

# Value-at-Risk Forecasting Using Volatility Measures

Levani METREVELI  
Sai ZHANG

## Abstract

This project focuses on forecasting the one-day-ahead Value-at-Risk (VaR) of SPY log-returns at confidence levels 1%, 5%, and 10%. Using realized volatility ( $rv5$ ) and bipower variation ( $bv$ ), we first document key stylized facts such as volatility clustering and long-memory behavior. We then implement and compare different models for VaR forecasting. Model performance is assessed using coverage diagnostics, the quantile loss, and Kupiec proportion-of-failures tests.

**Keywords.** Value-at-Risk; realized volatility; bipower variation; jump detection; HAR-RV; GARCH; quantile regression; Kupiec proportion-of-failures test.

## Contents

<b>1 Notation and experimental protocol</b>	<b>3</b>
1.1 Notation . . . . .	3
1.2 Out-of-sample evaluation and hyperparameter tuning . . . . .	3
<b>2 Data analysis</b>	<b>3</b>
2.1 ACF of $ \log_{\text{ret}} $ . . . . .	4
2.2 ACF of $\log(rv5)$ . . . . .	5
<b>3 Evaluation metrics and VaR backtesting</b>	<b>5</b>
3.1 Violation indicator and empirical coverage . . . . .	5
3.2 Quantile pinball loss . . . . .	6
3.3 Kupiec test . . . . .	6
3.4 Composite score for model selection . . . . .	7
<b>4 GARCH model</b>	<b>7</b>
4.1 VaR forecasts . . . . .	8
4.2 Parameter tuning . . . . .	9
4.3 Results . . . . .	9
<b>5 Feature engineering</b>	<b>10</b>
5.1 Volatility features . . . . .	10
5.2 Jump proxy . . . . .	10
<b>6 HAR-RV model</b>	<b>11</b>
6.1 HAR-RV specification . . . . .	11
6.2 VaR forecasts . . . . .	11
6.3 Parameter tuning . . . . .	11
6.4 Results . . . . .	12
<b>7 Quantile regression</b>	<b>12</b>

7.1	Parameter tuning . . . . .	12
7.2	Results . . . . .	13
<b>8</b>	<b>Model comparison</b>	<b>13</b>
8.1	Calibration . . . . .	13
8.2	Accuracy . . . . .	13
<b>9</b>	<b>Conclusion</b>	<b>13</b>

# 1 Notation and experimental protocol

## 1.1 Notation

We consider a daily time series indexed by  $t = 1, \dots, T$ . The main variable of interest is the daily log-return

$$r_t = \log\text{-}\mathbf{ret}_t.$$

From intraday data, we also observe two volatility measures: realized volatility  $\mathbf{rv5}_t$  and bipower variation  $\mathbf{bv}_t$ . Throughout the report,  $X_t$  denotes the vector of predictors available at the end of day  $t$ .

For a confidence level  $\alpha \in \{0.01, 0.05, 0.10\}$ , the conditional  $\alpha$ -quantile of next-day returns is

$$Q_\alpha(r_{t+1} | X_t) = \inf\{q \in \mathbb{R} : \mathbb{P}(r_{t+1} \leq q | X_t) \geq \alpha\}.$$

We denote by  $\widehat{q}_{t,\alpha}$  the one-step-ahead forecast of this lower-tail quantile. In this report, the VaR forecast at level  $\alpha$  is this lower-tail quantile of returns:

$$\widehat{\text{VaR}}_{t+1,\alpha} \equiv \widehat{q}_{t,\alpha}.$$

## 1.2 Out-of-sample evaluation and hyperparameter tuning

To avoid look-ahead bias, we keep the chronological order of the sample and split the dataset into two consecutive parts:

- (i) Training set (first 80%) used to estimate model parameters and to tune hyperparameters via a grid search.
- (ii) Test set (last 20%) kept untouched during model selection and used only for the final out-of-sample backtesting results reported in this project.

Hyperparameters are selected on the training set using a grid-search process. For each candidate tuple of hyperparameters, we simulate a realistic forecasting procedure: at each day  $t$  in the validation period, the model is fitted using only past data and then used to produce the forecast  $\widehat{q}_{t,\alpha}$  for  $r_{t+1}$ . The best hyperparameters are chosen based on metrics which we define later. Finally, the model is refitted on the full 80% training sample with the selected hyperparameters and evaluated on the 20% test sample.

# 2 Data analysis

Our dataset contains daily observations for SPY. The main target is the daily log-return `log_ret`. Before choosing a Value-at-Risk (VaR) forecasting model, we study basic time-series properties related to volatility dynamics.

