EXTREME VALUE ANALYSIS OF FINANCIAL DATA
PROJECT REPORT FOR STOR 834

AKSHAY SAKANAVEETI AND SOUMYAJYOTI KUNDU

1. Introduction

Modeling extreme events in financial markets is important for risk management. In this project, we apply methods from *Extreme Value Theory* (EVT) to analyze extreme events in the S&P 500 index. The S&P 500 tracks 500 major U.S. companies, reflecting overall market performance. S&P 500 index performance encapsulates the long history of market performance, making it a natural choice for studying rare and impactful market crashes.

2. Data and Objective

We consider the S&P 500 daily closing prices from January 2, 1985 to April 22, 2025. There are a total of $N = 10154$ observations, denoted by $X_1, X_2, \cdots, X_N$, where each X_n is the closing price on day n. We work with log returns, defined by

$$R_n = \log\left(\frac{X_n}{X_{n-1}}\right),$$

because they stabilize the variance. The returns are also additive over time, and better capture relative changes in price. Our focus is on analyzing rare and extreme crashes (i.e the left tail distribution) of R_n .

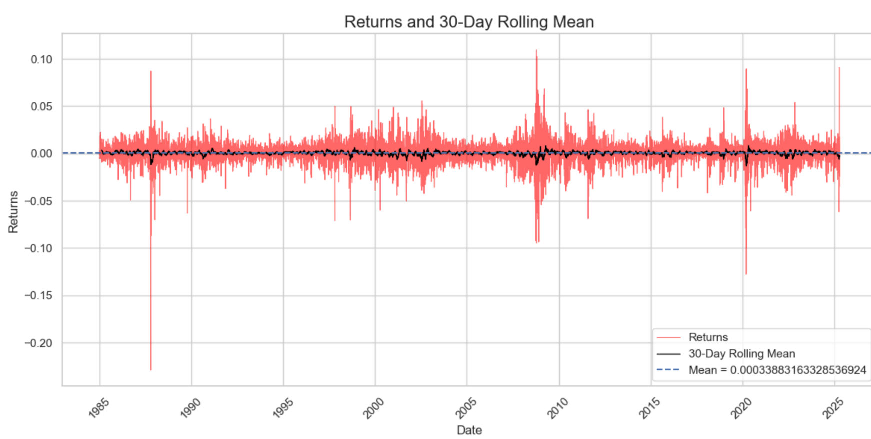


Figure 2.1. Log-Returns with time

Department of Statistics and Operations Research, 304 Hanes Hall, University of North Carolina, Chapel Hill, NC 27599

E-mail addresses: sakshay@unc.edu, kundus@unc.edu.

3. Overview of analysis

Since the returns do not satisfy the $i.i.d$ assumption, one cannot expect GPD models to fit well. One reason for the violation of iid assumption is because volatility clustering induces conditional heteroskedasticity, violating the independence assumption critical for Extreme Value Theory. We follow the approach proposed by [1], which addresses these issues in two steps:

(i) **Time Series analysis:** Fit a GARCH(p, q) model to the returns
(ii) **Extreme Value analysis:** Fit a GPD model to the standardized residuals of GARCH model in step (i).

4. Time Series Analysis

In this section, we briefly introduce GARCH model and fit a GARCH model to the returns, $\{R_n : 1 \le n \le N\}$.

The returns is observed to have "Volatility clustering" - big moves (up or down) are often followed by more big moves — and calm periods tend to follow calm periods. GARCH models captures this behavior well.

Definition 4.1 (GARCH(p, q))**.** *We say time series $Y_1, Y_2, \cdots$ follows GARCH(p, q) if*

$$Y_n = \mu + \sigma_n \epsilon_n, \qquad \sigma_n^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{n-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{n-j}^2$$

where ϵ_n are i.i.d mean 0 random variables, $\alpha_i, \beta_j \ge 0$ and $\omega > 0$.

