# Risk Quantification in a Low-Dimensional Portfolio

Georg Khella, Applied Statistics, Project 5

May 2025

**Abstract**

This report investigates the estimation and evaluation of Value-at-Risk (VaR) using a semi-parametric approach that combines GARCH models with Extreme Value Theory (EVT). Daily log-losses of an equally weighted portfolio of three US stocks (GE, KO, MMM) are analyzed over the period 2000–2024. An AR(0)–GARCH(1,1) model is fitted to the loss series, and the tail of standardized innovations is modeled using a Generalized Pareto Distribution. Out-of-sample one-step-ahead VaR forecasts are computed through a rolling window scheme and validated using binomial backtesting procedures. Results indicate that the model provides reliable VaR estimates, with violation frequencies statistically consistent with the nominal confidence levels.

## 1 Introduction

This report examines the use of univariate time series modelling and Extreme Value Theory to estimate and evaluate Value-at-Risk for a financial portfolio composed of three large-cap U.S. companies listed in the Dow Jones Industrial Average (DJIA): General Electric (GE), Coca-Cola (KO), and 3M (MMM). The analysis uses daily adjusted stock prices from January 3, 2000, to December 31, 2024, divided into a calibration phase (2000–2018) and a testing phase (2018–2024). The aim is to capture the tail risk of portfolio losses using a semi-parametric approach combining GARCH models for conditional volatility and EVT for rare events.

The portfolio assumes fixed weights: 55% in GE, 25% in KO, and 20% in MMM. We focus on the time series of portfolio log-losses, computed as a weighted average of the log-losses of the individual assets.

## 2 Dataset Description

The dataset contains four elements: a matrix of daily stock prices, a vector of portfolio weights, and two date intervals defining the calibration and testing periods. The calibration period spans from January 3, 2000, to July 29, 2018, and the testing period from August 2, 2018, to December 31, 2024. The stock prices include over 6,000 daily observations for each asset. In this section, we restrict the analysis to the calibration interval.

# 3 Exploratory Analysis of Portfolio Log-Losses

The logarithmic loss of the portfolio at time $t$ is computed as the weighted average of the individual assets' log-losses. Formally, we define:

$$\ell_{p,t} = \sum_{i=1}^{3} w_i \ell_{i,t} = \sum_{i=1}^{3} w_i (\log P_{i,t-1} - \log P_{i,t}) \tag{1}$$

where $w_i$ represents the fixed portfolio allocation for asset $i$, and $P_{i,t}$ denotes the adjusted closing price of asset $i$ at trading day $t$. This formulation corresponds to the negative logarithmic return, a standard measure of financial loss used in risk modelling.

Figure 1 illustrates the time evolution of daily log-losses for the portfolio during the calibration period from 2000 to 2018. The plot highlights several stylized features typical of financial time series. Most notably, we observe volatility clustering: periods of relative calm alternate with intervals of elevated variability, particularly evident during the 2008–2009 financial crisis. This non-constant variance structure suggests the presence of heteroskedasticity, motivating the use of GARCH-type models in subsequent analysis. Additionally, the presence of sharp negative spikes reflects the heavy-tailed nature of loss distributions, where extreme events are more likely than in a Gaussian framework.

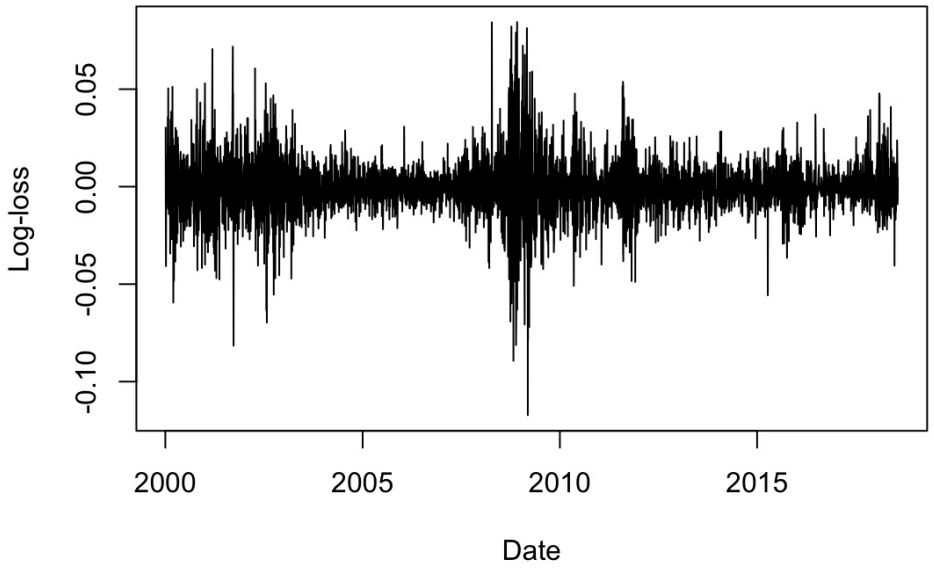



Figure 1: Daily log-losses of the portfolio during the calibration period (2000–2018)