## 2.1 ACF of $|\log_{\text{ret}}|$

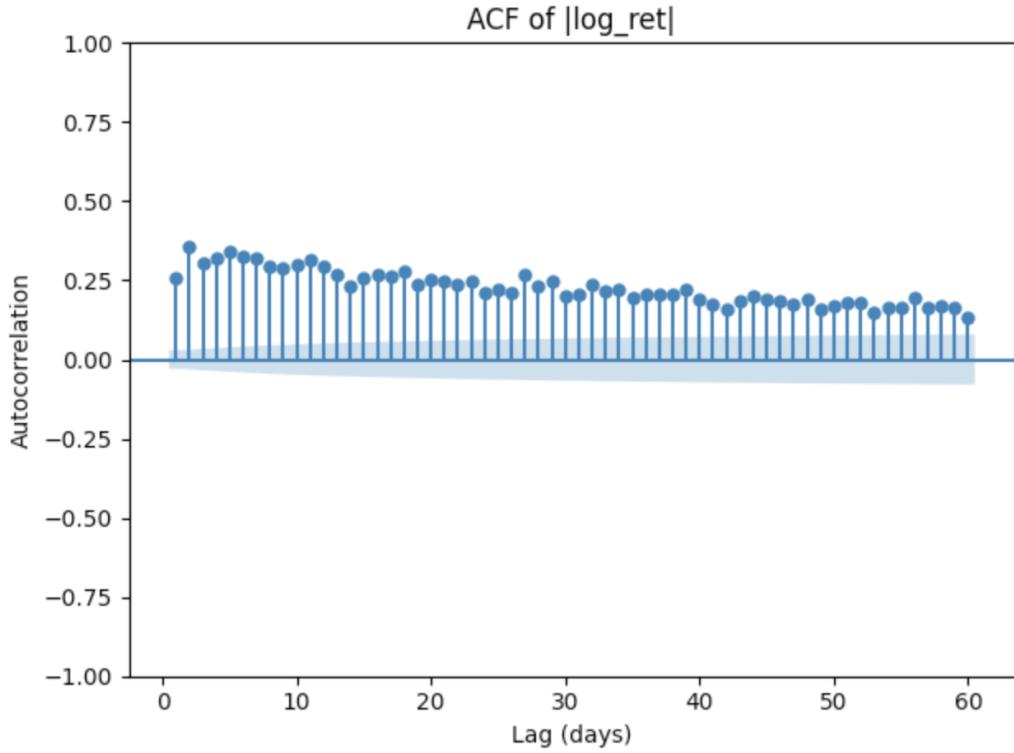


Figure 1: Autocorrelation function of  $|\log_{\text{ret}}|$ .

We first plot the autocorrelation function (ACF) of the absolute log-returns  $|\log_{\text{ret}}|$ . The ACF is clearly positive for many lags and decreases slowly. This is a typical volatility clustering pattern, large moves tend to be followed by large moves, and calm periods tend to persist. As a consequence, the volatility is predictable even if the mean return is not. This suggests that volatility contains exploitable time dependence, which motivates modeling and forecasting VaR.

## 2.2 ACF of log(rv5)

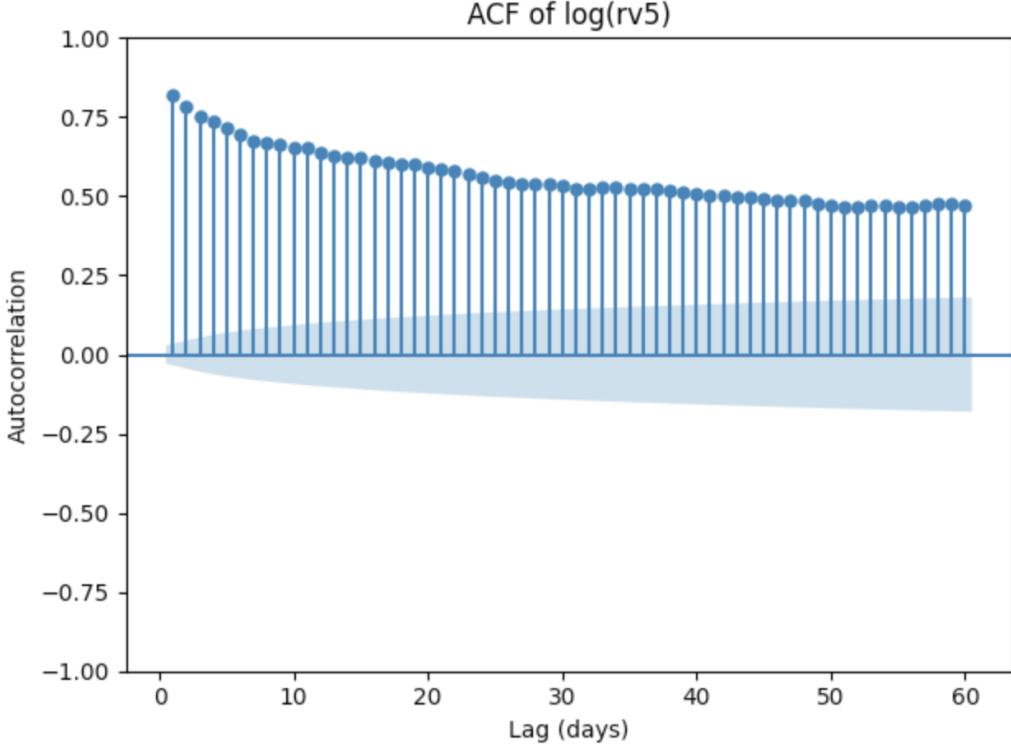


Figure 2: Autocorrelation function (ACF) of log(rv5).

