-statistic	p-value
$\hat{\omega}$	0.1122	1.614e-02	6.953	3.584e-12
$\hat{\alpha}$	0.0909	8.038e-03	11.304	1.255e-29
$\hat{\beta}$	0.8184	1.768e-02	46.299	0.000

Table 5: Summary of the GARCH model with simulated data sampled from $\omega = 0.02, \alpha = 0.1, \beta = 0.9, e_t \sim N(0, 1)$