Iranian Journal of Finance

Print ISSN 2676-6337
Online ISSN 2676-6345

EVT-Based Value-at-Risk Estimation in Iranian financial markets Using the GUB Threshold Identification Method

Narges Ebadinezhad

Master's Student in Financial Management, Faculty of Financial Sciences, Kharazmi University, Tehran, Iran. (Email: nargesebadi7808@gmail.com)

Mohammad Reza Dehghani Ahmadabad*

*Corresponding Author, Assistant Prof., Faculty of Financial Sciences, Kharazmi University, Tehran, Iran. (Email: mr.dehghani@khu.ac.ir)

Iranian Journal of Finance, 2026, Vol. 10, No.2, pp. 1-31.
Publisher: Iran Finance Association
doi: https://doi.org/10.66224/ijf.2026.565753.1562
Article Type: Original Article
© Copyright: Author(s)
Type of License: Creative Commons License (CC-BY 4.0)

Received: January 13, 2026
Received in revised form: February 09, 2026
Accepted: March 07, 2026
Published online: April 01, 2026

Abstract

In classical risk management models, the assumption of normally distributed returns leads to an underestimation of the true probability of extreme losses. Extreme Value Theory (EVT), by focusing on tail behavior, provides an effective framework for modeling heavy-tailed risks; however, determining the threshold beyond which tail behavior begins remains one of its main challenges. In this study, the Good–Usual–Bad (GUB) method introduced by (González-Sánchez, 2021) and (González-Sánchez & Nave Pineda (2023) is employed to separate returns into central and extreme components, enabling the automatic identification of the tail threshold. Subsequently, Value-at-Risk (VaR) is estimated using the Hill estimator and the Generalized Pareto

Distribution (GPD). The dataset consists of daily returns for four asset classes in the Iranian market over the period 2015-2025: 18-carat gold, the Tehran Stock Exchange Overall Index (TEDPIX), the USD/IRR exchange rate, and Bitcoin in Iranian Rial. Empirical results indicate that, at the univariate level, the GUB method provides accurate estimates for heavy-tailed assets; however, at the portfolio level, its sensitivity to dependence structures makes it conservative and prone to overestimating risk. In contrast, dependence-based models such as the t-student copula yield more reliable portfolio VaR estimates, and hybrid approaches combining GUB with parametric models and the modified GUB-based procedures proposed in this study produce more balanced and efficient risk assessments.

Keywords: Extreme Value Theory (EVT), Value-at-Risk (VaR), GUB method, Hill estimator, t-student copula

JEL Classification: G21, C53, C45, C38

Introduction

The tail behavior of return distributions plays a crucial role in market risk assessment. Empirical studies have shown that financial returns generally exhibit heavier tails than the normal distribution, implying that the likelihood of large losses is higher than what classical models assume. Consequently, methods based on the normal distribution tend to underestimate risk, particularly during periods of market stress.

To address this issue, Extreme Value Theory (EVT) and in particular the Generalized Pareto Distribution (GPD), have been shown better results for modeling tail behavior (Jansen et al., 2000). By focusing on extreme observations, this approach can provide more accurate estimates of Value-at-Risk (VaR). The GPD is characterized by two key parameters: shape and scale, which represent the tail thickness and the dispersion of the tail, respectively. According to Brooks et al. (2005), these parameters are interrelated, meaning that one can be inferred from the other. The shape parameter is especially important for heavy-tailed distributions and is sometimes referred to as the tail index. Several methods exist to estimate the shape parameter, among which the Hill estimator (Hill, 1975) is recognized as one of the most effective according to Fedotenkov (2020).

The performance of EVT, however, heavily depends on the appropriate selection of the threshold. An incorrect threshold leads to unstable estimates of the tail index and scale parameter, undermining the reliability of the results. Moreover, as Koedijk et al. (1990) note, determining the appropriate number of extreme observations poses a trade-off: including more tail observations reduces variance but increases systematic bias, whereas considering fewer extreme points or focusing on the central part of the distribution reduces bias but greatly increases variance.

Various approaches have been proposed to determine the threshold. Some studies use fixed thresholds, such as the 95th percentile (Rhee & Wu, 2020). While simple, this method has a key limitation: it does not adapt to the characteristics of the actual data, potentially resulting in too few or too many extreme observations. Other studies adopt data-driven methods, such as bootstrap techniques (Németh & Zempléni, 2017), which can optimize the threshold by minimizing estimation error. However, these methods are computationally intensive and sensitive to large fluctuations, sometimes resulting in unstable performance. Alternative approaches, such as minimizing the Kolmogorov–Smirnov (KS) distance (Drees et al., 2018), aim to systematically select the optimal threshold for fitting the GPD. Although more objective than visual inspection, these approaches still rely on a limited number of extreme observations, making their results sensitive in financial series with unstable structures or large shocks. Recent studies further indicate that even widely used tail index estimators, such as the Hill estimator, remain sensitive to threshold selection, and much of the estimation error arises from misidentifying this point. Therefore, threshold determination remains a major challenge in applying EVT.

In this context, the Good–Usual–Bad (GUB) method introduced by (González-Sánchez and Nave Pineda (2023) provides a distinct approach. The method divides the return series into three segments: "good," "usual," and "bad." By iteratively removing the most extreme points until the central segment approximates a normal behavior, the tail threshold is automatically identified. This procedure enables a more appropriate selection of extreme observations, resulting in more stable and reliable estimates of the tail index and VaR.

In this study, the GUB method is applied to assess the risk of four asset classes in the Iranian market—18-carat gold, the Tehran Stock Exchange Overall Index (TEDPIX), the USD/IRR exchange rate, and Bitcoin (IRR)—over the period 2015-2024. After determining the threshold and estimating the tail index using the Hill estimator, VaR is calculated through the GPD distribution and compared with historical, parametric, and Student-t copula approaches. This analysis allows evaluation of the GUB method's performance under emerging-market conditions. It demonstrates how data-driven threshold selection can enhance the accuracy and stability of risk estimates.

Research Methodology

In this study, the market risk is assessed using the Extreme Value Theory (EVT) approach. EVT focuses on the behavior of extreme points in a distribution and is particularly useful for analyzing large losses occurring in the left tail of return distributions, especially during periods of market stress.

Generalized Pareto Distribution (GPD)

Within this framework, exceedances beyond a specified threshold are modeled using the Generalized Pareto Distribution (GPD). The cumulative distribution function (CDF) of the GPD is defined as (Gençay & Selçuk, 2004):

$$F(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, y > 0, \sigma > 0 \tag{1}$$

where:

- ξ: shape parameter, representing the tail thickness of the distribution,
- σ: scale parameter,
- $y = X - u$: the exceedance over the selected threshold u.

Accurate estimation of these parameters critically depends on the choice of the threshold that marks the start of the left tail, as an incorrect threshold may introduce systematic bias in risk estimation.

Shape Parameter Estimation

A crucial step in implementing the GPD framework is estimating the shape parameter (or tail index) ξ. In this study, rather than relying on the classical Hill estimator (Hill, 1975), we employ an improved, GUB-adjusted version of the Hill estimator. The detailed mathematical formulation and the implementation steps of this GUB-adjusted tail-index estimator are presented in the following subsection.

Let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$represent the order statistics of the financial return series sorted in descending order. If k denotes the number of observations exceeding the threshold u(such that $u = X_{(k+1)}$), the Hill estimator for the tail index is defined as:

$$\hat{\xi}_{\text{Hill}} = \frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{X_{(i)}}{X_{(k+1)}}\right)$$

The choice of k (and consequently the threshold u) is a critical preprocessing step, as it controls the bias-variance trade-off of the estimator. In this paper, the optimal threshold $u$and the corresponding number of tail events $k$are dynamically identified using the Good–Usual–Bad (GUB) algorithm. Once $\hat{\xi}_{\text{Hill}}$is is computed, it is directly integrated into the GPD cumulative distribution to estimate the tail probabilities and calculate the final Value-at-Risk (VaR) measures.

Good–Usual–Bad (GUB) Algorithm

To address the threshold selection problem, this study employs the Good–Usual–Bad (GUB) method, introduced by (González-Sánchez and Nave Pineda (2023). The GUB method is designed to determine the threshold for the onset of the distribution tail in a data-driven manner. In this approach, daily returns are classified into three main categories:

- Good: extreme positive returns,
- Usual: ordinary returns exhibiting stable statistical behavior,
- Bad: extreme negative returns or large losses.