We now turn to the autocorrelation of  $\log(\text{rv5})$ , a direct proxy for daily volatility computed from intraday data. Compared to  $|\log_{\text{ret}}|$ ,  $\log(\text{rv5})$  shows an even stronger dependence structure, with high autocorrelations over long lags. This indicates that volatility evolves over multiple horizons rather than reverting quickly. In the next section, we exploit this property by constructing a small set of multi-horizon volatility predictors (daily, weekly, and monthly components) based on  $\log(\text{rv5})$ .

## 3 Evaluation metrics and VaR backtesting

We assess out-of-sample performance using empirical coverage, the pinball (quantile) loss, and the Kupiec unconditional coverage test [5].

### 3.1 Violation indicator and empirical coverage

The fundamental object in VaR backtesting is the *violation* indicator:

$$I_{t+1}^{(\alpha)} = \mathbf{1}\{r_{t+1} < \hat{q}_{t,\alpha}\}.$$

A violation occurs when the realized return falls below the predicted lower-tail quantile. If the VaR model is correctly calibrated, then we expect

$$\mathbb{P}(I_{t+1}^{(\alpha)} = 1) = \alpha,$$

meaning that violations should occur about  $100\alpha\%$  of the time.

Over a test sample of length  $T$ , we define the empirical (actual) coverage:

$$\hat{\pi}_\alpha = \frac{1}{T} \sum_{t=1}^T I_{t+1}^{(\alpha)}.$$

Interpretation:

- (i) If  $\hat{\pi}_\alpha \gg \alpha$ , violations happen too often: the model *underestimates risk*.
- (ii) If  $\hat{\pi}_\alpha \ll \alpha$ , violations are too rare: the model *overestimates risk*.

Coverage is an essential sanity check, but it does not fully rank competing models because two forecasts can have similar violation rates while having very different precision.

### 3.2 Quantile pinball loss

To compare quantile forecasts beyond mere coverage, we use the *quantile loss* (also called *pinball loss*). For an error  $u_{t+1} = r_{t+1} - \hat{q}_{t,\alpha}$ , the loss function is

$$\rho_\alpha(u) = u(\alpha - \mathbf{1}_{\{u < 0\}}) = \begin{cases} \alpha u, & u \geq 0, \\ (1-\alpha)(-u), & u < 0. \end{cases}$$

The out-of-sample quantile score is the average:

$$\text{QL}_\alpha = \frac{1}{T} \sum_{t=1}^T \rho_\alpha(r_{t+1} - \hat{q}_{t,\alpha}).$$

Interpretation: when  $r_{t+1}$  falls below the forecast (a violation), the penalty is weighted by  $(1-\alpha)$ , which is large for small  $\alpha$ . Importantly,  $\text{QL}_\alpha$  is minimized when  $\hat{q}_{t,\alpha}$  equals the true conditional quantile. Therefore, among models with reasonable coverage, a smaller  $\text{QL}_\alpha$  indicates a more accurate VaR forecast.

### 3.3 Kupiec test

A standard statistical test for VaR calibration is the Kupiec (1995) *proportion-of-failures* test. Let

$$X = \sum_{t=1}^T I_{t+1}^{(\alpha)} \quad \text{and} \quad \hat{\pi}_\alpha = \frac{X}{T}.$$

Under correct unconditional calibration, violations should behave like  $X \sim \text{Binomial}(T, \alpha)$ . The Kupiec test compares:

$$H_0 : \mathbb{P}(I_{t+1}^{(\alpha)} = 1) = \alpha \quad \text{vs} \quad H_1 : \mathbb{P}(I_{t+1}^{(\alpha)} = 1) \neq \alpha.$$

The likelihood ratio statistic is

$$LR_{\text{uc}} = -2 \log \left( \frac{(1-\alpha)^{T-X} \alpha^X}{(1-\hat{\pi}_\alpha)^{T-X} \hat{\pi}_\alpha^X} \right).$$

Under  $H_0$ , one has approximately  $LR_{\text{uc}} \sim \chi^2(1)$ , which yields a  $p$ -value. In this project, we choose the decision rule  $\delta$  equal to 5%.

Kupiec's test checks whether the overall violation rate is correct by comparing the  $p$ -value to  $\delta$ .

### 3.4 Composite score for model selection

During hyperparameter tuning, we need a single criterion to rank candidate models across the three confidence levels  $\alpha \in \{0.01, 0.05, 0.10\}$ . We therefore minimize a composite score that balances the accuracy of the quantile forecasts through the pinball loss  $QL_\alpha$  and the calibration of the VaR through a penalty on the coverage deviation  $|\hat{\pi}_\alpha - \alpha|$ :

$$S(\lambda) = \frac{1}{|\mathcal{A}|} \sum_{\alpha \in \mathcal{A}} (QL_\alpha + \lambda |\hat{\pi}_\alpha - \alpha|), \quad \mathcal{A} = \{0.01, 0.05, 0.10\}.$$

The parameter  $\lambda > 0$  controls the trade-off: a larger  $\lambda$  puts more emphasis on matching the nominal violation rate (coverage), while a smaller  $\lambda$  focuses more on pure quantile accuracy. In our grid search, we set  $\lambda = 0.1$  and select the hyperparameters that minimize  $S(\lambda)$  on the validation period.