4.1. **Fitting the GARCH model:** In general, **GARCH(1,1)** is often considered effective. However, to find out the choice of p, q that fits the data well, we follow the approach outlined below:

(i) Analyze **ACF and PACF** of **squared returns**:
  - PACF cutoff suggests **ARCH (p)**
  - ACF tailing off suggests **GARCH (q)**

(ii) Compare models using:
  - **AIC**, **BIC** (lower is better)

The following plots of ACF and PACF of R_n^2 suggests that we need to compare models with GARCH(p, q) models with $1 \le p, q \le 5$.

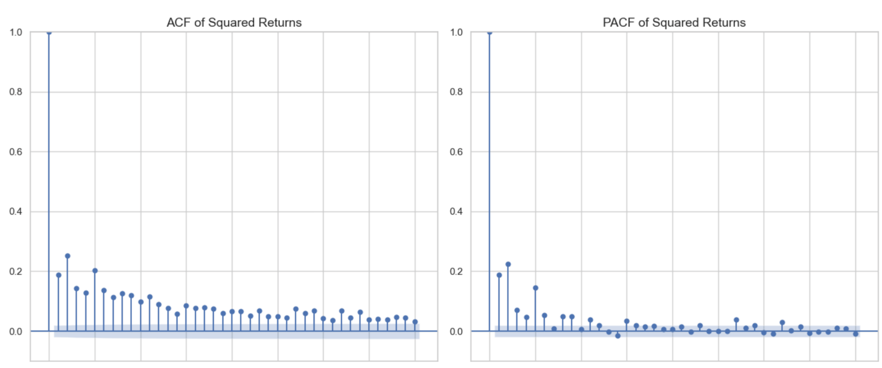


Figure 4.1. Plots of ACF and PACF of R_n^2.

Following are AIC and BIC plots for GARCH(p, q) models with $1 \le p, q \le 5$.

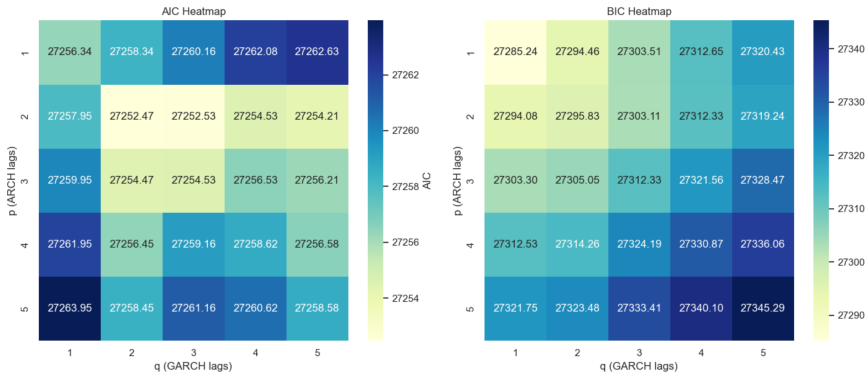


FIGURE 4.2. AIC and BIC heatmaps

Considering the scale of AIC and BIC values, it appears that all the models have similar performance. GARCH(1,1) is known to adequately capture volatility dynamics in financial time series, and further complexity does not significantly improve AIC/BIC scores here. Therefore, we use the GARCH$(1,1)$ model, which has four parameters given by

$$R_n = \mu + \sigma_n \, \epsilon_n \quad \text{where} \quad \sigma_n^2 = \omega + \alpha_1 \epsilon_{n-1}^2 + \beta_1 \sigma_{n-1}^2.$$

We first fit the model with $\epsilon_i \sim \mathcal{N}(0,1)$ and then analyze if the normality assumption holds. We use the *maximum likelihood estimator* (MLE) to estimate the parameters. Following are estimates, their standard errors and their significance level p-values.

Parameter	**Estimate**	**Std. Error**	**p-value**
$\hat{\mu}$	0.0649	8.325e-03	6.647e-15
$\hat{\omega}$	0.0219	4.969e-03	1.073e-05
$\hat{\alpha}_1$	0.114	1.776e-02	3.550e-10
$\hat{\beta}_1$	0.08729	1.798e-02	0.000

All parameters seem significant. We next focus our attention on the normality assumption.