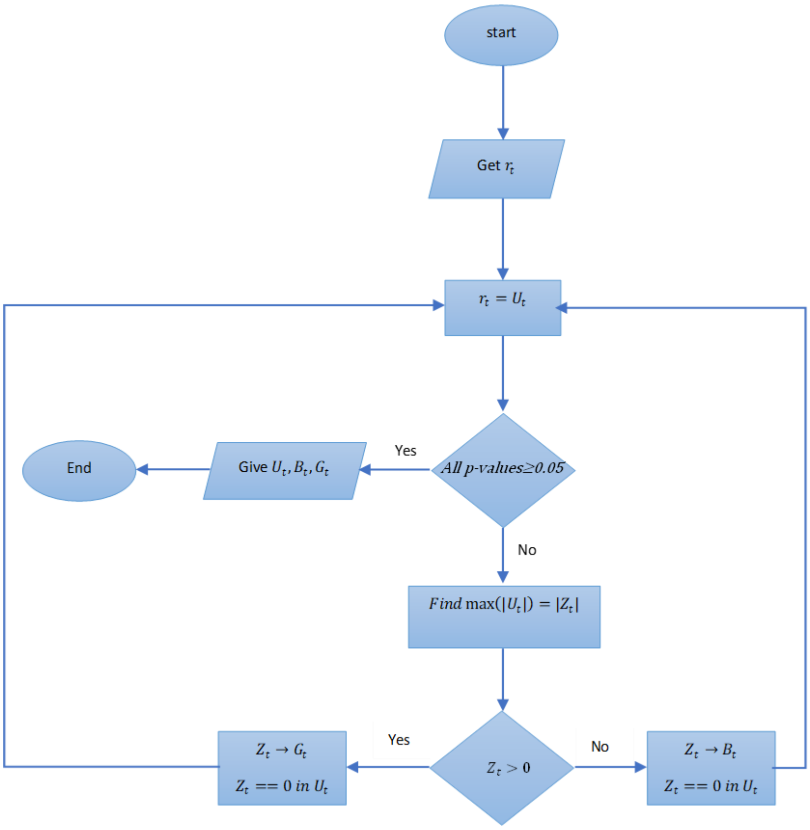

Figure 1. GUB Algorithm schematically

As shown in Figure 1, at the beginning of the algorithm, all returns are assigned to the Usual category. To assess the statistical stability of this set, three statistical tests are performed:

1. Jarque–Bera test for normality of returns, where the null hypothesis assumes a normal distribution,
2. LM–ARCH test for variance heterogeneity, ensuring the variance is stable over time,
3. Box–Pierce test for autocorrelation, where the null hypothesis assumes no serial correlation in returns.

If all three tests are satisfied at the chosen significance level, the returns in the Usual category are considered statistically stable, and the algorithm stops.

However, if any test fails, the algorithm iteratively identifies the return with the largest absolute value. If the sign of this return is positive, it is moved to the Good category; if negative, it is added to the Bad category. The three statistical tests are then repeated on the remaining Usual data. This process continues until the Usual set passes all statistical tests successfully.

At the end of this procedure, three non-overlapping time series are obtained:

- Good series: extreme positive returns,
- Usual series: representing normal and stable market behavior,
- Bad series: extreme negative returns.

The boundary between the Usual and Bad series is defined as the empirical threshold u (i.e., the largest value in the Bad series), marking the start of the left tail of the distribution. This threshold is subsequently used to fit the GPD within the EVT framework.

After determining the threshold, exceedances beyond u are used to estimate the GPD parameters. The shape parameter ξ is estimated using the Hill estimator, defined as:

$$H_{GUB} = \frac{1}{m}\sum_{i=1}^{m} ln\frac{B_i}{u} \quad , \ \alpha_{GUB} = H_{GUB}^{-1} \tag{2}$$

where:

- u: the threshold of the Bad series or the left-tail threshold,
- m: the number of exceedances in the tail region,
- B_i: members of the Bad series or returns in the Bad series.

The value of ξindicates the thickness of the negative tail; larger values imply a higher probability of extremely large losses.

The scale parameter $\sigma$is calculated following the relations proposed by Brooks et al. (2005):

$$\sigma = \left(\frac{1}{m}\sum_{i=1}^{m} B_i^{\frac{1}{\xi}}\right)^{\xi} \tag{3}$$

As shown in Equation (3), the scale parameter is calculated using the shape parameter. The definitions of the other variables are the same as in Equation (2).

Value-at-Risk (VaR) Calculation

After estimating the parameters ξand σ, the Value-at-Risk (VaR) at a given confidence level is obtained using the following Equation (Gençay & Selçuk, 2004):

$$VaR_{\gamma} = u + \frac{\sigma}{\xi}\cdot\left[\left(\frac{T}{m}.\gamma\right)^{-\xi} - 1\right] \tag{4}$$

where:

- T: total number of observations,
- m: number of exceedances beyond the threshold,
- γ: desired confidence level (in this study, 95% & 99%).

Method Comparison and Portfolio Risk Estimation

To evaluate the accuracy of the GUB method, its results are compared with those of three other conventional approaches. The first method is the parametric normal approach, which assumes that returns follow a normal distribution and calculates VaR parametrically. The second method is Historical Simulation, where VaR is defined as the empirical percentile of past returns at the desired confidence level. The third method is the t-student copula, used to model extreme dependencies among assets in a portfolio and capable of capturing joint heavy-tailed behavior (Stulajter, 2009).

Using these methods, the risk of each asset and the portfolio risk with equal weights are calculated and compared. This approach provides a comprehensive assessment of the GUB method's performance in market risk estimation, particularly under extreme market conditions and high volatility.

Given the GUB method's conservative nature at the portfolio level, adjustments are applied to compute portfolio risk. For example, in the GPD(Total (Bad) corr) approach, the tail threshold and the scale and shape parameters are estimated using the same GUB procedure, and the appropriate GPD is then fitted to the distribution's tail to calculate risk directly.

The portfolio risk under the GUB method is computed as follows (González-Sánchez & Nave Pineda, 2023), where the correlation matrix is calculated for the Bad series of the different assets, as the focus is on estimating tail risk and considering extreme observations:

$$VaR^{\gamma}_{P,GUB}(w) = \sum_{i=1}^{N} \frac{w_i}{w_T} x^{\gamma}_{i,GUB} \tag{5}$$

$$X^{\gamma}_{GUB} = \begin{pmatrix} A_B^{1,1} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_B^{1,i} & \cdots & A_B^{2,i} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_B^{1,N} & \cdots & A_B^{i,N} & \cdots & A_B^{N,N} \end{pmatrix} \cdot \begin{pmatrix} VaR^{\gamma}_{1,GUB} \\ \vdots \\ VaR^{\gamma}_{i,GUB} \\ \vdots \\ VaR^{\gamma}_{N,GUB} \end{pmatrix} \tag{6}$$

where:

- $\mathrm{VaR}^{\gamma}_{i,\,\mathrm{GUB}}$: univariate VaR of each asset at confidence level γ,
- w_T: total weight,
- w_i: weight allocated to asset i,
- Ω_B: correlation matrix among assets in the **Bad series**,

and via Cholesky decomposition: $\Omega_B = A_B A'_B$.

Data and Empirical Application

In this study, daily price data for four asset classes (18-carat gold, the USD/IRR exchange rate, Bitcoin (IRR), and the Tehran Stock Exchange Overall Index (TEDPIX)) were used. All data span the period from March 21, 2015, to March 19, 2025. Gold, USD/IRR exchange, and Bitcoin (IRR) prices in Iranian Rial were obtained from the reputable website Tgju.org, while the TEDPIX data were extracted from the FinPy library in Python.

To ensure precise comparability across markets, only days on which all markets were simultaneously active were retained. As a result, the number of observations for each asset amounts to 2,362 daily data points.

All prices are expressed in Iranian Rial, and daily returns were computed using the logarithmic relation:

$$r_t = \ln(P_t) - \ln(P_{t-1})$$

where P_t denotes the asset price on day t.