## 3.1 Summary Statistics

Descriptive statistics of the log-loss series:

- Mean: $-4.47 \times 10^{-5}$, defined as $\mu = \frac{1}{n}\sum_{t=1}^{n} \ell_t$
- Median: $-1.70 \times 10^{-4}$
- Standard deviation: 0.0139, computed as $\sigma = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(\ell_t - \mu)^2}$
- Skewness: 0.11, where Skew $= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\ell_t - \mu}{\sigma}\right)^3$
- Kurtosis: 6.50, where Kurt $= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\ell_t - \mu}{\sigma}\right)^4$

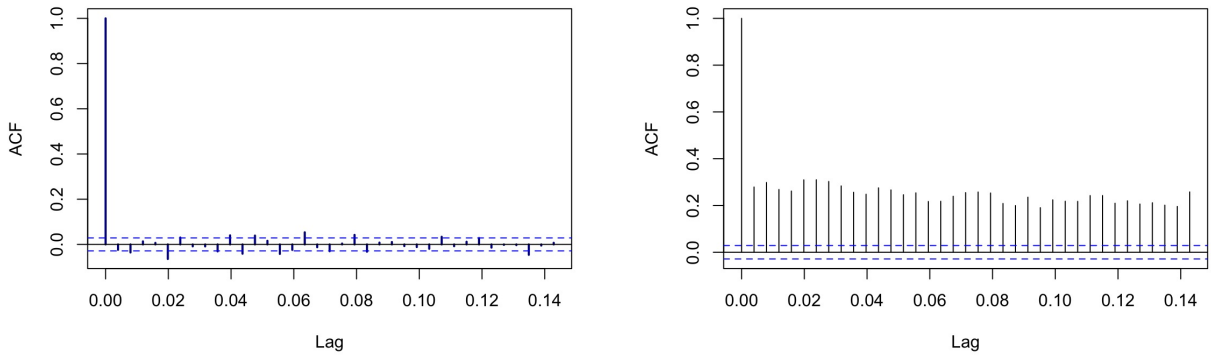
These values indicate a slightly asymmetric distribution with heavy tails and high excess kurtosis.

## 3.2 Autocorrelation Analysis

Autocorrelation analysis is a standard tool in time series modelling to assess the presence of temporal dependence. In the context of financial data, it is well known that raw returns (or losses) often exhibit negligible autocorrelation, while their absolute or squared values tend to show significant persistence — a phenomenon associated with volatility clustering.

Figure 2a shows the autocorrelation function (ACF) of the portfolio log-losses. The lack of statistically significant autocorrelations beyond lag zero confirms that the log-loss series behaves like a white noise process in the mean, i.e., individual losses are largely uncorrelated over time. This is consistent with the efficient market hypothesis, which posits that price changes should be unpredictable and not serially dependent.

In contrast, Figure 2b presents the ACF of the absolute log-losses. Here, we observe a slow decay of autocorrelations across several lags, indicating the presence of strong temporal dependence in the magnitude of losses. This confirms volatility clustering: large losses tend to be followed by large losses (of either sign), and small losses by small ones. Such dynamics are typical in financial time series and motivate the use of GARCH-type models that explicitly model conditional variance.




(a) Autocorrelation of log-losses

(b) Autocorrelation of absolute log-losses

Figure 2: Correlograms of the portfolio's log-loss time series during the calibration period

### 3.3 Ljung–Box Tests

To formally assess the presence of autocorrelation in the time series, we apply the Ljung–Box test, which belongs to the family of portmanteau tests. These tests are designed to detect deviations from randomness by examining whether a group of autocorrelations are jointly different from zero. The original version was introduced by Box and Pierce [1] and later refined by Ljung and Box [2] to improve performance in finite samples.

Given a time series $\{x_t\}$, the Ljung–Box test evaluates the joint null hypothesis that the first $h$ autocorrelations are all equal to zero. The test statistic is computed as:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k} \tag{2}$$

where $n$ is the sample size and $\hat{\rho}_k$ is the sample autocorrelation at lag $k$. Under the null hypothesis of no autocorrelation, $Q$ asymptotically follows a chi-squared distribution with $h$ degrees of freedom.