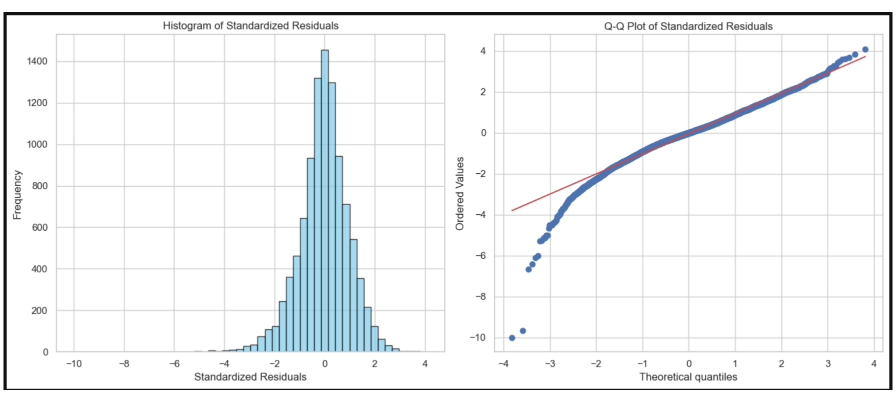


FIGURE 4.3. Histogram and Q-Q plot of standardized residuals

The Q-Q plot suggests that the tails of standardized residuals are heavy. One common suggestion to fix this issue is use $t-$ distributed ϵ's.

We now fit the model with $\epsilon_n \sim t_\nu$ distribution. We again estimate parameters using MLE.

Parameter	Estimate	Std. Error	p-value
$\hat{\mu}$	3.3883e+04	7243.760	2.903e-06
$\hat{\omega}$	2.6740e+10	7.964e+09	7.857e-04
$\hat{\alpha}_1$	0.1135	1.373e-02	1.417e-16
$\hat{\beta}_1$	0.8674	1.868e-02	0.000
$\hat{\nu}$	5.6969	0.392	7.544e-48

We use this GARCH model to obtain standardized residuals for our extreme value analysis.

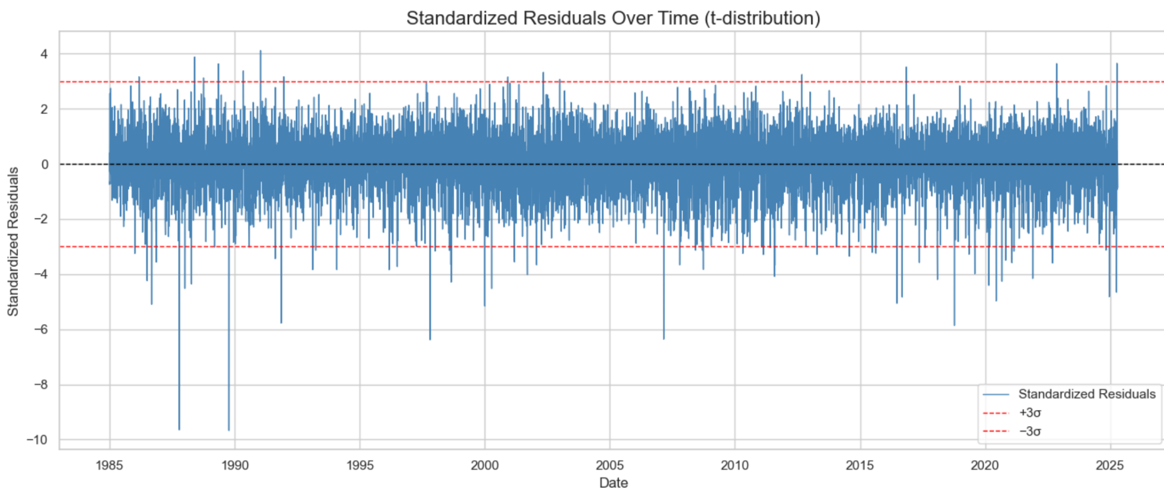


FIGURE 4.4. Standardized residuals

5. PRELIMINARY ANALYSIS

We use Peaks of Thresholds approach to fit the GPD models. We are interested in the crashes, which is the left tails of standardized residuals. For the preliminary analysis, we use a threshold of 0.01 quantile, and fit the GPD model. Choosing the 0.01 quantile ensures we have a sufficient number of exceedances for stable parameter estimation, while still focusing on the extreme left tail events. Following are MLEs of parameters in GPD model

