commodities

MDPI

Extreme Value Theory and Gold Price Extremes, 1975–2025: Long-Term Evidence on Value-at-Risk and Expected Shortfall

1. Abstract

We analyze extreme gold price movements between 1975 and 2025 using Extreme Value Theory (EVT). Using both the Block-Maxima and Peaks-over-Threshold approaches on a daily return basis, we estimate Value-at-Risk (VaR) and Expected Shortfall (ES) for the entire distribution focusing on a long-term view. Our results demonstrate that models based on the standard normal distribution systematically underestimate extreme risks, whereas EVT provides more reliable measures. In particular, EVT captures not only rare losses, but also sudden positive rallies, highlighting gold's dual function as a risk and opportunity asset. Asymmetries emerge in the analysis: at the 0.99 quantile, losses appear larger in absolute value than gains. At the 0.995 quantile, in some episodes, upside extremes dominate. Furthermore, we find that geopolitical and economic shocks, including the oil crises, the 2008 financial crisis and the COVID-19 pandemic, leave distinct signatures in the extremes. By covering five decades, our study provides the most extensive EVT-based assessment of gold risks to date. Our findings contribute to debates on financial stability and provide practical guidance for investors seeking to manage tail risks while recognizing gold's potential as both a safe haven and a speculative asset.

Keywords: Gold, Extreme Value Theory (EVT), Value-at-Risk (VaR), Expected Shortfall (ES)

1. Introduction

Throughout history, gold has served as both a store of value and a refuge asset during episodes of monetary, political, and economic turbulence. Unlike fiat currencies or corporate assets, gold is not tied directly to the performance of any single economy. This makes it particularly appealing during periods of crisis, including armed conflicts, inflation, financial crises, and pandemics. (Gold's perceived stability and universal acceptability have established it as a hedge against systemic risk and currency devaluation (Capie et al., 2005; Beckmann et al., 2015).

Over the past five decades, the global economy has experienced multiple severe disruptions, including the collapse of the Bretton Woods system, the 1970s oil shocks, the dot-com bubble, the 2008 global financial crisis, the 2020s pandemic, and, more recently, geopolitical tensions resulting from the war in Ukraine and broader shifts in the global economic order (Eichengreen, 2008; Cheema et al., 2020; Smales, 2021). In each of these periods, gold has experienced significant price movements, driven in part by heightened investor demand for safe assets (Batten et al. 2010; Gkillas and Longin, 2019).

In addition to private and institutional investors, central banks have emerged as key actors in the global gold market, significantly influencing price dynamics. Particularly in recent years, large-scale gold purchases by central banks—especially in emerging economies seeking to diversify away from U.S. dollar holdings—have contributed to both the upward trajectory and the volatility of gold prices (Aizenman and Inoue, 2013; Baur and McDermott, 2016; Ilesanmi and Tewari, 2020). Strategic factors—such as concerns over monetary sovereignty, inflation hedging, and geopolitical risk exposure—appear to increasingly drive central bank demand for gold.

Despite its centrality in financial markets, the behavior of gold prices under extreme market conditions remains insufficiently explored, especially over long historical timeframes. While various studies have examined short- to medium-term gold price volatility, few have applied Extreme Value Theory (EVT) to comprehensively analyze both the downside risks and the upside potential (e.g., explosive rallies) associated with gold investments (Chinhamu et al., 2014; Khan et al., 2021; Gkillas and Longin, 2019). Furthermore, existing EVT-based analyses tend to focus on short time periods or specific geographic markets. This overlooks broader structural shifts in the global financial system.

Given the dual function of gold—as both a source of potential loss during deflationary downturns and a vehicle for capital preservation or appreciation during crises—there is a compelling need to adopt a tail-focused statistical approach that concentrates on the most extreme price changes—those that occur only under extraordinary market stress. EVT shifts the analytical focus away from the average case to the most extreme outcomes, offering a well-suited statistical tool for evaluating rare but consequential events in financial markets (McNeil et al., 2015; Gilli and Këllezi, 2006). When applied across a comprehensive time horizon, EVT can offer unique insights into the evolving risk-return profile of gold under varying macroeconomic and geopolitical regimes.

In response to the limitations of the existing literature, we propose an expanded empirical analysis that takes a long-term view of statistical analysis. We will apply EVT to gold price data over an extended period from 1975 to 2024. This approach will account for multiple economic cycles, crises, and structural transformations. In doing so, it aims not only to assess the extreme risks associated with gold investments but also to evaluate the opportunities, such as gold's performance as a hedge or safe haven in times of systemic stress. Additionally, this analysis considers the influence of central bank demand and macro-financial dynamics, offering a nuanced understanding of gold's behavior under financial stress and extreme tail events.