To obtain a more dynamic and realistic assessment of risk behavior, computations were performed using a five-year rolling window. At each step, a fixed five-year data window was considered and shifted forward by one day at a time. Accordingly, for each day in the sample period, market risk was estimated using the preceding five years of data, enabling dynamic measurement of Value-at-Risk over time. Finally, the average VaR across all rolling windows was reported as the final Value-at-Risk.

Descriptive Statistics of Returns

Table 1 presents the descriptive statistics for the daily logarithmic returns of gold, USD/IRR, TEDPIX, and Bitcoin/IRR. All four series demonstrate significant departures from normality, as confirmed by the Jarque-Bera, Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling tests. The p-values for all tests are effectively zero, providing empirical evidence to reject the Gaussian hypothesis for all assets under study.

The distributional properties further highlight the inadequacy of normal distribution models. Each asset exhibits substantial excess kurtosis—ranging from 2.36721in TEDPIX to a pronounced 15.85131in USD/IRR—which is indicative of heavy-tailed distributions and a high frequency of extreme price movements. Regarding asymmetry, TEDPIX displays mild positive skewness, whereas gold, USD/IRR, and Bitcoin/IRR exhibit negative skewness, suggesting a higher propensity for large downward price shocks. In terms of market risk, Bitcoin/IRR has the highest standard deviation (0.0211), making it the most volatile series, followed by gold, USD/IRR, and TEDPIX.

These stylized facts—specifically the heavy tails and the empirical rejection of normality—serve as the primary justification for employing Extreme Value Theory (EVT). Given that classical Value-at-Risk (VaR) models based on Gaussian assumptions are likely to underestimate tail risk in these markets, the subsequent analysis utilizes EVT to estimate VaR and capture the dynamics of extreme returns more accurately.

Table 1. Descriptive statistics of daily asset returns (2015–2025)

Asset	Observations	Mean	Std. Dev.	Kurtosis	Skewness	Jarque-Bera Stat	Jarque-Bera p-value	Shapiro-Wilk Stat	Shapiro-Wilk p-value	KS Stat	KS p-value	Anderson-Darling Stat
Gold	2,362	0.00081	0.00836	10.53543	–0.21468	10,890.5	0.0000	0.8572	0.0000	0.1187	0.0000	79.1
USD/IRR	2,362	0.00062	0.00792	15.85131	–0.56349	24,741.0	0.0000	0.7925	0.0000	0.1468	0.0000	124.7
Bitcoin (IRR)	2,362	0.00169	0.02115	7.20471	–0.55060	572.8	0.0000	0.9356	0.0000	0.1184	0.0000	61.4
TEDPIX	2,362	0.00069	0.00535	2.36721	0.25127	5,202.6	0.0000	0.9185	0.0000	0.0945	0.0000	42.4

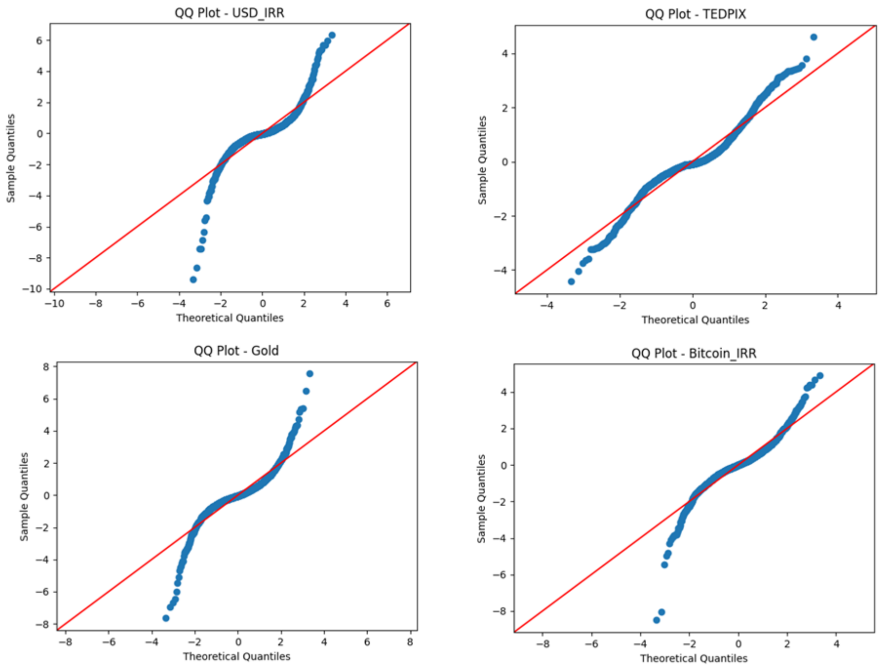

Figure 2. Q-Q.Plots for all four assets

The Q-Q plots for all four assets (Figure 2) confirm significant departures from normality. While a Gaussian distribution would align with the 45-degree reference line, the plots exhibit an S-shaped pattern, indicating heavy tails, with extreme returns occurring more frequently than predicted. Notably, TEDPIX shows the smallest deviation from the reference line, suggesting that its distribution is closer to normal than those of the other assets. However, all series—particularly USD/IRR and Bitcoin/IRR—display pronounced deviations in the tails. These findings indicate that standard Gaussian assumptions are inadequate for modeling the distribution of returns, thereby supporting the use of an EVT-based framework to more accurately capture tail risk.

Results and Discussion

Results of Series Segmentation Using the GUB Method

After applying the Good–Usual–Bad (GUB) procedure to the daily returns of the four assets, each return series was decomposed into three components: Good, Usual, and Bad. Using the observations in the Bad series, the parameters of the Generalized Pareto Distribution (GPD) including the scale parameter (σ), the shape parameter (ξ), the tail threshold (u), and the number of exceedances (m)—were estimated for each asset. The results are summarized in Table 2.

Table 2. Estimated GUB parameters for the assets

Asset	Scale (σ)	Shape (ξ)	Threshold (u)	Bad Series Count (m)
Gold	0.01838	0.17306	–0.02050	216
USD/IRR	0.01104	0.38309	–0.01096	282
Bitcoin (IRR)	0.03596	0.22491	–0.09371	72
TEDPIX	0.00893	0.06564	0	506

The results in Table 2 highlight substantial differences in the tail behavior of the assets. The shape parameter (ξ), which measures tail heaviness, varies significantly across markets. The USD/IRR exhibits the thickest tail ($\xi = 0.383$), indicating that extreme negative returns and large shocks are more likely in Iran's currency market than in other assets. Bitcoin (IRR) follows with a relatively high value ($\xi = 0.225$), consistent with the well-documented volatility of the cryptocurrency market. Gold also displays a moderately heavy tail ($\xi = 0.173$), suggesting that although extreme movements do occur, they are less frequent than in Bitcoin (IRR) or the USD/IRR.

In contrast, the Tehran Stock Exchange Index (TEDPIX) shows a very small shape parameter ($\xi = 0.0656$), indicating a thin-tailed distribution that is closer to normality. This result aligns with the market's daily price limits, which naturally constrain extreme fluctuations, and is consistent with the earlier descriptive analysis, where the kurtosis was below 3.

The scale parameter (σ), which reflects the dispersion of observations within the tail region, is highest for Bitcoin (IRR) (0.036) and lowest for TEDPIX (0.0089). This further confirms the substantially greater magnitude of extreme losses in the cryptocurrency market than in the Iranian stock market.

Additionally, the tail threshold (u) is notably more negative for Bitcoin (IRR) than for the other assets, indicating that extremely large negative returns occur more frequently in this market. In contrast, the TEDPIX threshold is zero, implying that extreme events are relatively rare due to structural trading constraints.

Estimated Value at Risk (VaR) and Comparison of Methods

To evaluate the performance of the GUB–EVT framework relative to established benchmarks, the Value at Risk (VaR) for each asset was estimated at the 95% and 99% confidence levels using several widely applied risk-measurement techniques. These approaches range from traditional models (such as the parametric VaR based on the assumption of normally distributed returns and the historical simulation method derived from empirical quantiles) to more advanced models that capture extreme losses and nonlinear dependence, including the EVT model with an optimally selected threshold determined by the GUB algorithm, as well as the t-Student copula, which accounts for tail dependence among assets during periods of market stress. This comparative analysis aims to identify the specific market conditions and asset classes in which the GUB-based approach offers a distinct advantage or demonstrates performance comparable to alternative methodologies.