Threshold (u)	Shape (ξ)	Scale (β)
-2.7831	0.2561	0.7333

This suggests that the tails are heavier than exponential, but not as extreme as power-law. Also, note that $\xi > 0$ implies the distribution has no left tails have no lower bounds— large events are possible. Following is the plot of

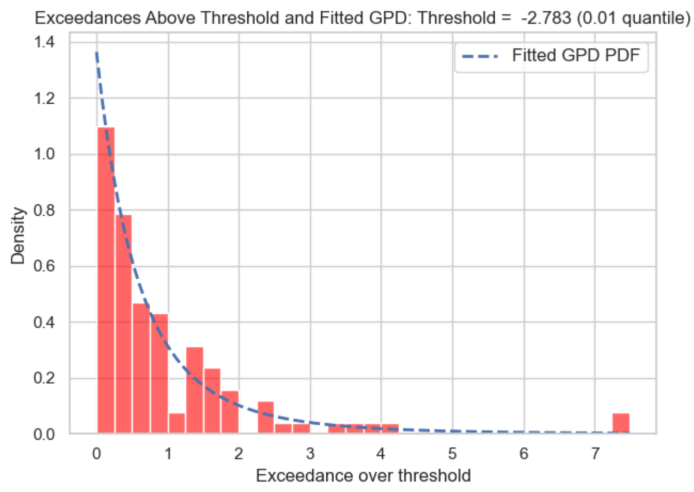


FIGURE 5.1. Exceedances below threshold

We next look at *VaR: Value at Risk* values. The following plot shows 95% and 99% VaR values. Recall the interpretation of VaR values, for example 99% VaR is interpreted as follows: in the worst 1% of cases, there is a 1% chance that the residual drops below -9.112.

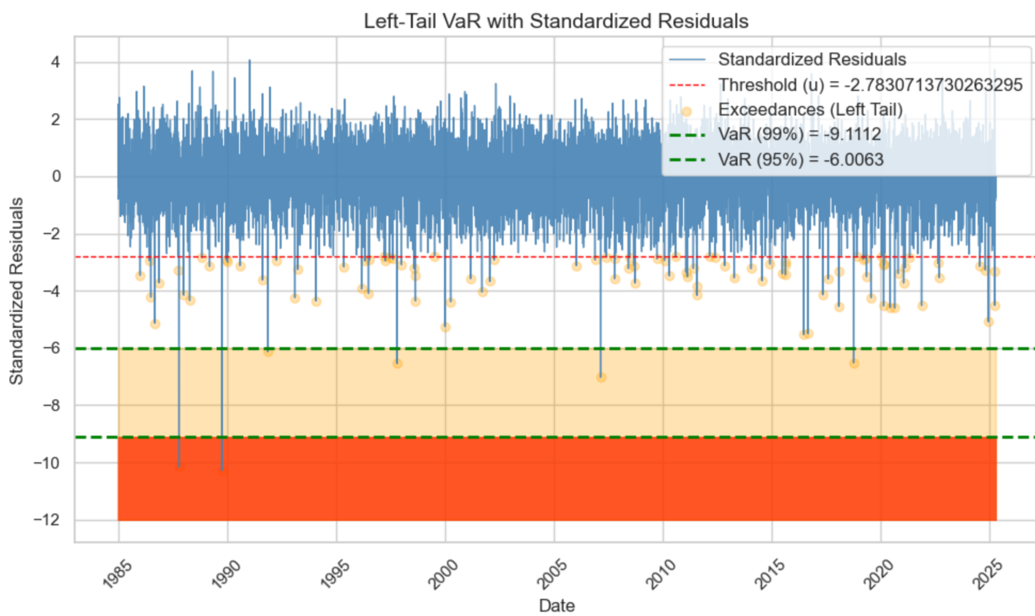


FIGURE 5.2. VaR plot with 0.01-quantile threshold

We note some extreme events in the above plot.