2. Literature Review

Extreme Valute Theory is a statistical framework developed for modeling rare events, especially extreme price fluctuations in financial markets. Unlike traditional models, EVT focuses on the tails of distributions, which are critical for capturing the full extent of risk. EVT is useful for estimating risk measures such as Vlaue at Risk (VaR) and Expected Shortfall (ES). These measures are critical for portfolio management and risk assessment (Khan et al., 2021; Chinhamu et al., 2014; Chaithep et al., 2012). EVT comprises, in particular, the following models:

1. Block Maxima (BM) Method: This approach involves dividing the data into blocks (e.g., annual blocks) and analyzing the maximum or minimum values within each block. The generalized extreme value (GEV) distribution is often used to model these block maxima (Khan et al., 2021; Chaithep et al., 2012).
2. Peak Over Threshold (POT) Method: This method focuses on extreme values above (or below) a certain threshold. The generalized Pareto distribution (GPD) is commonly used to model these exceedances (Chinhamu et al., 2014; Giles & Chen, 2014).

The BM approach is widely used to model extreme gold price movements. For example, one study analyzed daily gold prices in the Pakistan Bullion Market from 2011 to 2021 using this approach. The results showed that the GEV distribution effectively captured extreme price movements, enabling accurate VaR and ES estimates (Khan et al., 2021).

Similarly, another study used the BM method to forecast extreme price levels over the next 20 years based on historical gold prices (Chaithep et al., 2012).

The POT approach has been used to model extreme tail behavior in gold prices. For instance, a study of gold, silver, and platinum prices employed the POT method to estimate VaR and ES for extreme daily price fluctuations. The results showed that silver had the highest risk of the three metals (Giles & Chen, 2014).

Some studies have combined EVT with other models to improve the accuracy of gold price analysis. For example, one study integrated EVT with GARCH models to account for volatility persistence in gold prices. The results revealed that the hybrid model yielded more precise VaR and ES estimates than EVT models operating independently (Khan et al., 2023; Khemawanit & Tansuchat, 2016). Another study combined POT with GARCH models to estimate VaR and ES for gold prices, demonstrating EVT's superiority over traditional models (Chinhamu et al., 2014).

Multivariate extreme value theory (MVEVT) has been used to analyze the joint behavior of gold prices with other assets or indices. For instance, one study employed extreme value copulas to investigate the dependence structure between gold prices and the U.S. dollar index. The results showed that gold and the U.S. dollar index were independent of each other in extreme market conditions (Kaewkheaw et al., 2014).

EVT is primarily used to estimate two key risk measures in gold price analysis.

1. Value-at-Risk
2. Expected Shortfall

VaR is a measure of the maximum potential loss that a portfolio could incur within a given time frame and confidence level. EVT has been widely used to estimate VaR for gold prices. For instance, the study of Khan et. al. of gold prices in the Pakistan Bullion Market employed the GEV distribution to estimate VaR for extreme daily losses and gains (Khan et al., 2021). Another study used the generalized Pareto distribution (GPD) to model extreme gold price returns, demonstrating EVT's effectiveness in VaR estimation (Chinhamu et al., 2014).

ES is a measure of the average potential loss in the worst (1 - confidence level) percent of cases. EVT has been used to estimate ES for gold prices, often in combination with VaR. For example, one study examined gold, silver, and platinum prices and used the GPD to estimate the expected shortfall for extreme daily price changes. The results showed platinum to be riskier than gold for negative returns and gold to be riskier than platinum for positive returns (Giles and Chen, 2014).

Several studies have used EVT to provide valuable insights into the behavior of gold prices.

1. Volatility and Tail Behavior: Gold prices exhibit significant volatility with fat-tailed distributions that traditional models cannot capture. EVT has been used to effectively model these fat tails. For instance, one study of gold price returns showed that EVT models outperformed Gaussian and Student's t models in capturing tail behavior (Chinhamu et al., 2014).

2. Extreme Risk in Gold Markets: EVT has been used to analyze extreme risk in gold markets, particularly during periods of high volatility. For instance, a study of gold futures markets found that incorporating tail risk indicators improved the predictive accuracy of models of gold price volatility (Tang and Zhong, 2023).

3. Comparison with Other Assets: EVT has also been used to compare the extreme risk of gold with that of other assets. For example, a study of bitcoin and gold prices found that both assets exhibited low extreme correlation, indicating their potential as diversification tools during market turbulence (Gkillas and Longin, 2019).