We apply the test to both the log-loss series and its absolute value using a lag order of 10. The resulting p-values are:

- Log-losses: $p = 1.43 \times 10^{-6}$
- Absolute log-losses: $p < 2.2 \times 10^{-16}$

The very small p-value for the raw log-losses leads us to reject the null hypothesis of no autocorrelation, despite the visual impression of white noise. This discrepancy may be due to the sensitivity of the test in large samples. More importantly, the result for the absolute log-losses confirms significant autocorrelation in the magnitude of returns, in line with the earlier correlogram analysis. This reinforces the presence of conditional heteroskedasticity.

## 4 Model Fitting and Diagnostics

To model the conditional volatility of the portfolio's log-losses, we adopt the **Autoregressive Generalized Autoregressive Conditional Heteroskedasticity** framework, commonly denoted as AR($k$)–GARCH($p$,$q$). The GARCH model, first introduced by [3], extends the ARCH model of [4] by incorporating lagged conditional variances, allowing it to capture persistent volatility clustering — a typical feature of financial time series.

Let $r_t$ denote the portfolio log-loss at time $t$. In the AR–GARCH setup:

$$r_t = \mu + \sum_{i=1}^{k}\phi_i r_{t-i} + \epsilon_t, \tag{3}$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim \mathcal{N}(0,1), \tag{4}$$

$$\sigma_t^2 = \omega + \sum_{j=1}^{q}\alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^{p}\beta_i \sigma_{t-i}^2, \tag{5}$$

where $\sigma_t^2$ is the conditional variance of the process, $\omega > 0$, $\alpha_j, \beta_i \geq 0$, and the condition $\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1$ is typically imposed to ensure weak stationarity of the variance.

The parameter $k$ refers to the order of the autoregressive (AR) component in the mean equation: it models the dependence of the current log-loss on its own past values. In contrast, the parameter $q$ controls the number of lagged squared innovations (i.e., past shocks) included in the conditional variance equation, capturing the short-term impact of past volatility. The parameter $p$ determines the number of lagged conditional variances, accounting for the persistence of volatility over time. The special case with $k = 0$ reduces the model to a pure GARCH process with no autoregressive dynamics in the mean.

We estimate and compare three models: AR(0)–GARCH(1,1), AR(1)–GARCH(1,1), and AR(2)–GARCH(1,1). Table 1 reports the information criteria for each model.

Model selection is based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), two standard penalized-likelihood criteria used to balance goodness of fit and model complexity. Given a model with log-likelihood $\mathcal{L}$, number of parameters $k$, and number of observations $n$, these criteria are defined as:

$$\begin{aligned} \text{AIC} &= -2\mathcal{L} + 2k, \\ \text{BIC} &= -2\mathcal{L} + k \log n. \end{aligned}$$

Lower values indicate a better trade-off between fit and parsimony. While AIC tends to favor more complex models, BIC penalizes model complexity more heavily and is consistent under regularity conditions.

Table 1: Information Criteria for Competing AR–GARCH Models

| Model | AIC | BIC |
|---|---|---|
| AR(0)–GARCH(1,1) | **-6.1315** | **-6.1260** |
| AR(1)–GARCH(1,1) | -6.1311 | -6.1242 |
| AR(2)–GARCH(1,1) | -6.1310 | -6.1227 |

Although the differences are small, the AR(0) model has the lowest values of both AIC and BIC. Furthermore, the autoregressive coefficients in the other models are statistically insignificant:

- **AR(1)**: $\phi_1$ has $p = 0.817$
- **AR(2)**: $\phi_1$ has $p = 0.86$, $\phi_2$ has $p = 0.12$

This suggests that the log-loss series does not benefit from including autoregressive terms in the conditional mean equation.

All models produce strongly significant GARCH coefficients. In the selected AR(0)–GARCH(1,1) model:

- $\alpha_1 = 0.0636$, $\beta_1 = 0.9296$
- Persistence: $\alpha_1 + \beta_1 = 0.9932$

indicating a high level of volatility persistence, typical of financial return series.

Diagnostic tests on standardized residuals further support the model's adequacy:

- Ljung–Box on residuals: $p = 0.1245$
- Ljung–Box on squared residuals: $p = 0.8559$

These confirm that the AR(0)–GARCH(1,1) adequately captures both the autocorrelation and conditional variance structure. We therefore retain this model for the subsequent residual-based analysis using Extreme Value Theory.

## 4.1 Residual Diagnostics

To assess model adequacy, we analyze standardized residuals from the selected AR(0)–GARCH(1,1) model. Figure 3 displays the time series of standardized residuals. The residuals appear homoskedastic and roughly centered around zero, with occasional spikes, which is expected in financial return series.