## 4 GARCH model

We estimate a standard GARCH(1,1) model to capture time-varying volatility in daily returns. In this framework, returns are written as

$$r_t = \mu + \sigma_t z_t,$$

where  $\sigma_t$  is the conditional volatility and  $z_t$  follows a given distribution  $F$ . The choice of this distribution is important because the VaR corresponds to a lower-tail quantile of  $r_{t+1}$ .

To assess whether a Normal distribution is appropriate, we analyze the standardized residuals  $\hat{z}_t = \hat{\varepsilon}_t / \hat{\sigma}_t$ .

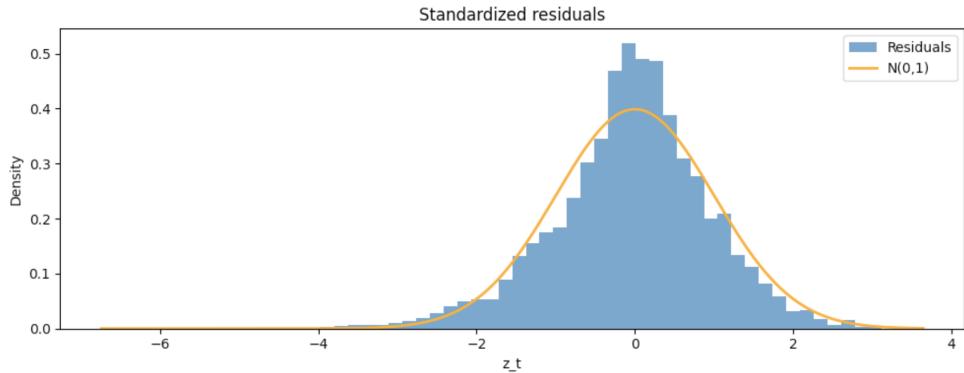


Figure 3: Histogram of standardized residuals with Normal density overlay.

The histogram shows that the empirical distribution is more peaked around zero and exhibits heavier tails than the Normal curve. In particular, extreme negative values appear more frequently than predicted by a Gaussian distribution.

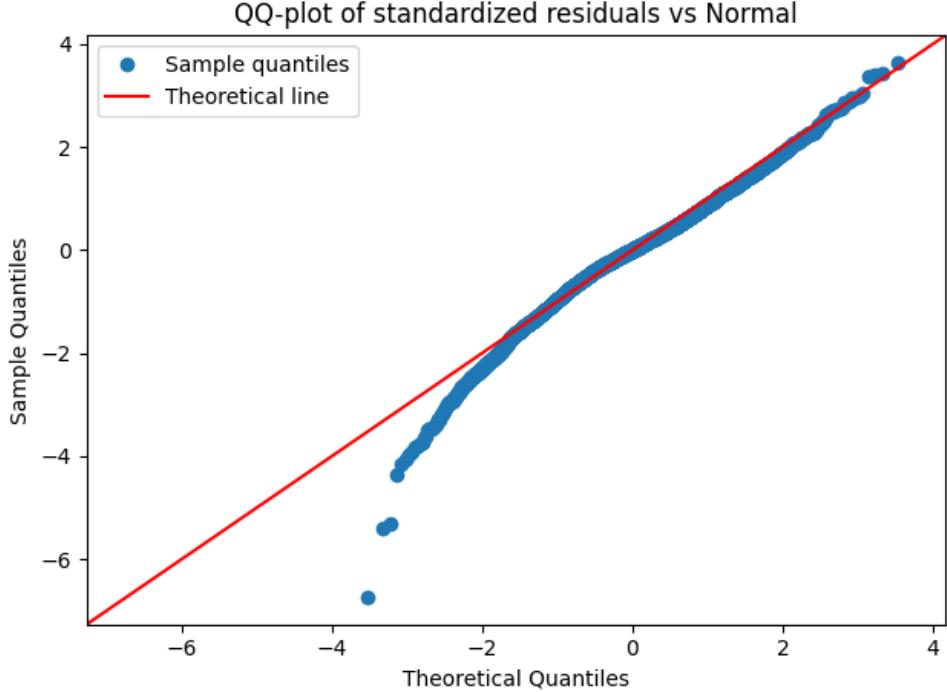


Figure 4: QQ-plot of standardized residuals versus the Normal distribution.

The QQ-plot confirms this result. While the residuals follow the straight line in the central region, they deviate significantly in the tails, especially on the left side. This indicates the presence of fat tails. Therefore, a Student-t distribution is more appropriate than a Normal distribution for modeling the innovation term in the GARCH model, which is particularly important for accurate VaR estimation.

#### 4.1 VaR forecasts

In a GARCH(1,1) model, as mentioned earlier, returns are written as

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim F,$$

where  $\mu$  is the conditional mean,  $\sigma_t$  is the conditional volatility, and  $F$  is the distribution of the standardized innovations  $z_t$ . Based on Figures 3 and 4, we take  $F$  to be Student-t.

The parameter  $\mu$  is estimated on the training sample and is assumed constant in time in our model, hence its estimate  $\hat{\mu}$  does not carry an index  $t$ . The time variation comes from the one-step-ahead volatility forecast  $\hat{\sigma}_{t+1|t}$ , which is updated every day using the GARCH recursion.