- **1987** – Black Monday crash (October 19, 1987), S&P 500 dropped about 20% in a single day.
- **2008** – Global Financial Crisis, major declines after Lehman Brothers bankruptcy (September 2008).
- **2020** – COVID-19 pandemic market crash (March 2020), extreme volatility and sharp sell-offs.

6. Extreme Value Analysis

We focus on the lower tail of the standardized residuals by fitting a Generalized Pareto Distribution (GPD), selecting the threshold (denote by u) via two methods namely - Mean excess plot which visually identifies the point above which the empirical mean excess is approximately linear and Weissman bootstrap-MSE which computes a bootstrap estimate of the MSE of the tail-index estimator over a grid of candidate thresholds and picks the one minimizing that MSE and then we compare the quantiles of the resulting fitted GPDs. To read about GPD see https://en.wikipedia.org/wiki/Generalized_Pareto_distribution.

The choice of threshold is important because, according to the Pickands-de Haan theorem, once the threshold u is sufficiently low, the excess distribution $u - X$ given $X < u$ can be approximated by a GPD. If u is too low, the GPD assumption is violated, leading to biased tail estimates. If u is too high, there are too few exceedances, making estimates noisy. Thus, we use two methods to choose optimal thresholds (i) Mean-Excess plots and (ii) Weissmann method.

6.1. Mean-excess plot.

The mean-excess plot helps identify the threshold beyond which the distribution of exceedances becomes approximately linear, indicating the optimal threshold for fitting extreme value models such as the GPD. We outline the steps below

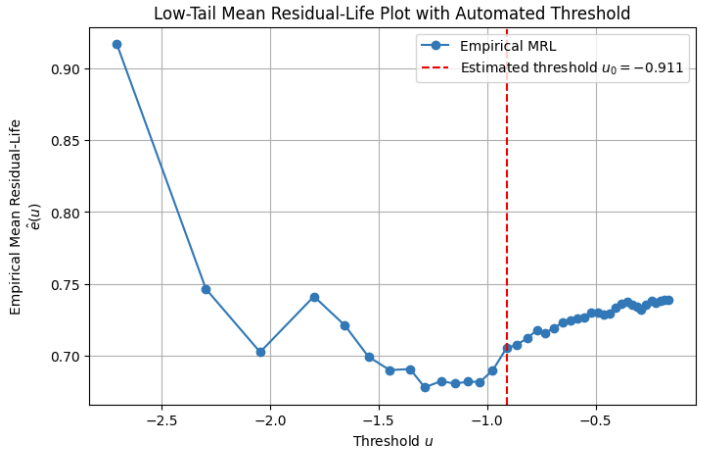
- **Threshold grid:** Select $u_1 < u_2 < \cdots < u_K$ to cover the plausible tail region—e.g. from the 50th up to the 90th percentile of your data.
- **Empirical mean excess:** For each u_j, let
$$n_j = \#\{x_i : x_i < u_j\},$$
$$\hat{e}(u_j) = \frac{1}{n_j} \sum_{i:x_i<u_j} (u_j - x_i).$$
Here $\hat{e}(u_j)$ is the average excess above u_j.
- **Plotting:** Draw the "mean-excess plot" by plotting the points
$$\big(u_j,\, \hat{e}(u_j)\big), \quad j = 1, \ldots, K.$$
Under a GPD tail model, this curve should be approximately linear in u.
- **Linearity criterion:** Scan the plot from left to right and identify the smallest u_j beyond which the points lie roughly on a straight line. That linearity indicates the regime where the Pickands–Balkema–de Haan theorem applies.
- **Threshold choice:** Set
$$u^* = \min\{\, u_j : \hat{e}(u) \text{ is approximately linear for all } u \geq u_j\}.$$
This u^* balances including enough data (to reduce variance) against entering the true Pareto tail (to avoid bias).