Table 3. Estimated VaR values using alternative methods

Confidence level	Asset	Historical VaR	Parametric VaR	t-Copula VaR	EVT (GUB) VaR
99%	Gold	–0.0653	–0.0475	–0.069601	–0.0665
	USD/IRR	–0.0608	–0.0468	–0.068424	–0.0651
	Bitcoin (IRR)	–0.1514	–0.1193	–0.174269	–0.1481
	TEDPIX	–0.0357	–0.0306	–0.045319	–0.0366
95%	Gold	–0.0281	–0.0330	–0.040898	–0.0221
	USD/IRR	–0.0268	–0.0326	–0.040343	–0.0261
	Bitcoin (IRR)	–0.0747	–0.0832	–0.102791	–0.0622
	TEDPIX	–0.0221	–0.0200	–0.026237	–0.0196

The results, summarized in Table 3, reveal substantial differences in VaR estimates across these methods, particularly at the 99% confidence level. For assets such as Bitcoin (IRR) and the USD/IRR exchange rate, the tail-focused approaches generate VaR values of larger absolute magnitude compared with the parametric and historical models, highlighting their heightened sensitivity to extreme returns and severe market fluctuations; however, it also indicates a more conservative risk posture that may lead to higher capital requirements. In contrast, for assets such as gold and, notably, the Tehran Stock Exchange overall index, the estimates produced by the different methods are more closely aligned, suggesting a lower degree of tail risk in these assets, and indicating that for such portfolios, the choice of threshold selection method may have a secondary impact on the final risk estimate.

At the 95% confidence level, the dispersion among the methods becomes smaller, and the resulting VaR figures for most assets fall within a relatively narrow range. While this stability in ranking confirms the robustness of the general risk profiles, the empirical evidence shows that the GUB-based approach is not universally superior. In several instances—specifically at the 95% confidence level—alternative methods such as the t-Copula or Historical Simulation yield comparable or even lower VaR estimates. These findings suggest that the GUB method should be viewed as a specialized tool for robust threshold selection in high-volatility environments rather than a one-size-fits-all replacement.

The conservative nature of the GUB approach, as observed in the 99% VaR estimates, reflects its objective to minimize underestimation during extreme stress. However, as the results indicate, this can lead to an overestimation of risk in more stable periods, where simpler models provide sufficient coverage with less complexity. Therefore, GUB's superiority is context-dependent, excelling primarily at capturing high-magnitude tail events.

To provide a clearer understanding of how the estimated VaR values relate to the empirical distribution of returns, the density functions of asset returns were plotted, and the VaR estimates from each method were superimposed as vertical reference lines, as illustrated in Figure 3Figure 4.

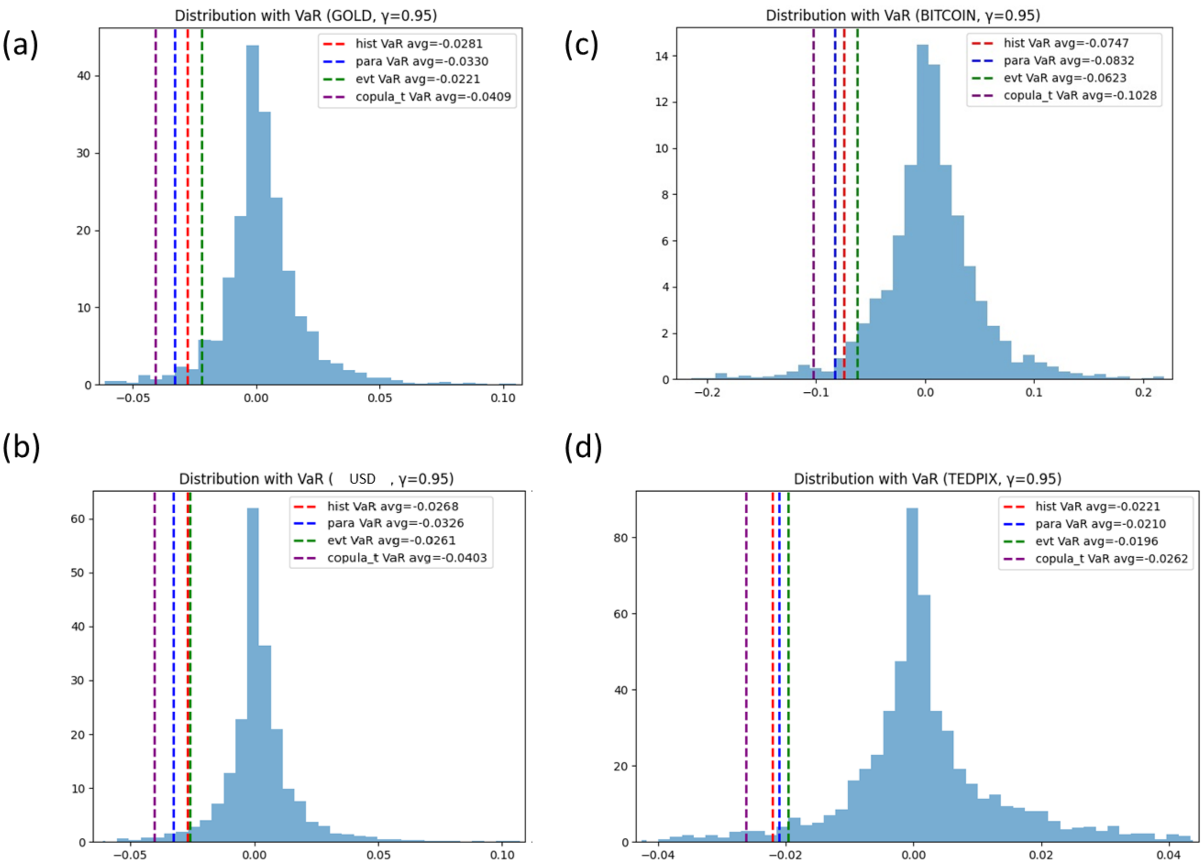

Figure 3. Empirical return distributions for the four asset classes ((a) gold, (b) USD/IRR exchange rate, (c) Bitcoin (IRR), and (d) the Tehran Stock Exchange (TSE) total index) together with the estimated Value at Risk (VaR) levels at the 95% confidence level. The vertical lines correspond to the VaR obtained from different methodologies: historical simulation (red), parametric normal VaR (blue), GUB–EVT (green), and the t-copula model (purple)