After estimation, we compute the filtered residual

$$\hat{\varepsilon}_t = r_t - \hat{\mu},$$

and update the conditional variance forecast via GARCH(1,1) recursion:

$$\hat{\sigma}_{t+1|t}^2 = \omega + \alpha \hat{\varepsilon}_t^2 + \beta \hat{\sigma}_t^2.$$

Let  $q_\alpha(F)$  denote the  $\alpha$ -quantile of  $z_{t+1}$  under  $F$ . Since  $r_{t+1} = \mu + \sigma_{t+1} z_{t+1}$ , the conditional  $\alpha$ -quantile of  $r_{t+1}$  is

$$Q_\alpha(r_{t+1} | X_t) = \mu + \sigma_{t+1} q_\alpha(F).$$

Therefore, VaR forecast at level  $\alpha$  is

$$\widehat{\text{VaR}}_{t+1,\alpha} = \hat{\mu} + \hat{\sigma}_{t+1|t} q_\alpha(\hat{F}),$$

where  $q_\alpha(\hat{F})$  is the Student-t quantile implied by the fitted degrees of freedom.

## 4.2 Parameter tuning

To select a well-calibrated GARCH for VaR forecasting, we perform a grid search on the training period. For each candidate model, we generate VaR forecasts for  $\alpha \in \{0.01, 0.05, 0.10\}$  and rank candidates using the composite score  $S(\lambda)$  introduced in Section 3.4.

We explore the following hyperparameters:

- (i) GARCH orders  $(p, q) \in \{1, 2\} \times \{1, 2\}$ ,
- (ii) innovation distribution  $F \in \{\text{Normal}, \text{Student-}t\}$ ,
- (iii) estimation window:  $W \in \{\text{None}, 500, 1000, 1500\}$  trading days.

The best configuration on the validation set is obtained for

$$(p, q) = (1, 1), \quad F = \text{Student-}t, \quad W = 1500,$$

with validation score  $S(\lambda) = 0.0916$ . As expected, using Student-t innovations yields better results than the Normal distribution.

## 4.3 Results

Table 1 reports the final test-set backtesting results of the selected GARCH model. All three confidence levels pass the Kupiec unconditional coverage test at the 5% significance level, and the overall composite score on the test set is 0.0884. Figure 5 illustrates the  $\widehat{\text{VaR}}_{t+1, 0.05}$  forecasts on the test set. Violations (red markers) correspond to days where  $r_{t+1} < \widehat{\text{VaR}}_{t+1, 0.05}$ . The observed violation frequency is slightly above the nominal 5% level (5.39%), indicating a slight underestimation of downside risk.

$\alpha$	Exp. cov.	Act. cov.	$QL_\alpha$	Kupiec LR	Kupiec p-value	Result
1%	1.0%	1.29%	0.0272	1.11	0.2917	PASS
5%	5.0%	5.39%	0.0914	0.44	0.5077	PASS
10%	10.0%	10.57%	0.1453	0.49	0.4838	PASS

Table 1: Test-set VaR backtesting results for the selected GARCH model  $(p, q) = (1, 1)$  with Student- $t$  innovations and rolling window  $W = 1500$ .

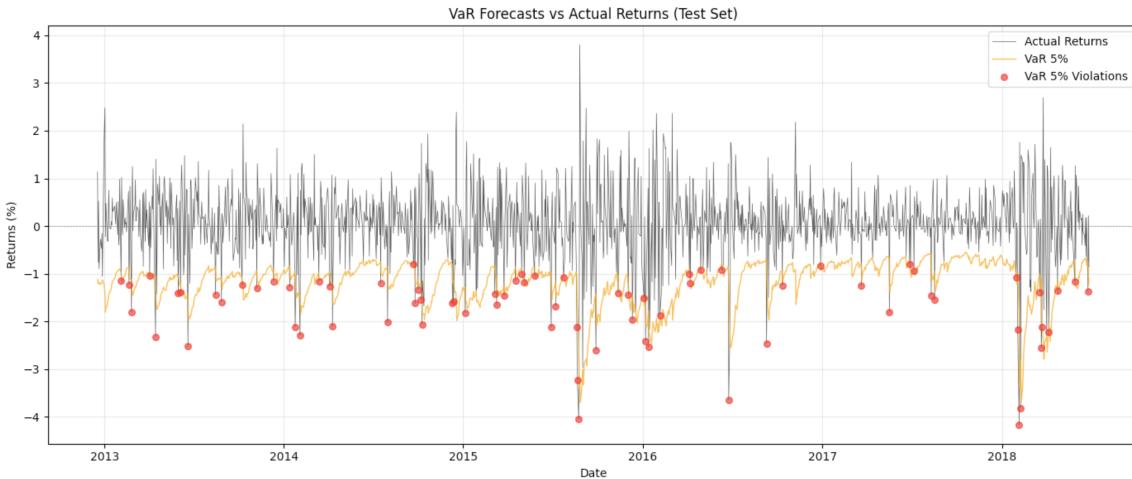


Figure 5: VaR forecasts versus realized returns on the test set (with alpha=5%).