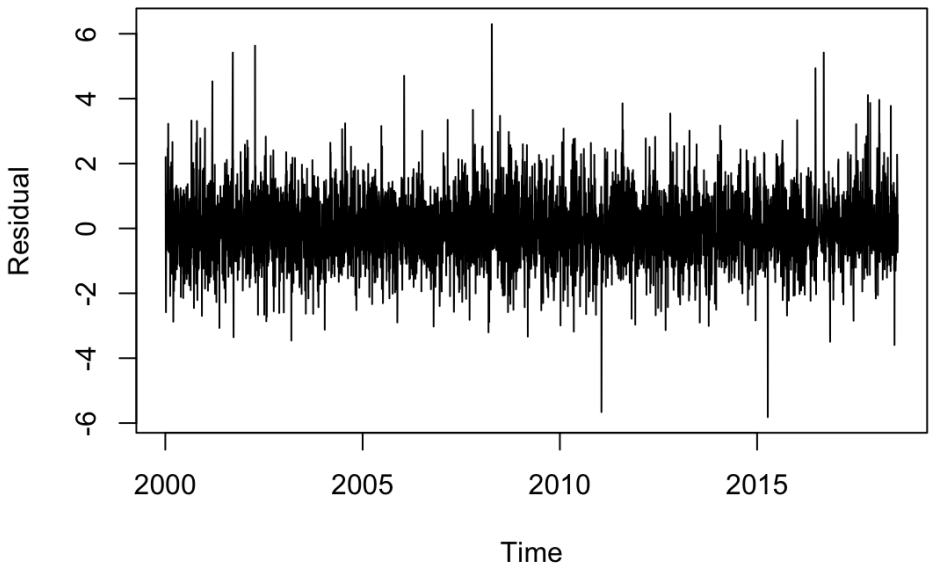



Figure 3: Standardized residuals from AR(0)–GARCH(1,1) model

Next, we investigate the normality of the residuals.

To assess the assumption of normality, we rely on both graphical and statistical tools. The Q–Q plot compares the empirical quantiles of the standardized residuals to those of a standard normal distribution. Under normality, the points should lie approximately on a 45-degree line; systematic deviations indicate departures from normality, such as skewness or heavy tails.

In addition, we apply the Jarque–Bera (JB) test [6], which formally tests whether a sample's skewness and kurtosis match those of a normal distribution. The test statistic is given by:

$$\mathrm{JB} = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right), \tag{6}$$

where $n$ is the sample size, $S$ is the sample skewness, and $K$ the sample kurtosis. Under the null hypothesis of normality, the JB statistic asymptotically follows a chi-squared distribution with 2 degrees of freedom.

Figure 4 presents the Q–Q plot of standardized residuals against the theoretical quantiles of the standard normal distribution. Deviations from the 45-degree reference line, particularly in the tails, indicate that residuals exhibit excess kurtosis. This is confirmed by the **Jarque–Bera test** [6], which strongly rejects the null hypothesis of normality ($p < 2.2 \times 10^{-16}$).

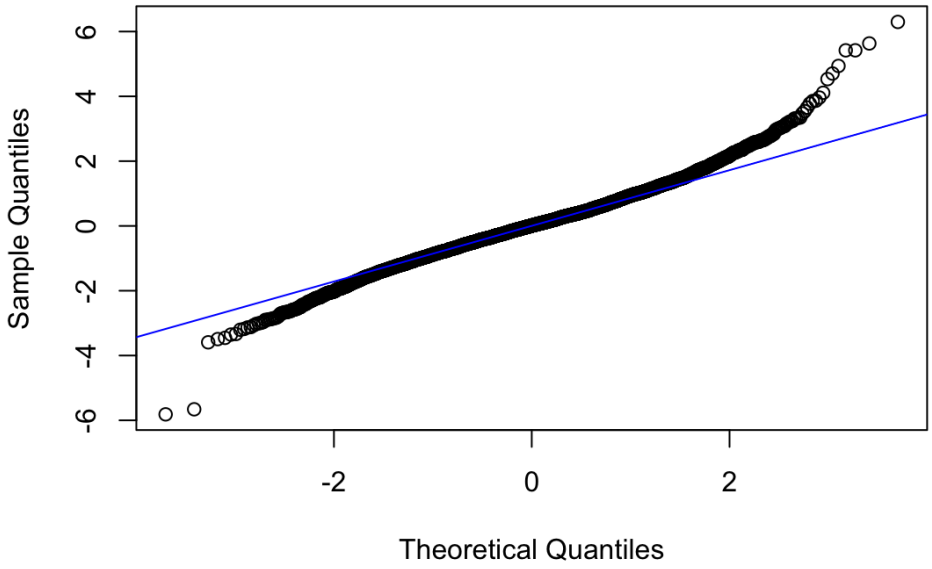



Figure 4: Q–Q plot of standardized residuals from AR(0)–GARCH(1,1)

These findings highlight a limitation of the normal-GARCH model: while it captures volatility dynamics well, it fails to accommodate heavy-tailed behavior. In the next section, we address this by applying EVT to model the distribution of extreme losses.