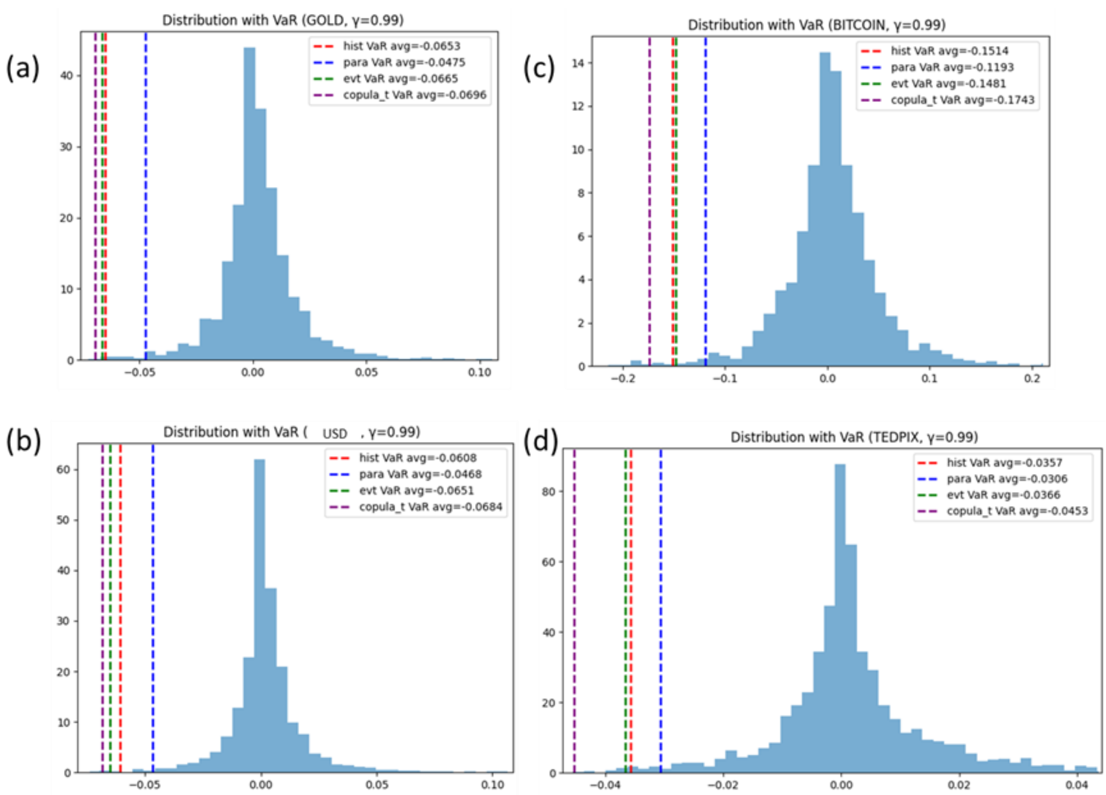

Figure 4. Empirical return distributions for the four asset classes ((a) gold, (b) USD/IRR exchange rate, (c) Bitcoin (IRR), and (d) the Tehran Stock Exchange (TSE) total index) together with the estimated Value at Risk (VaR) levels at the 99% confidence level. The vertical lines correspond to the VaR obtained from different methodologies: historical simulation (red), parametric normal VaR (blue), GUB–EVT (green), and the t-copula model (purple)

In these charts, the VaR lines obtained from the t-copula method appear further out in the left tail compared with the parametric estimates, indicating a better fit to heavy-tailed behavior. In contrast, for assets with relatively low kurtosis (such as gold and the stock market index), the GUB-based estimates closely overlap with those of the classical methods, suggesting that the choice of threshold-selection strategy has only a limited effect in these cases. However, for assets with pronounced tail heaviness, including Bitcoin (IRR) and the USD/IRR exchange rate, the GUB approach yields more conservative VaR levels, reflecting its tendency to place greater weight on extreme downside risk. This pattern suggests that the GUB model is particularly useful when the objective is to capture tail risk in highly volatile assets. However, its advantage is less pronounced for assets with milder distributional asymmetry or lower kurtosis.

Analysis of Backtesting Results and Model Performance Evaluation

To assess the accuracy of the different risk estimation approaches, a comprehensive set of standard backtesting metrics was employed. These include the total number of VaR violations, the hit ratio (the empirical failure rate), the average shortfall, the unconditional coverage test of Kupiec (Kupiec), the conditional coverage test of Christoffersen (1998), and the Lopez L1 loss function (Lopez, 1999). Each of these indicators captures a distinct aspect of model performance. The hit ratio evaluates how closely the observed failure frequency aligns with the theoretical confidence level α. Kupiec's test assesses whether the number of violations is statistically consistent with the target coverage. Christoffersen's test extends this by examining the independence of violations over time. Meanwhile, the Lopez L1 measure quantifies the magnitude of model misspecification relative to the benchmark α.

To facilitate a unified comparison across the different VaR models, a composite performance score was constructed by aggregating several key backtesting metrics into a single normalized index. The index integrates information from model accuracy, violation behavior, and loss-based performance. The construction procedure is detailed below to ensure transparency and full reproducibility.

A. Normalization Procedure

Because the backtesting metrics operate on different numerical scales (e.g., violation counts, likelihood-ratio test statistics, shortfall magnitudes), each metric was first transformed using a min–max normalization:

$$\widetilde{M}_{i,j} = \frac{M_{i,j} - \min(M_{\cdot,j})}{\max(M_{\cdot,j}) - \min(M_{\cdot,j})}$$

where

- $M_{i,j}$ is the value of metric j for model i,
- $\widetilde{M}_{i,j}$ is the normalized value in the interval $[0, 1]$,
- $\min(M_{\cdot,j})$ and $\max(M_{\cdot,j})$ are the minimum and maximum values of metric j across all competing models.

This transformation ensures comparability across metrics without affecting model rankings.

B. Weighting Scheme

Three components were included in the composite performance score, based on the metrics emphasized in the manuscript:

- **Violation-based performance** (e.g., violation count, hit ratio)
- **Statistical backtesting accuracy** (Kupiec UC and Christoffersen CC tests)
- **Loss-based performance** (Lopez L1 loss, average shortfall)

Given their equal importance for assessing VaR quality, each component was assigned an equal weight:

$$w_1 = w_2 = w_3 = \frac{1}{3}$$

Composite Performance Score

The final score for model i is computed as the weighted sum of the normalized components:

$$S_i = \sum_{j=1}^{3} w_j \ \tilde{C}_{i,j}$$

where

- $\tilde{C}_{i,j}$ is the normalized value of component j for model i,
- $S_i \in [0,1]$, with higher values indicating better overall VaR performance.

C. Interpretation

A higher composite score reflects a superior balance across:

- accurate violation frequency,
- passing statistical coverage tests,
- and minimizing tail losses when violations occur.

Thus, the score provides a concise and transparent ranking of the competing VaR models under a unified evaluation framework.

The results presented in Table 4 reveal that model performance varies significantly across confidence levels. At the 99% confidence level, the parametric model achieves the highest composite scores for gold, the USD/IRR exchange rate, and the TEDPIX index. This suggests that, for these assets, the return distributions in the extreme 1% tail may not deviate sufficiently from normality to justify the added complexity of EVT-based models, or that the GUB-selected threshold leads to a more conservative posture that penalizes the hit ratio at this level. In contrast, for Bitcoin (IRR)—whose returns exhibit very heavy-tailed behavior the EVT–GUB approach outperforms all competing models and achieves the highest score, demonstrating its superior ability to capture extreme downside risks in highly volatile markets.

At the 95% confidence level, a different pattern emerges. Here, the EVT–GUB method attains the highest composite score for all four assets. This indicates that when evaluating risk in a slightly broader tail area, the GUB algorithm strikes an optimal balance between coverage and precision, outperforming both overly simplistic parametric models and the t-copula.

Across most cases, the t-copula model records the lowest composite scores. While copula models are powerful tools for modelling tail dependence in multivariate contexts, the results indicate that the t-copula performs poorly for single-asset one-day VaR forecasting and fails to meet the statistical backtesting criteria. Conversely, the EVT–GUB model consistently delivers strong performance, particularly at the 95% level and in markets characterized by heavy-tailed return distributions, such as Bitcoin (IRR).

Overall, these findings underscore that there is no "universal winner." The parametric VaR remains efficient for moderately heavy-tailed assets at the 99% level. In contrast, the EVT–GUB model shows superior adaptability at the 95% level and for assets with pronounced tail risk (e.g., Bitcoin). This context-dependent performance justifies the use of GUB-EVT as a specialized tool for markets prone to extreme shocks.