## 5 Feature engineering

Since Section 2 shows strong persistence in volatility, we design a small set of interpretable features that summarize volatility dynamics at different horizons and capture tail-risk episodes.

### 5.1 Volatility features

Realized volatility  $\text{rv5}_t$  is strictly positive and typically right-skewed, so we work with its logarithm to obtain a more stable signal. As we can observe in Figure 6.

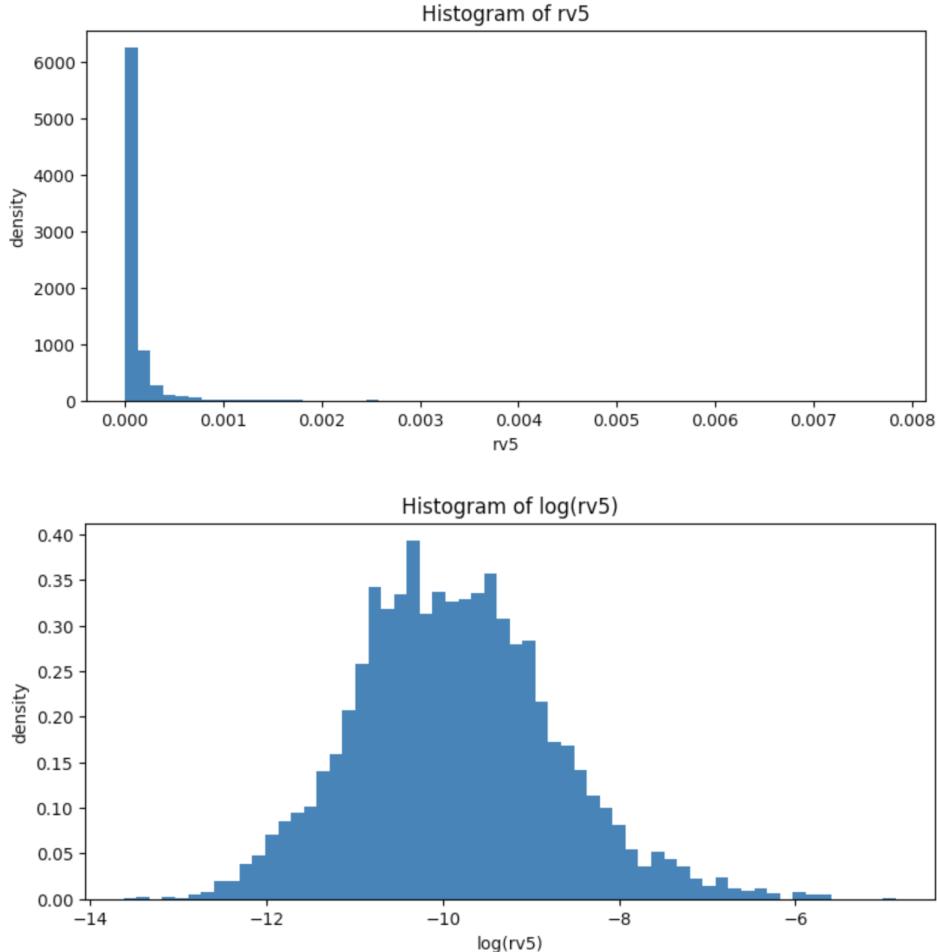


Figure 6: Histograms of  $\text{rv5}$  (top) and  $\log(\text{rv5})$  (bottom).

To reflect the slow decay of autocorrelation observed in  $\log(\text{rv5})$ , we construct features that approximate short (daily), medium (weekly), and long-run (monthly) volatility levels:

$$x_t^{(d)} = \log(\text{rv5}_t), \quad x_t^{(w)} = \frac{1}{5} \sum_{j=0}^4 \log(\text{rv5}_{t-j}), \quad x_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} \log(\text{rv5}_{t-j}).$$

These three variables provide a parsimonious summary of volatility clustering without introducing many parameters, which helps avoid overfitting.

### 5.2 Jump proxy

Realized volatility can spike because of jumps[3]. Bipower variation  $\text{bv}_t$  is less sensitive to jumps, so the difference  $\text{rv5}_t - \text{bv}_t$  isolates discontinuous moves. We therefore define a non-negative

jump indicator:

$$J_t = \max(\text{rv5}_t - \text{bv}_t, 0).$$

## 6 HAR-RV model

Section 2.2 showed that  $\log(\text{rv5}_t)$  has a slowly decaying autocorrelation, which is a typical *long-memory* pattern in volatility. The HAR-RV model (Heterogeneous Autoregressive model of Realized Volatility) captures this persistence in a parsimonious way by combining volatility components computed over multiple horizons (daily, weekly, monthly)[1].

### 6.1 HAR-RV specification

The HAR-RV(3) model then forecasts next-day log realized volatility as

$$\log(\text{rv5}_{t+1}) = \beta_0 + \beta_d x_t^{(d)} + \beta_w x_t^{(w)} + \beta_m x_t^{(m)} + \varepsilon_{t+1}. \quad (1)$$

This is an AR-type model reparameterized through economically meaningful horizon averages, which allows it to reproduce long-memory behavior with only a few coefficients.