# 5 Extreme Value Theory for Tail Modelling

Despite the adequacy of the AR(0)–GARCH(1,1) model in capturing the conditional heteroskedasticity of the portfolio log-losses, diagnostic tests reveal non-normality in the stan-

dardized residuals, especially in the tails. To better capture extreme losses, we apply Extreme Value Theory (EVT), which provides a rigorous framework for modeling tail risks.

Specifically, we use the *Peaks-Over-Threshold* (POT) method, which approximates the distribution of exceedances above a high threshold $u$ by a Generalized Pareto Distribution (GPD), under general conditions [7]. Formally, let $X$ be a random variable with cumulative distribution function $F$. For a sufficiently high threshold $u$, the conditional distribution of the excesses $Y = X - u \mid X > u$ converges to the GPD as $u \to \infty$:

$$\lim_{u \to x_F} \mathbb{P}(X - u \leq y \mid X > u) = G(y), \tag{7}$$

where $x_F = \sup x : F(x) < 1$ is the right endpoint of the support of $F$, and $G(y)$ is the GPD.

The GPD cumulative distribution function is given by:

$$G(y) = 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-1/\xi}, \quad y > 0, \quad \text{for } \xi \neq 0, \tag{8}$$

where $\xi$ is the shape parameter, $\beta > 0$ is the scale, and $y = z_t - u$ is the excess.

The POT approach is particularly suited to financial time series, which often display heavy tails, volatility clustering, and skewness — features inadequately captured by Gaussian models [8]. Combining GARCH for conditional volatility with EVT for tail behavior yields a more accurate semi-parametric model for the innovations: empirical CDF on $(-\infty, u]$, and GPD on $(u, +\infty)$.

## 5.1 Threshold Selection

A key step in EVT modeling is the choice of threshold $u$, which must balance bias and variance. A low threshold results in bias due to poor GPD approximation, while a high threshold leads to high variance in parameter estimates due to few exceedances.

To inform this choice, we consider two graphical diagnostics:

- The **Mean Excess Plot** examines the average excess over thresholds. A roughly linear trend suggests the validity of a GPD model beyond that threshold.
- The **Shape Parameter Stability Plot** plots the estimated GPD shape parameter $\xi$ across a range of thresholds. A plateau in this plot indicates a region of stable tail behavior.

These diagnostics reveal a stable region for $u$ in the range of the 90th to 97th percentiles. We select $u = \text{quantile}(z_t, 0.95)$, corresponding to approximately 1.65, as it falls within the stable region of $\xi$ estimates and ensures a sufficient number of exceedances (12) for reliable estimation.

## 5.2 GPD Fit and Diagnostics

We fit a Generalized Pareto Distribution to the exceedances over $u = 1.65$ using maximum likelihood estimation. The shape parameter $\xi$ governs the heaviness of the tail, with $\xi > 0$ indicating a heavy tail, while the scale parameter $\beta$ controls the dispersion of the exceedances. In our estimation, we obtain:

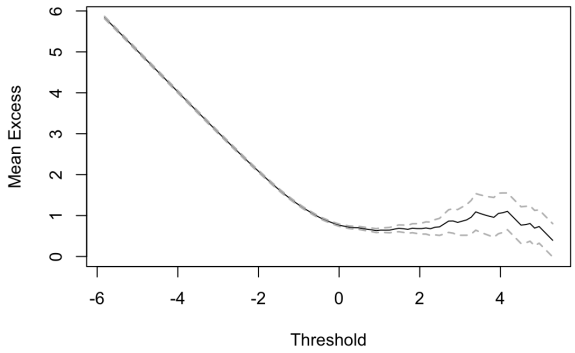



Figure 5: Mean Excess Plot for standardized residuals

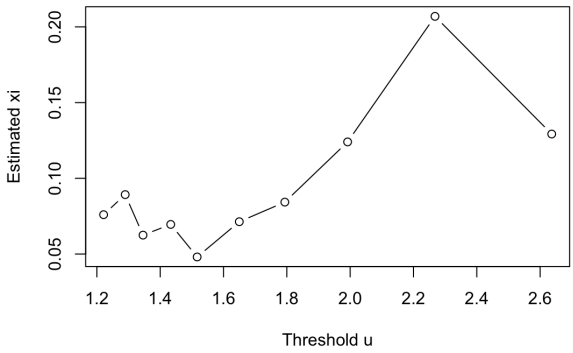



Figure 6: Shape parameter estimates $\xi$ across thresholds

- Shape parameter $\hat{\xi} \approx 0.071$
- Scale parameter $\hat{\beta} \approx 0.626$

To assess goodness-of-fit, we examine the following diagnostic plots:

- The **Quantile-Quantile Plot** compares empirical quantiles of exceedances with theoretical GPD quantiles.
- The **Histogram with Fitted Density** shows the empirical distribution overlaid with the GPD fit.