The mean-excess plot suggests that $u = -0.911$ is a good choice for the threshold. In Figure 6.1, the red line shows this threshold, and we can see that the plot becomes almost linear afterward.

Next, we find the MLE of GPD model with the above threshold. The estimates for parameters are

Threshold (u)	**Shape** (ξ)	**Scale** (σ)
-0.911	0.039	0.678

Table 6.1. Estimated GPD parameters and selected threshold.



H

FIGURE 6.1. Mean-excess plot

Finally we obtain the exceedances below the threshold plot. The shape parameter $\xi = 0.039$ indicates a slightly heavy-tailed distribution with *no finite lower bound* for the residuals beyond the threshold $u = -0.911$. Also, the QQ plot shows a good fit for most of the data, but some points at the higher end deviate, suggesting a lower threshold might better capture the extreme residuals.

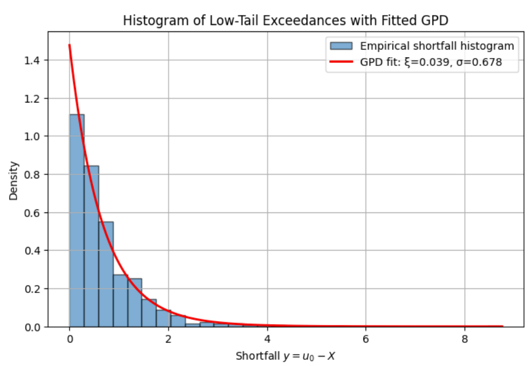


FIGURE 6.2. Fitted GPD

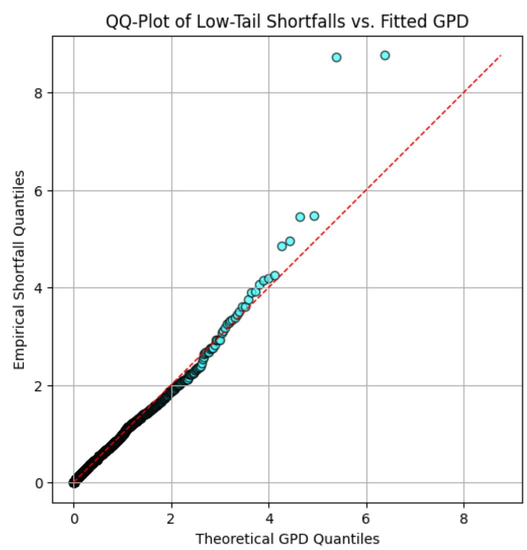


FIGURE 6.3. Q-Q-plot

6.2. **Weismann Bootstrap Mean Squared Error.** The Weismann Bootstrap Mean Squared Error (MSE) method provides a way to select an optimal threshold by minimizing the mean squared error of the tail-index estimator. This method relies on bootstrapping techniques to resample the exceedances above candidate thresholds and fit the Generalized Pareto Distribution (GPD) to each bootstrap sample. Below we describe steps in the implementation of the method.

(i) **Threshold grid:** Select a sequence of candidate thresholds spanning the tail region, say $u_1 < u_2 < \cdots < u_K$.
(ii) **Initial GPD fit:** For each $1 \leq k \leq K$, let $n_k = |\{i : x_i < u_k\}|$, i.e., the number of observations greater than u_k. For each k and $i \in \{i : x_i < u_k\}$, define the exceedances $Y_{i,k} = u_k - x_i$. Finally, for each k, obtain $\hat{\theta}_k =$ MLE estimate of the GPD shape parameter from $\{Y_{i,k}\}$.