Equation (1) is estimated by linear regression, since all regressors are directly observable at the end of day  $t$ .

### 6.2 VaR forecasts

The HAR-RV model produces a forecast of next-day realized volatility:

$$\widehat{\text{rv5}}_{t+1} = \exp\left(\widehat{\beta}_0 + \widehat{\beta}_d x_t^{(d)} + \widehat{\beta}_w x_t^{(w)} + \widehat{\beta}_m x_t^{(m)}\right).$$

We convert it into a volatility scale via  $\widehat{\sigma}_{t+1} = \sqrt{\widehat{\text{rv5}}_{t+1}}$  and then forecast VaR by specifying the  $\alpha$ -quantile of standardized innovations  $z_{t+1} = r_{t+1}/\widehat{\sigma}_{t+1}$ :

$$\widehat{\text{VaR}}_{t+1,\alpha} = \widehat{\sigma}_{t+1} q_\alpha,$$

where  $q_\alpha$  is either the Normal quantile  $\Phi^{-1}(\alpha)$  or the Student- $t$  quantile fitted on training standardized returns.

### 6.3 Parameter tuning

We perform a grid search and we explore the following hyperparameters:

- (i) horizon choices for the HAR components:  $(w, m) \in \{(5, 22), (10, 22), (5, 63), (10, 63)\}$ ,
- (ii) inclusion of a jump proxy: `include_jump`  $\in \{\text{False}, \text{True}\}$ , where the jump measure is  $J_t = \max(\text{rv5}_t - \text{bv}_t, 0)$ ,
- (iii) regularization: ridge regression with  $\alpha_{\text{reg}} \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ ,
- (iv) innovation quantile method: `q_method`  $\in \{\text{normal}, \text{t}\}$ .

The best configuration on the validation set is obtained for

$$w = 10, \quad m = 63, \quad \text{include\_jump} = \text{True}, \quad \alpha_{\text{reg}} = 10^{-3}, \quad \text{q\_method} = \text{t},$$

with validation score  $S(\lambda) = 0.0022$ . This indicates that using longer horizons and a jump proxy, together with Student- $t$  quantiles for innovations, improves VaR calibration and quantile accuracy.

## 6.4 Results

Table 2 reports the final test-set VaR backtesting results for the selected HAR-RV configuration. All three confidence levels pass the Kupiec unconditional coverage test at the 5% significance level. Empirical coverages are slightly above the nominal levels for  $\alpha = 1\%$  and  $\alpha = 5\%$ , suggesting a mild underestimation of downside risk, while the  $\alpha = 10\%$  coverage is very close to target. The overall composite score on the test set is 0.0014.

$\alpha$	Exp. cov.	Act. cov.	QL $_\alpha$	Kupiec LR	Kupiec p-value	Result
1%	1.0%	1.60%	$3.06 \times 10^{-4}$	2.12	0.1453	PASS
5%	5.0%	5.82%	$8.56 \times 10^{-4}$	0.93	0.3345	PASS
10%	10.0%	10.19%	$1.351 \times 10^{-3}$	0.03	0.8690	PASS

Table 2: Test-set VaR backtesting results for the selected HAR-RV model with  $(w, m) = (10, 63)$ , jump proxy enabled, ridge regularization  $\alpha_{\text{reg}} = 10^{-3}$ , and Student- $t$  innovation quantiles.

## 7 Quantile regression

To forecast Value-at-Risk (VaR), we can directly model conditional quantiles of next-day returns instead of modeling the conditional mean and variance [4].

We can estimate VaR by fitting a quantile regression model:

$$Q_\alpha(r_{t+1} | X_t) = \beta_{0,\alpha} + \beta_\alpha^\top X_t.$$

The parameter  $(\beta_{0,\alpha}, \beta_\alpha)$  is obtained by minimizing the quantile loss

$$(\hat{\beta}_{0,\alpha}, \hat{\beta}_\alpha) = \arg \min_{\beta_0, \beta} \sum_t \rho_\alpha(r_{t+1} - \beta_0 - \beta^\top X_t), \quad \rho_\alpha(u) = u(\alpha - \mathbf{1}_{\{u < 0\}}).$$

This loss is such that where  $r_{t+1}$  falls below the fitted quantile, it penalize differently from where  $r_{t+1}$  is above the fitted quantile, which forces the fitted line to target the  $\alpha$ -quantile. After estimation, the VaR forecast at level  $\alpha$  is simply

$$\widehat{\text{VaR}}_{t+1,\alpha} = \hat{\beta}_{0,\alpha} + \hat{\beta}_\alpha^\top X_t.$$

By fitting separate models for  $\alpha = 1\%, 5\%, 10\%$ , we obtain VaR forecasts at the required confidence levels.

### 7.1 Parameter tuning

We perform a grid search on the training period by exploring the following hyperparameters:

- (i) rolling estimation window size:  $W \in \{500, 1000\}$  trading days,
- (ii) rebalancing frequency:  $K \in \{21, 63, 123\}$ ,
- (iii) ridge regularization strength:  $\alpha_{\text{reg}} \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ .

The best configuration on the validation set is obtained for

$$W = 500, \quad K = 21, \quad \alpha_{\text{reg}} = 0.01,$$

which corresponds to refitting the model monthly (every 21 trading days) using the most recent 500 observations, with moderate ridge regularization to limit overfitting.

## 7.2 Results

Table 3 reports the final test-set VaR backtesting results for the selected quantile regression. All three confidence levels pass the Kupiec unconditional coverage test at the 5% significance level. The empirical coverages are slightly above the nominal levels, indicating a slight underestimation of downside risk, but the deviations remain small and statistically acceptable. The overall composite score on the test set is 0.0018

$\alpha$	Exp. cov.	Act. cov.	$QL_\alpha$	Kupiec LR	Kupiec $p$ -value	Result
1%	1.0%	1.19%	$3.57 \times 10^{-4}$	0.319	0.572	PASS
5%	5.0%	5.41%	$9.54 \times 10^{-4}$	0.321	0.571	PASS
10%	10.0%	10.82%	$1.50 \times 10^{-3}$	0.678	0.410	PASS

Table 3: Test-set VaR backtesting results for the selected quantile regression with rolling window  $W = 500$ , rebalancing every  $K = 21$  days, and ridge penalty  $\alpha_{\text{reg}} = 0.01$ .

## 8 Model comparison

We now compare the three VaR forecasting approaches on the *test set* using the same backtesting metrics reported in Tables 1, 2, and 3. The consolidated results are shown in Table 4.

### 8.1 Calibration

All three models pass the Kupiec unconditional coverage test at the 5% significance level for  $\alpha \in \{1\%, 5\%, 10\%\}$ . Empirical coverages are slightly above nominal levels overall, meaning that violations occur somewhat more often than expected, which indicates a mild underestimation of downside risk. A possible improvement is to strengthen the left-tail fit by allowing for asymmetric and heavier-tailed innovations to capture more skewness.

### 8.2 Accuracy

Comparing HAR-RV and quantile regression, HAR-RV achieves lower quantile loss at all confidence levels, suggesting sharper and more accurate VaR forecasts among these two models.

Model	$\alpha$	Exp. cov.	Act. cov.	$QL_\alpha$	Kupiec LR	Kupiec $p$ -value	Result
GARCH	1%	1.0%	1.29%	0.0272	1.11	0.2917	PASS
	5%	5.0%	5.39%	0.0914	0.44	0.5077	PASS
	10%	10.0%	10.57%	0.1453	0.49	0.4838	PASS
HAR-RV	1%	1.0%	1.60%	$3.06 \times 10^{-4}$	2.12	0.1453	PASS
	5%	5.0%	5.82%	$8.56 \times 10^{-4}$	0.93	0.3345	PASS
	10%	10.0%	10.19%	$1.351 \times 10^{-3}$	0.03	0.8690	PASS
Quantile regression	1%	1.0%	1.19%	$3.57 \times 10^{-4}$	0.319	0.572	PASS
	5%	5.0%	5.41%	$9.54 \times 10^{-4}$	0.321	0.571	PASS
	10%	10.0%	10.82%	$1.50 \times 10^{-3}$	0.678	0.410	PASS

Table 4: Test-set VaR backtesting comparison across models.

## 9 Conclusion

In this project, we studied Value-at-Risk forecasting for SPY log-returns at  $\alpha \in \{1\%, 5\%, 10\%\}$ . The preliminary analysis highlighted the usual stylized facts of financial markets with persistent volatility and a slowly decaying autocorrelation, suggesting multi-horizon dynamics in volatility.

We then implemented three VaR forecasting approaches. First, a GARCH(1,1) model with Student- $t$  standardized residuals captured time-varying volatility and heavy tails. Second, the HAR-RV model exploited the long-memory structure of realized volatility through daily/weekly/monthly components and achieved the best overall performance in our experiments. Finally, quantile regression provided a direct, flexible way to estimate conditional quantiles of returns from the same set of features. Its backtesting results were also satisfactory (Kupiec tests passed for all  $\alpha$ ), although slightly less competitive than HAR-RV in our composite ranking.

Overall, our results confirm that combining realized-volatility measures with multi-horizon predictors is particularly effective for short-horizon VaR forecasting. Possible extensions include engineering more robust features (e.g., richer jump proxies), and applying additional tests, like conditional coverage tests such as Christoffersen's test assess not only the correct violation rate but also the independence of violations [6]. Finally, with a larger dataset, it would be interesting to explore more flexible model classes, including deep learning approaches, to capture more complex nonlinear patterns.

## References

- [1] Fulvio Corsi (2002). A Simple Long Memory Model of Realized Volatility.
- [2] Sanel Halilbegovic, Mia Vehabovic (2016). Backtesting Value at Risk Forecast: the Case of Kupiec Pof-Test. *Russian Federation European Journal of Economic Studies* Vol.(17).
- [3] Fulvio Corsi, Davide Pirino, Roberto Renò (2012). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*
- [4] Robert F. Engle, Simone Manganelli (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles.
- [5] T. Gneiting, A. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation.
- [6] Y. Zhang, S. Nadarajah (2017). A review of backtesting for value at risk. *Communications in Statistics - Theory and Methods*