Table 4. Results of Backtesting and Performance Evaluation at the 95% and 99% Confidence Levels

Asset	Confidence level	method	Breaches	Hit-ratio	Average shortfall	Kupic-POF LR(p.val)	Conditional Converge LR (p.val)	Lopez L1 mean	composite score
Gold	99%	historical	12	0.0051	0.028	7.045(0.008**)	10.913(0.004**)	0.0051	0.961
		parametric	19	0.0080	0.031	0.978(0.323)	7.944(0.019*)	0.008	1.196
		T-copula	9	0.0038	0.032	11.924(0.001**)	16.993(0.000**)	0.0038	0.941
		GUB	11	0.0047	0.030	8.496(0.004**)	12.71(0.002**)	0.0047	0.953
	95%	historical	7	0.0030	0.060	16.331(0.000**)	16.373(0.000**)	0.003	0.935
		parametric	12	0.0051	0.046	7.045(0.008**)	10.913(0.004**)	0.0051	0.961
		T-copula	7	0.0030	0.053	16.331(0.000**)	16.373(0.000**)	0.003	0.935
		GUB	7	0.0030	0.051	16.331(0.000**)	16.373(0.000**)	0.003	0.935
USD /IRR	99%	historical	4	0.0017	0.027	25.198(0.000**)	25.211(0.000**)	0.0017	0.954
		parametric	13	0.0055	0.026	5.762(0.016*)	9.316(0.009**)	0.0055	0.974
		T-copula	1	0.0004	0.046	39.134(0.000**)	39.135(0.000**)	0.0004	1.358
		GUB	5	0.0021	0.030	21.861(0.000**)	21.883(0.000**)	0.00021	1.581
	95%	historical	11	0.0047	0.006	8.496(0.000**)	19.996(0.000**)	0.0047	0.951
		parametric	29	0.0123	0.007	1.154(0.283)	55.981(0.000**)	0.0123	1.157
		T-copula	5	0.0021	0.004	21.861(0.000**)	29.4(0.000**)	0.0021	0.939
		GUB	14	0.0059	0.006	4.635(0.031*)	21.647(0.000**)	0.0059	0.982
Bitcoin	99%	historical	45	0.0191	0.027	61.718((0.000**)	80.636(0.000**)	0.0191	0.735
		parame	38	0.01	0.026	76.846(0.00	95.117(0.00	0.016	0.714

(IRR)		tric		61		0**)	0**)	1	
		T-copula	27	0.0114	0.027	106.162(0.000**)	114.987(0.000**)	0.0114	0.690
		GUB	81	0.0343	0.020	13.723(0.000**)	36.239(0.000**)	0.0343	0.855
	95%	historical	39	0.0165	0.027	74.534(0.000**)	87.352(0.000**)	0.0165	0.717
		parametric	31	0.0131	0.028	94.61(0.000**)	101.865(0.000**)	0.0131	0.697
		T-copula	22	0.0093	0.030	122.317(0.000**)	128.144(0.000**)	0.0093	0.690
		GUB	41	0.0174	0.027	70.067(0.000**)	81.937(0.000**)	0.0174	0.723
Stock index	99%	historical	36	0.0152	0.040	81.631(0.000**)	81.94(0.000**)	0.0152	0.708
		parametric	31	0.0131	0.037	94.61(0.000**)	95.245(0.000**)	0.0131	0.697
		T-copula	21	0.0089	0.029	125.808(0.000**)	127.602(0.000**)	0.0089	0.691
		GUB	52	0.0220	0.037	48.818(0.000**)	53.444(0.000**)	0.022	0.756
	95%	historical	64	0.0271	0.010	31.075(0.000**)	117.407(0.000**)	0.0271	0.795
		parametric	67	0.0284	0.009	27.399(0.000**)	115.529(0.000**)	0.0284	0.806
		T-copula	44	0.0186	0.008	63.734(0.000**)	133.201(0.000**)	0.0186	0.731
		GUB	85	0.0360	0.009	10776(0.001**)	126.894(0.000**)	0.036	0.870

Note: **: rejected at 1% level, *: rejected at 5% level.

To examine the dynamics of risk over time, the daily actual returns and the corresponding VaR estimates obtained from the four approaches—parametric, historical, t-Copula, and GUB–EVT—are plotted in Figure 5Figure 6. In these figures, the breaches, i.e., instances where the actual returns fell below the VaR estimates, are highlighted.

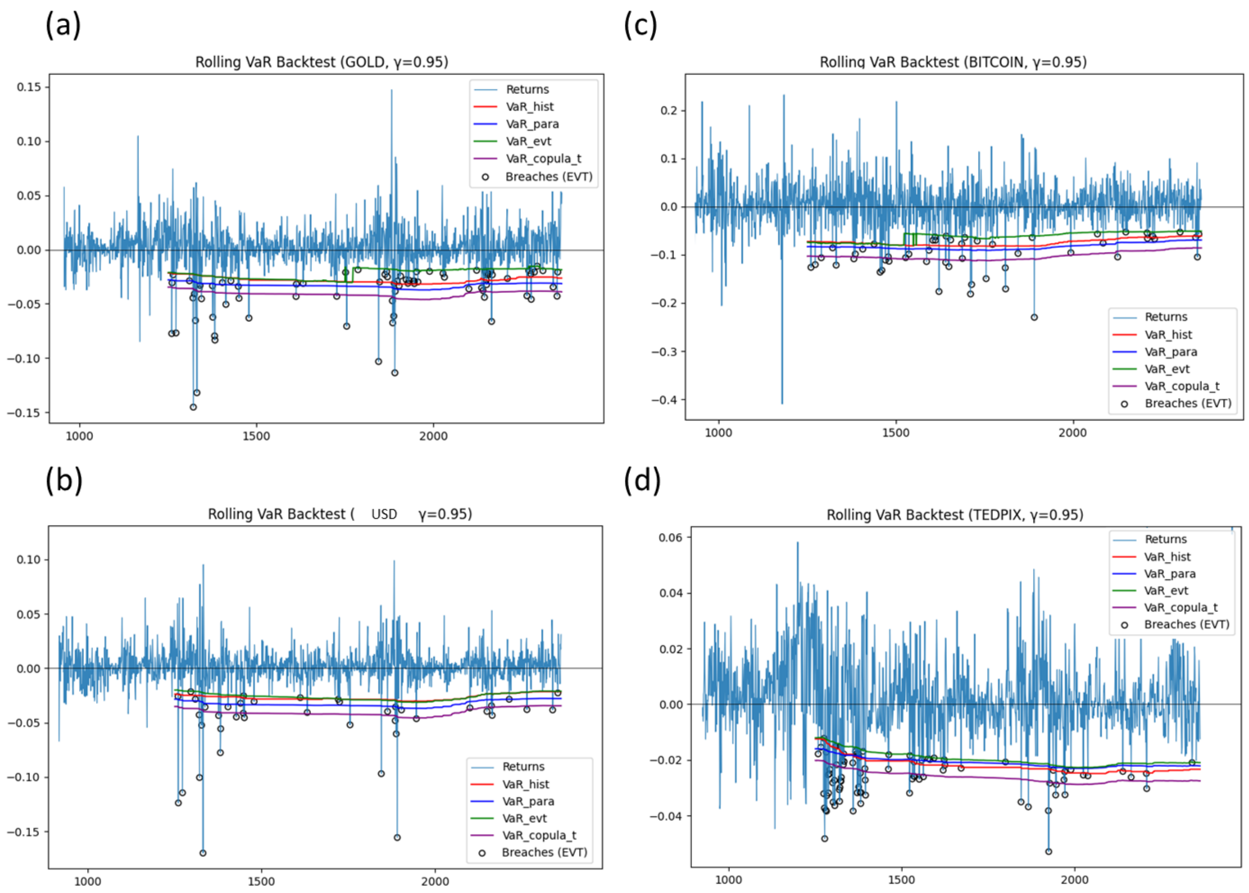

Figure 5. Performance evaluation plots for asset classes: (a) Gold, (b) USD/IRR Exchange Rate, (c) Bitcoin (IRR), (d) TEDPIX. The plots show VaR lines (red: Historical, blue: Parametric, green: EVT–GUB, purple: t-Copula) at the 95% confidence level, with the breaches clearly marked.