These diagnostics confirm a good fit of the GPD model to the data above the selected threshold.

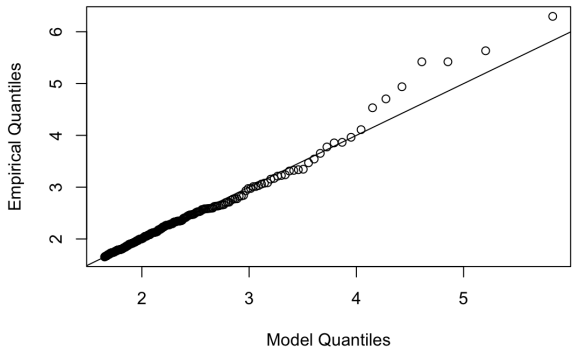



Figure 7: QQ plot of exceedances over $u = 1.65$ against GPD

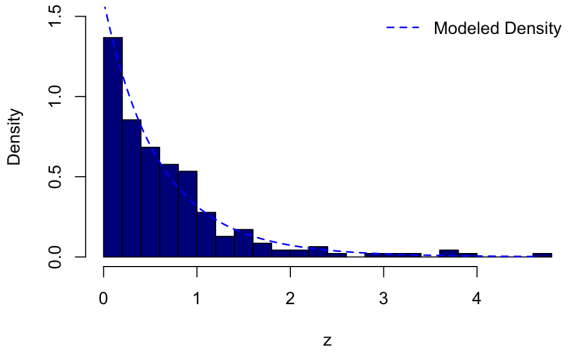



Figure 8: Histogram of exceedances with GPD fit

## 5.3 Semi-Parametric Quantile Estimation

To estimate quantiles of the standardized innovations, we construct a semi-parametric approximation of their distribution. Following the Peaks-Over-Threshold (POT) framework, the distribution function $\hat{F}_u(x)$ is defined as:

$$\hat{F}_u(x) = \begin{cases} \frac{1}{n}\sum_{j=1}^{n} I(x_j \le x), & x \le u \\ 1 - \frac{n_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, & x > u \end{cases} \tag{9}$$

where $n$ is the sample size, $n_u = \#\{j : x_j > u\}$ is the number of exceedances over the threshold $u$, $\hat{\xi}$ and $\hat{\beta}$ are the maximum likelihood estimates of the shape and scale parameters from the GPD fit, and $I(\cdot)$ is the indicator function.

This construction yields a semi-parametric estimator that combines the empirical distribution function for values below the threshold $u$ and a Generalized Pareto tail above it, as discussed in [7, 8].

To estimate quantiles above the threshold, we invert the GPD tail approximation in Eq. (9). For a given level $\alpha$, the estimated $\alpha$-quantile $v_\alpha$ satisfies $\hat{F}u(v\alpha) = \alpha$, leading to the expression:

$$v_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\frac{1-\alpha}{\hat{F}(u)}\right)^{-\hat{\xi}} - 1\right], \tag{10}$$

where $\hat{F}(u) = n_u/n$ is the empirical exceedance probability.

Using our estimated parameters and threshold $u = 1.65$, we obtain:

- $v_{0.95} \approx 1.65$
- $v_{0.99} \approx 2.72$

These semi-parametric quantiles are essential for computing the in-sample Value-at-Risk. Given that the standardized innovations $z_t$ follow the distribution $\hat{F}_u$, and under the assumption of a zero conditional mean ($\mu_t = 0$), the one-step-ahead in-sample Value-at-Risk at level $\alpha$ is computed as:

$$\text{VaR}_{\alpha,t} = -\sigma_t \cdot v_\alpha, \tag{11}$$

where $\sigma_t$ is the conditional standard deviation obtained from the AR(0)-GARCH(1,1) model, and $v_\alpha$ is the semi-parametric innovation quantile estimated above. The assumption $\mu_t = 0$ is justified by the selection of an AR(0) model, where the conditional mean component is constant and statistically negligible, hence excluded from the risk forecasting formula.

Applying this to the entire calibration period, we compare the log-loss observations with the computed $\text{VaR}_{\alpha,t}$ thresholds. The number of exceedances observed is:

- 204 exceedances over $\text{VaR}_{0.95}$ out of 4671 observations
- 26 exceedances over $\text{VaR}_{0.99}$ out of 4671 observations

These exceedances will be benchmarked in the backtesting analysis of the next section to assess the out-of-sample predictive performance of the VaR estimates.

# 6 Out-of-Sample VaR Forecasting

Having estimated the semi-parametric quantiles of standardized innovations using EVT, we now turn to the computation of one-step-ahead Value-at-Risk (VaR) forecasts over the testing period (August 2018–December 2024). The goal is to evaluate the out-of-sample predictive performance of the AR(0)–GARCH(1,1) + EVT model, with dynamic refitting based on a rolling calibration window.