(iii) **Bootstrap MSE computation:** For each u_k:

1. Resample its exceedances B times (with replacement) to get bootstrap samples $\{Y_{i,k}^{*(b)}\}_{b=1}^{B}$.
2. Fit the GPD to each bootstrap sample to obtain

$$\{\hat{\theta}_k^{*(1)}, \hat{\theta}_k^{*(2)}, \ldots, \hat{\theta}_k^{*(B)}\}.$$

3. Compute

$$\mathrm{MSE}(u_k) = \frac{1}{B}\sum_{b=1}^{B}\big(\hat{\theta}_k^{*(b)} - \hat{\theta}_k\big)^2.$$

(iv) **Threshold selection:** Choose

$$u^* = \arg\min_{1\le k\le K} \mathrm{MSE}(u_k),$$

i.e. the threshold that minimizes the bootstrap-estimated mean squared error of the tail-index estimator.

As before, we first select the threshold. The below plot suggest that the optimal threshold using Weissmann method is $u = -0.267$.

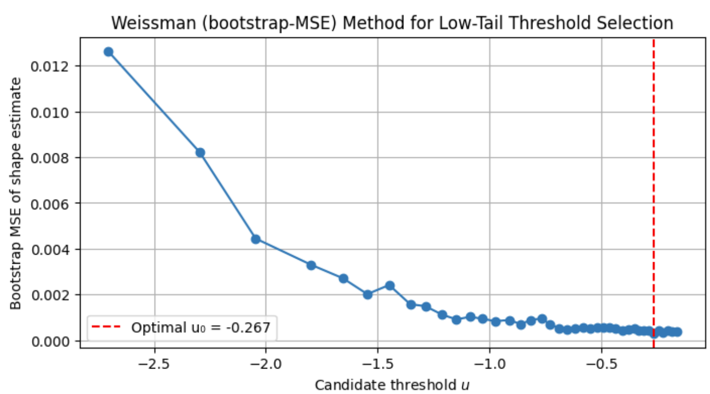


FIGURE 6.4. Threshold selection

For this threshold, the GPD estimates are

Threshold (u)	**Shape (ξ)**	**Scale (σ)**
-0.267	0.001	0.734

The shape parameter $\xi = 0.001$ suggests an almost exponential tail,with no lower bound. The QQ plot is similar to before, with a slight improvement, indicating a better fit for extreme values.

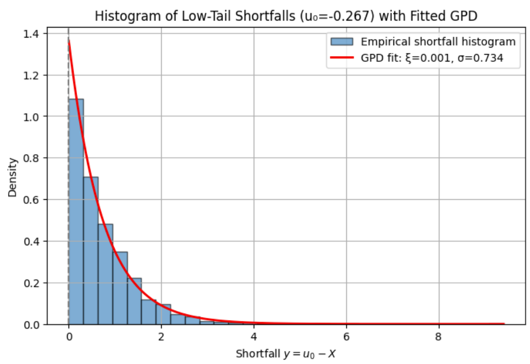


FIGURE 6.5. Fitted GPD

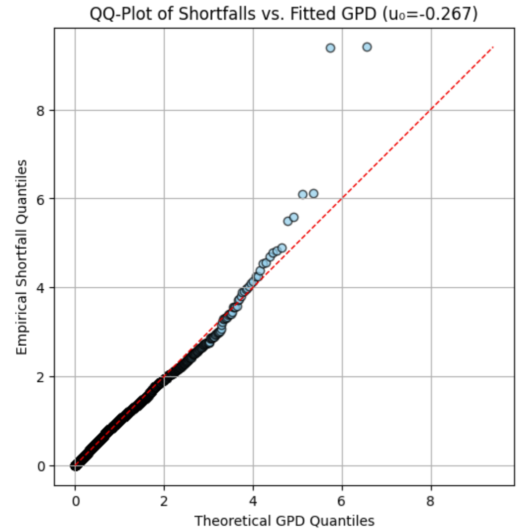


FIGURE 6.6. Q-Q plot

6.3. **Comparison of the quantiles.**

- **Mean-excess threshold:** With more exceedances, the Q-Q plot is smooth and dense in the mid-tail, but the highest points bend slightly above the 45° line, indicating a small bias.
- **Weissman bootstrap-MSE threshold:** Fewer exceedances lead to slightly more scatter in the mid-tail, but the top two or three points lie almost exactly on the 45° line, indicating very low bias.
- **Trade-off:** Mean-excess gives lower variance overall, Weissman gives lower bias at the extreme tail.