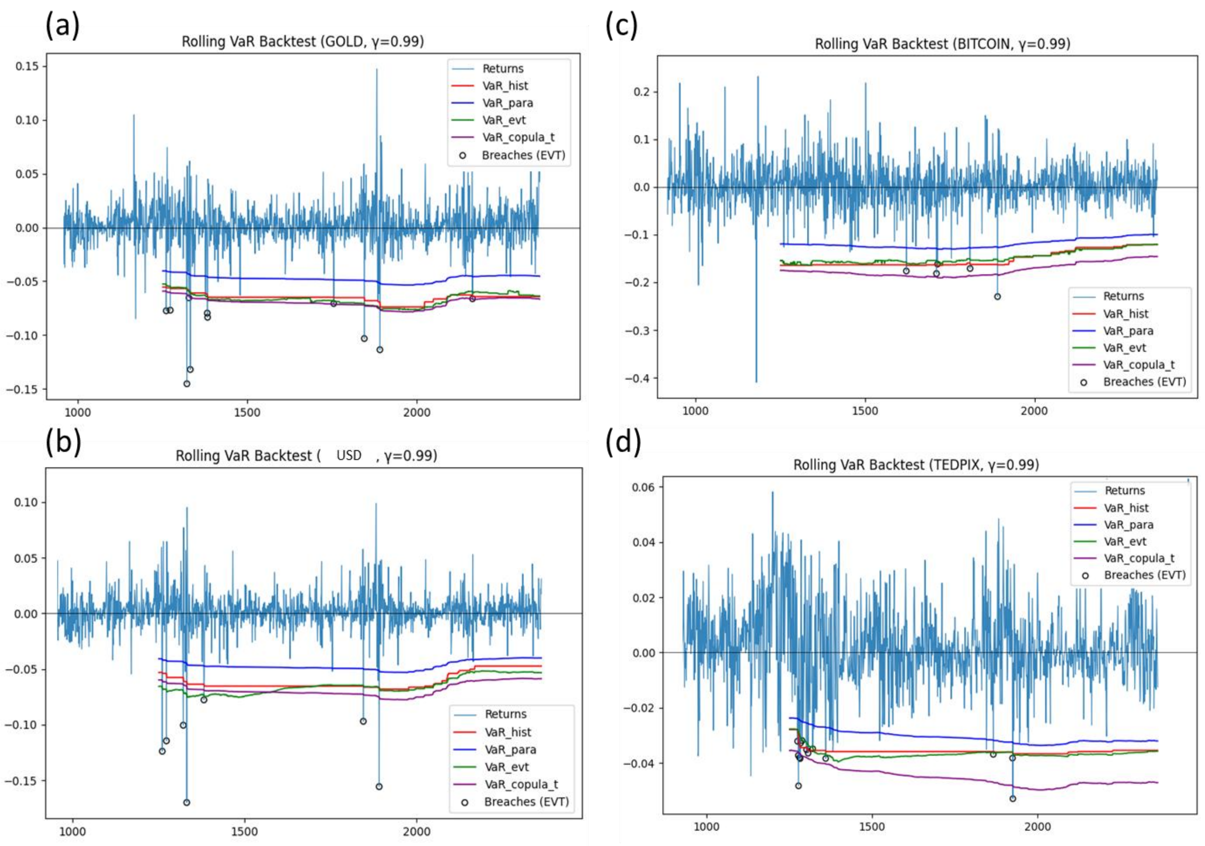

Figure 6. Performance evaluation plots for asset classes: (a) Gold, (b) USD/IRR Exchange Rate, (c) Bitcoin (IRR), (d) TEDPIX. The plots show VaR lines (red: Historical, blue: Parametric, green: EVT–GUB, purple: t-Copula) at the 99% confidence level, with the breaches clearly marked.

Portfolio VaR Results

Table 7 presents the results of portfolio risk estimation with equal weights across several methods, which are explained in detail below. First, to examine Value at Risk (VaR) estimation based on the GUB approach, the method proposed by (González-Sánchez and Nave Pineda (2023) was implemented. According to this method, the return series of each asset is divided into three categories: "Good," "Usual," and "Bad," and the tail threshold is automatically determined by separating the "Bad" segment. Then, for the returns in the "Bad" segment, the Generalized Pareto Distribution (GPD) is fitted, and the single-asset VaR is extracted.

To extend these single-asset estimates to the portfolio level, two correlation matrices were applied: the "Bad returns" correlation matrix (Table 5), reflecting dependence only among extreme losses, and the "Total returns" correlation matrix (Table 6), based on the full return series. These two versions were calculated as GUB (Bad Corr) and GUB (Total Corr), respectively. As shown in

Table 7, although the GUB algorithm accurately identifies the tail threshold, the resulting single-asset VaR from the GPD is highly conservative, and when aggregated at the portfolio level, it produces very large VaR values and almost zero breaches. This behavior leads to the rejection of conditional coverage tests. It indicates that directly transferring single-asset GUB VaR to the portfolio level, as with many EVT-based methods, can be overly cautious.

To overcome this limitation, two new methods were defined in this study. These approaches still use the GUB algorithm to determine thresholds, but calculate portfolio VaR directly via simulation using the GPD distributions. In these two methods, called GPD (Bad Corr) and GPD (Total Corr), the thresholds are first determined using GUB; the marginal distributions of each asset are modeled with the GPD; and the portfolio is then simulated using the respective correlation matrices. By replacing direct aggregation with simulation-based portfolio construction, these methods substantially reduce the excessive conservatism of the original GUB-based portfolio approach and generate more stable and realistic VaR estimates.

For a comprehensive comparison, classical methods, including parametric VaR and historical simulation, were also implemented as benchmarks. The parametric approach relies on the assumption of normally distributed returns and the total correlation matrix, while the historical method extracts empirical percentiles of losses without assuming any distribution. Additionally, to model non-linear dependencies and joint extreme behavior in the tail of the distribution, the t-Student copula was applied, which is particularly suitable under stressed market conditions.

Table 5. Correlation Matrix of Bad Return Series for Asset Classes

	Gold	USD/IRR	Bitcoin (IRR)	TEDPIX
Gold	1	0.5258	0.0710	0.0299
USD/IRR	0.5258	1	0.0825	0.0292
Bitcoin (IRR)	0.0710	0.0825	1	0.0347
TEDPIX	0.0299	0.0292	0.0347	1

Table 6. Correlation Matrix of Total Return Series for Asset Classes

	Gold	USD/IRR	Bitcoin (IRR)	TEDPIX
Gold	1	0.4956	0.1913	0.0387
USD/IRR	0.4956	1	0.2183	0.0136
Bitcoin (IRR)	0.1913	0.2183	1	0.0145
TEDPIX	0.0387	0.0136	0.0145	1

Table 7. Portfolio Value-at-Risk (VaR) Results Using the Mentioned Methods

Confidence level	method	VaR	Breaches	Hit-ratio	CC-LR (p.val)	Kupic-POF LR(p.val)	Lopez L1 mean	composite score
95%	GUB (Bad Corr)	0.0359	8	0.003	182.5 (0.000**)	182.443 (0.000**)	-0.04661	0.627
	GUB (Total Corr)	0.0362	7	0.003	188.095 (0.000**)	188.053 (0.000**)	-0.04704	0.625
	Hybrid GUB	0.0334	8	0.003	182.5 (0.000**)	182.443 (0.000**)	-0.04661	0.627
	GPD (Bad Corr)	0.0102	131	0.055	24.46 4.9 E-) (06**	1.435 (0.231)	0.00546	1.209
	GPD (Total Corr)	0.0107	119	0.050	13.27 0.00131**) (	1.029 (0.310)	0.00461	1.308
	Gaussian	0.0128	81	0.034	20.92 2.86E-) (05**	13.723 (0.0002**)	-0.01571	0.857
	Historical	0.0107	119	0.050	13.27 0.00131**) (	0.007 (0.932)	0.00038	2.331
	t-Copula	0.0103	131	0.055	24.46 4.9 E-) (06**	1.435 (0.231)	0.00546	1.209
99%	GUB (Bad Corr)	0.0892	0	0.000	47.48 4.9E-) (11**	47.478 (5.56E-12**)	-0.01000	0.905

	GUB (Total Corr)	0.090 4	0	0.00 0	47.48 4.9E-) (11**	47.478 (5.56E- 12**)	- 0.0100 0	0.905
	Hybrid GUB	0.083 3	0	0.00 0	47.48 4.9E-) (11**	47.478 (5.56E- 12**)	- 0.0100 0	0.905
	GPD (Bad Corr)	0.018 5	39	0.01 7	10.374 0.00559**) (	8.456 (0.004**)	0.0065 1	0.944
	GPD (Total Corr)	0.019 2	33	0.01 4	6.288 (0.0431*)	3.349 (0.067)	0.0039 7	1.039
	Gaussian	0.018 1	39	0.01 7	10.37 0.00559**) (	8.456 (0.004**)	0.0065 1	0.944
	Historica l	0.020 5	24	0.01 0	1.36 (0.5055)	0.006 (0.938)	0.0001 6	2.495
	t-Copula	0.021 0	23	0.01 0	1.51 (0.46998)	0.017 (0.898)	- 0.0002 6	2.329

Note: **: rejected at 1% level, *: rejected at 5% level.