## 6.1 Methodology

We adopt a recursive rolling forecast procedure. At each time step $t$ in the testing period, an AR(0)–GARCH(1,1) model is fitted to the most recent $n = 4682$ log-loss observations, corresponding to the length of the original calibration period. The standardized residuals from this model are used to estimate the upper tail of the innovation distribution via a Generalized Pareto Distribution (GPD). The following steps are then performed:

1. Extract the standardized residuals $z_{t-n+1}, \ldots, z_t$ from the fitted model.

2. Estimate the conditional tail using the GPD applied to exceedances above the 95th percentile.

3. Compute the semi-parametric quantiles $v_{0.95}, v_{0.99}$ using the inverted GPD formula:

$$v_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left( \left( \frac{1-\alpha}{\hat{F}(u)} \right)^{-\hat{\xi}} - 1 \right), \tag{12}$$

   where $u$ is the threshold, $\hat{\beta}$ and $\hat{\xi}$ are the MLE estimates of the GPD parameters, and $\hat{F}(u)$ is the empirical exceedance probability.

4. Obtain the forecast for conditional standard deviation $\hat{\sigma}_{t+1}$ and conditional mean $\hat{\mu}_{t+1}$ from the GARCH model.

5. Compute the one-step-ahead VaR as:

$$\text{VaR}_{\alpha,t+1} = -(\hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \cdot v_\alpha). \tag{13}$$

This procedure is repeated at each point in the testing set, resulting in a time series of daily VaR predictions at both 95% and 99% confidence levels. Note that, unlike in the in-sample setting where the GPD parameters were estimated once, in the rolling out-of-sample procedure both the threshold exceedances and the corresponding GPD parameters are re-estimated dynamically at each time step. This allows the model to adapt to evolving tail behavior in the standardized innovations.

### 6.2 Forecast Results

Figure 9 displays the daily portfolio log-losses in the testing period alongside the predicted VaR levels. The black line shows actual losses, while the dashed lines represent the $\text{VaR}_{0.95}$ (blue) and $\text{VaR}_{0.99}$ (green). Red crosses highlight violations—instances where the realized loss exceeded the predicted threshold.

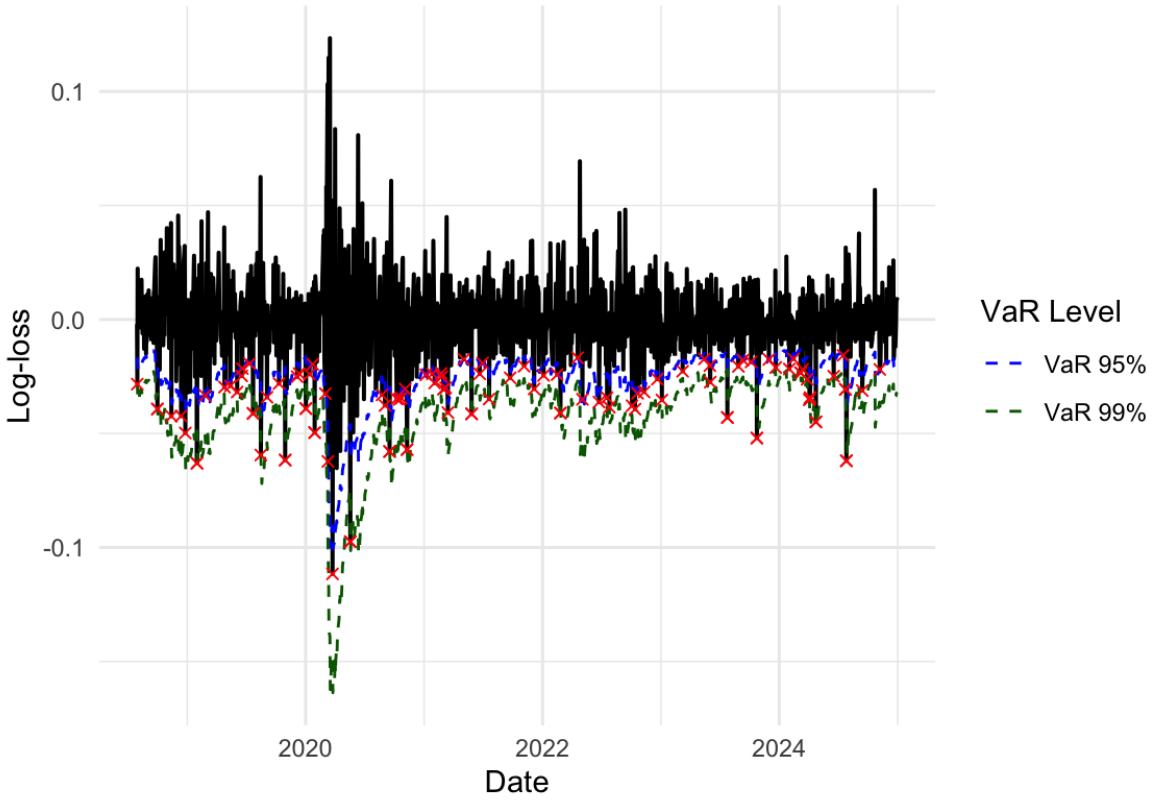



Figure 9: Out-of-sample log-losses and one-step-ahead VaR forecasts at 95% and 99% confidence levels.