6.4. **Comparision of VaR values.** Earlier, we plotted VaR values in the preliminary analysis with a threshold of 0.01−quantile (see figure 5.2). The following two plots show VaR values for extremes with two thresholds given by (i) Mean-excess method and (ii) Weissman-Method.

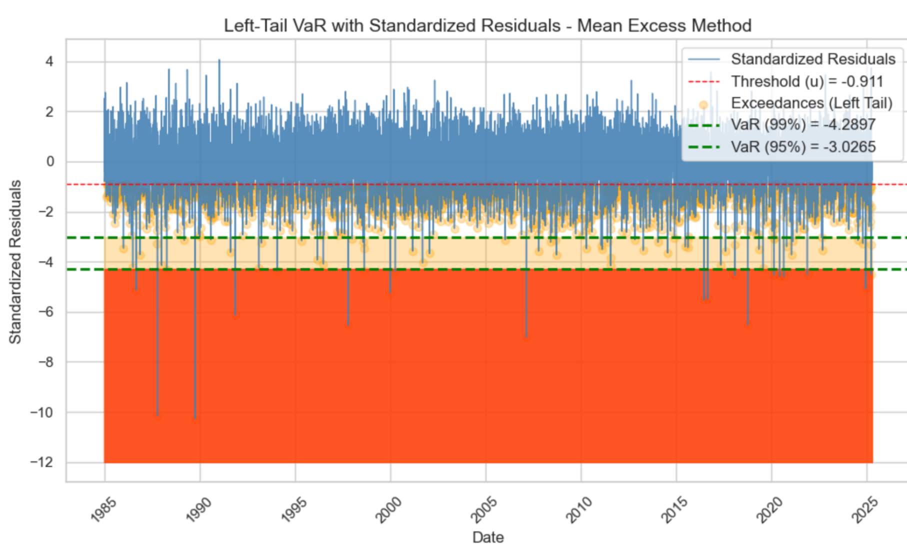


FIGURE 6.7. VaR with mean-excess threshold

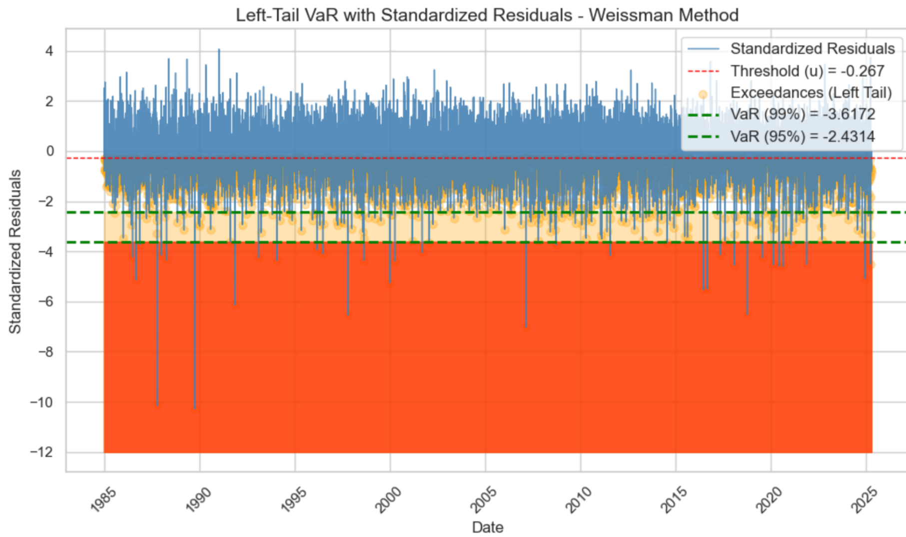


FIGURE 6.8. VaR with Weissman method threshold

The VaR values for the left tails are lower in the **mean-excess** method compared to the **Weissman** method, suggesting that the **mean-excess** method predicts smaller extreme losses in the lower tail.

References

[1] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: Concepts, techniques and tools - revised edition*, Princeton Series in Finance, Princeton University Press, 2015.