For the evaluation of the accuracy of different portfolio risk estimation models, similar to single-asset risk, a set of standard performance metrics was calculated, including the number of breaches, hit ratio, average shortfall, Kupiec's unconditional and conditional coverage tests (Kupiec), (Christoffersen, 1998), and Lopez's L1 measure (Lopez, 1999). To facilitate comparison among methods, a composite indicator, with the same concept as for single-asset VaR, was also defined.

The empirical results indicate that at the 95% confidence level, the Historical Simulation and GPD (Total Corr) methods exhibited balanced performance and achieved the highest composite scores. At the 99% confidence level, the Historical Simulation, t-Copula, and GPD (Total Corr) methods provided the best results, respectively. In contrast, the original GUB–EVT versions appeared highly conservative at the portfolio level, with the number of breaches close to zero, which caused them to fail the conditional coverage tests despite performing well at the single-asset level. This finding suggests that although the GUB method is highly effective in identifying the tail threshold, it requires adjustments for portfolio-level application. However, the GPD-based versions combined with the GUB threshold can simultaneously ensure statistical accuracy and economic efficiency. Nevertheless, in extreme and high-risk conditions, applying the original GUB approach (González-

Sánchez & Nave Pineda, 2023) can still be justified.

Conclusion

In this study, by leveraging the data-driven GUB algorithm within the framework of Extreme Value Theory (EVT), an attempt was made to address the challenge of threshold selection in modeling heavy-tailed asset returns. The results demonstrated that the GUB method, through its automatic classification of return series into "Good," "Usual," and "Bad" segments, has a high capability in identifying the true tail structure and can provide more accurate estimates of GPD parameters and, consequently, Value-at-Risk (VaR) for assets exhibiting extreme behaviors, such as Bitcoin (IRR) and USD/IRR exchange rates.

The model comparison showed that at the 95% confidence level, the GUB–EVT approach, by providing balanced estimates consistent with coverage tests, outperformed the historical, parametric, and copula-based methods. At the 99% confidence level, while the parametric method performed better for low-volatility assets like the TEDPIX index, the GUB model outperformed others for highly volatile assets such as Bitcoin (IRR). At the portfolio level, however, we observed that the original GUB versions exhibited significant compounding conservatism, leading to excessive risk estimates and failure in conditional coverage tests. To address this, our study introduced a simulation-based GPD framework that uses GUB-determined thresholds. This combined approach successfully mitigated the observed bias, offering portfolio risk estimates that better reconcile statistical accuracy with economic efficiency and outperform direct aggregation methods.

Overall, the findings suggest that flexible, data-driven methods like GUB–EVT can significantly improve the accuracy of risk measurement in volatile and emerging markets. The study highlights that while the GUB method's threshold detection capability is a powerful tool for tail estimation, its application to multi-asset portfolios requires simulation-based adjustments to avoid excessive conservatism. It is further recommended that future research explore the application of these models in optimal asset allocation processes and compare their effectiveness with classical methods.

Future Research Directions and Limitations

A limitation of this study lies in the selection of the threshold parameter. While threshold selection is a critical pre-processing step in Peak-Over-Threshold (POT) modeling under Extreme Value Theory, the primary objective of this research is to evaluate the practical performance of the GUB-EVT integrated framework in estimating Value-at-Risk (VaR) within highly volatile markets, rather than benchmarking threshold selection algorithms. Consequently, this study adopts the GUB method as the baseline thresholding mechanism. Future research could systematically compare the sensitivity of EVT-based VaR estimates and portfolio risk metrics under alternative threshold selection procedures, such as the Mean Excess Plot, Hill Stability Plot, and Kolmogorov-Smirnov test, to assess their relative efficiency in emerging markets.

Although the GUB procedure provides a useful data-driven mechanism for identifying tail thresholds and performs well for individual assets with pronounced tail behavior, its direct application to multi-asset portfolios requires caution. Aggregating single-asset GUB–EVT VaR estimates can introduce compounded conservatism, as the marginal tail estimates are combined with dependence information without explicitly simulating the joint portfolio distribution. In our empirical application, the direct GUB specifications yielded excessively large portfolio VaR estimates and almost no VaR violations, resulting in failures in the conditional coverage tests despite satisfactory single-asset performance. This behavior may unnecessarily increase capital requirements and reduce the economic usefulness of the risk measure. Accordingly, GUB is better viewed as a threshold-identification tool than as a complete portfolio aggregation rule. The proposed GPD-based procedures address this limitation by retaining GUB-selected thresholds while simulating portfolio losses from the fitted marginal GPDs and the relevant correlation structure. These procedures, particularly GPD (Total Corr), produced more balanced portfolio VaR estimates and improved the trade-off between statistical accuracy and economic efficiency. Nevertheless, the results remain conditional on the selected thresholding mechanism, and future research should therefore examine the sensitivity of portfolio VaR to alternative threshold-selection procedures.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest concerning the research, authorship and, or publication of this article.

Funding

The authors received no financial support for the research, authorship and, or publication of this article.

References

Brooks, C., Clare, A. D., Dalle Molle, J. W., & Persand, G. (2005). A comparison of extreme value theory approaches for determining value at risk. *Journal of Empirical Finance,* 12(2), 339–352. https://doi.org/https://doi.org/10.1016/j.jempfin.2004.01.004

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review,* 39(4), 841–862. https://doi.org/10.2307/2527341

Drees, H., Janssen, A., Resnick, S. I., & Wang, T. (2018). On a Minimum Distance Procedure for Threshold Selection in Tail Analysis. *SIAM J. Math. Data Sci*., 2, 75–102.

Fedotenkov, I. (2020). A Review of More than One Hundred Pareto-Tail Index Estimators. *Statistica,* 80(3), 245–299. https://doi.org/10.6092/issn.1973-2201/9533

Gençay, R., & Selçuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting*, 20(2), 287–303. https://doi.org/https://doi.org/10.1016/j.ijforecast.2003.09.005

González-Sánchez, M. (2021). Is there a relationship between the time scaling property of asset returns and the outliers? Evidence from international financial markets. *Finance Research Letters,* 38, 101510. https://doi.org/https://doi.org/10.1016/j.frl.2020.101510

González-Sánchez, M., & Nave Pineda, J. M. (2023). Where is the distribution tail threshold? A tale on tail and copulas in financial risk measurement. *International Review of Financial Analysis,* 86, 102512. https://doi.org/https://doi.org/10.1016/j.irfa.2023.102512

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5), 1163–1174.

Jansen, D. W., Koedijk, K. G., & de Vries, C. G. (2000). Portfolio selection with limited downside risk. *Journal of Empirical Finance*, 7(3), 247–269. https://doi.org/https://doi.org/10.1016/S0927-5398(00)00016-5

Koedijk, K. G., Schafgans, M. M. A., & de Vries, C. G. (1990). The tail index of exchange rate returns. *Journal of International Economics*, 29(1), 93–108. https://doi.org/https://doi.org/10.1016/0022-1996(90)90065-T

Kupiec, P. H. - Techniques for Verifying the Accuracy of Risk Measurement Models. https://doi.org/- 10.3905/jod.1995.407942

Lopez, J. A. (1999). Methods for evaluating Value-at-Risk estimates. *Federal Reserve Bank of New York Economic Policy Review,* 4(3), 119–124.

Németh, L., & Zempléni, A. (2017). Regression Estimator for the Tail Index. *Journal of Statistical Theory and Practice,* 14. https://doi.org/10.1007/s42519-020-00114-7

Rhee, S. G., & Wu, F. (2020). Conditional extreme risk, black swan hedging, and asset prices. *Journal of Empirical Finance*, 58, 412–435. https://doi.org/https://doi.org/10.1016/j.jempfin.2020.07.002

Stulajter, F. (2009). Comparison of Different Copula Assumptions and Their Application in Portfolio Construction. Capital Markets: Asset Pricing & Valuation.

Bibliographic information of this paper for citing:

Ebadinezhad, Narges & Dehghani Ahmadabad, Mohammad Reza (2026). EVT-Based Value-at-Risk Estimation in Iranian financial markets Using the GUB Threshold Identification Method. *Iranian Journal of Finance,* 10(2), 1-31.

Copyright © 2026, Narges Ebadinezhad and Mohammad Reza Dehghani Ahmadabad