The model captures most of the extreme losses in periods of high volatility, such as the early 2020 market turmoil. However, the presence of violations (especially for $\text{VaR}_{0.95}$) suggests room for improvement, which we analyze formally in the next section through statistical backtesting procedures.

# 7 Backtesting

To evaluate the predictive accuracy of the one-step-ahead VaR forecasts produced by the AR(0)–GARCH(1,1) + EVT model, we perform a formal backtesting analysis based on coverage and a binomial test. The goal is to assess whether the frequency of VaR violations observed in the out-of-sample period aligns with the nominal confidence levels of 95% and 99%.

## 7.1 Coverage Analysis

Out of 1616 out-of-sample observations (August 2018 to December 2024), we observe:

- **91** violations of the 95% VaR threshold (expected: 80.8),
- **18** violations of the 99% VaR threshold (expected: 16.16).

These observed frequencies are consistent with the theoretical expectations under the null hypothesis of correct model calibration.

## 7.2 Binomial Test

To formally test the adequacy of the VaR model, we perform a two-sided binomial test. Under the null hypothesis, the number of exceedances follows a Binomial$(n, \alpha)$ distribution, with $\alpha$ equal to 0.05 or 0.01 depending on the VaR level. Let $\ell_t$ denote the portfolio log-loss and $\text{VaR}_{\alpha,t}$ the predicted threshold at confidence level $1 - \alpha$. We define the violation indicator:

$$I_t = \begin{cases} 1 & \text{if } \ell_t > \text{VaR}_{\alpha,t} \\ 0 & \text{otherwise} \end{cases}$$

If the model is correctly specified, the sequence $\{I_t\}_{t=1}^{n}$ should consist of independent and identically distributed Bernoulli random variables with success probability $\alpha$. Accordingly, the total number of exceedances $X = \sum_{t=1}^{n} I_t$ should follow a binomial distribution $\text{Bin}(n, \alpha)$. The binomial test evaluates whether the observed value of $X$ is statistically consistent with this distribution [9].

The results are summarized below:

- For the 95% VaR: $p$-value = **0.253**, 95% CI: [0.0456, 0.0687],
- For the 99% VaR: $p$-value = **0.616**, 95% CI: [0.0066, 0.0175].

These results suggest that the observed number of violations is well within the range of statistical variation expected under the null hypothesis. In particular, the 95% confidence intervals indicate that the true exceedance probabilities are plausibly contained within the theoretical levels of 5% and 1%, respectively. The relatively high $p$-values further confirm that such deviations from the expected frequencies are not statistically significant and can be attributed to random fluctuation.

In both cases, the null hypothesis of correct coverage cannot be rejected at the 5% significance level. The observed frequency of violations is statistically compatible with the expected proportion of exceedances, supporting the validity of the model's predictive performance.

# 8 Conclusion

This report examined the estimation and evaluation of VaR using a semi-parametric model that integrates GARCH dynamics with EVT. By analyzing daily log-losses from a fixed-weight portfolio of three DJIA stocks over the period 2000–2024, we implemented a risk forecasting framework capable of capturing both conditional volatility and extreme tail behavior.

In-sample analysis showed that the AR(0)–GARCH(1,1) model successfully captures volatility clustering, while the use of a Generalized Pareto Distribution provided a good fit for the right tail of standardized innovations. The semi-parametric quantile estimation enabled us to compute time-varying VaR forecasts that reflect both volatility dynamics and extreme risks.

Out-of-sample forecasts were generated through a rolling window approach, and the predictive accuracy of the model was assessed using a binomial backtesting procedure. The results confirmed that the observed violation frequencies are statistically consistent with the nominal confidence levels, supporting the model's validity for practical risk management applications.

Future research could explore the use of time-varying thresholds for EVT, multivariate extensions to model joint risk factors, or comparisons with alternative distributional assumptions such as skewed or heavy-tailed innovations.

# References


[1] Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), 1509–1526.

[2] Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.

[3] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

[4] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987–1007.

[5] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.

[6] Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.

[7] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics.

[8] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

[9] Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, **3**(2), 73–84.