4. Safe Haven Properties Gold is often considered a safe haven during market crises. EVT has been used to analyze gold’s safe haven properties during extreme market

conditions. For instance, one study examined the relationship between gold and the U.S. dollar index and found that gold effectively acted as a safe haven against extreme U.S. dollar rate fluctuations (Reboredo, 2013). Other studies have come to different conclusions. For example, in a quantile-specific study using extreme quantile regression, Liu (2013) showed that gold acts as a safe haven only in the U.S. market at moderate extreme values (e.g., the 5% quantile). However, this protective function does not exist in other markets, such as those in France, Australia, and South Korea. In a later study, Liu (2020) found that, during extreme market events (e.g., the 0.1% quantile), neither gold nor government bonds act as reliable hedges; both asset classes fail as "ultimate" collateral.

The applications of EVT in gold price analysis are diverse and include the following:

1. Portfolio Risk Management: EVT can be used to estimate VaR and ES for gold portfolios, enabling investors to effectively manage tail risks. For instance, one study examined precious metal portfolios and used EVT to calculate the VaR and ES of an equally weighted portfolio of gold, silver, palladium, and platinum (Khemawanit and Tansuchat, 2016).

2. Return-Level Forecasting: EVT has been used to forecast extreme gold price levels over specific time horizons. For instance, a study of gold prices in the Pakistan Bullion Market forecasted extreme price levels for the next five and ten years using the GEV distribution (Khan et al., 2021).

3. Hedging and Safe Haven Strategies: EVT has also been used to analyze gold's role as a hedge or safe haven against other assets. For example, one study examined the relationship between gold and the U.S. dollar index and discovered that gold effectively hedges against fluctuations in the U.S. dollar exchange rate (Reboredo, 2013).

4. Volatility Forecasting: EVT has been used to improve the accuracy of gold price volatility forecasts. For instance, a study of gold futures markets revealed that incorporating tail risk indicators into volatility models significantly enhanced forecasting accuracy (Tang and Zhong, 2023).

3. Research Gap

Despite EVT being widely used to analyze extreme risks in the gold market, existing studies tend to focus on specific time periods, regions or risk perspectives (e.g. loss analysis without opportunity analysis).

A major limitation of existing research is the lack of a comprehensive EVT-based analysis of gold price extremes spanning the period from 1975 to 2024. This period encompasses major global economic and geopolitical events, including oil crises, the end of the gold standard, the 2008 financial crisis, the pandemic, the war in Ukraine, and recent economic changes driven by digitalization, energy and climate policy.

Moreover, existing studies tend to adopt a unidimensional focus on downside risks, with limited attention to potential gains, upside opportunities, and in particular sudden rallies.

Consequently, the research fails to consider either the dual perspective of risk and opportunity, or the long-term view and broad historical context that would be crucial for a sound investment assessment of gold.

Based on this research gap, the following key research questions can be derived:

1. How have extreme movements in gold prices evolved between 1975 and 2024, and to what extent can these be associated with major geopolitical and macroeconomic events?

Objective: The historically sound application of EVT to an exceptionally long period.

2. To what extent can EVT simultaneously capture downside risks and upside opportunities in gold investments, including safe haven behavior and

protective characteristics? What symmetries and asymmetries are present in long-term VaR and ES estimates based on EVT?

Objective: Extending the use of EVT to positive extreme values, not just extreme losses.

3. What methodological differences arise when analyzing long-term gold price extremes using the Block-Maxima method and the Peaks-over-Threshold method?

Objective: Comparing EVT methods in long-term use.

4. How do VaR and ES estimates based on EVT compare with a purely statistical approach in a long-term view?

Objective: Providing an in-sample test and validating EVT methods in long-term use.

5. What long-term time-dependence characteristics and statistical dependencies influencing the long-term dependence of extreme events can be deduced from EVT estimates?

Objective: Deriving long-term time dependencies of EVT methods.

6. How can EVT be used to derive long-term predications for VaR and ES measures?

Objective: Testing the hypothesis that EVT methods based on one decade can be employed to predict VaR and ES for the following decade.

7. How do the risk and opportunity assessments of gold differ across different geopolitical and economic phases (e.g. oil crises, financial crises, pandemics and wars)?

Objective: Contextualizing extreme value analyses across multiple economic eras.

Our paper is structured as follows: Building on the formulation of the quantiative research questions 2 to 6 , Chapter 4 outlines the methodological framework. Chapter 5 then introduces the underlying dataset and presents the empirical analysis, in which the main results are reported, systematically interpreted, and discussed in light of the research questions. This part constitutes the core of the paper, as it connects the theoretical considerations with the empirical evidence and tests the validity of the applied methods. Finally, Chapter 6 summarizes the central findings, highlights the implications for investors and regulators, and identifies potential avenues for future research.

4. Theoretical background: return statistics and Extreme Value Theory

In this section we outline the theoretical background and the fundamental results relevant for our empricial analysis in Chapter 5, in particular from EVT. For a more detailed and profound overview we refer to (Coles, 2001; Embrechts et al., 1997; McNeil et al., 2015).

4.1. Logarithmic returns

Our basic data is a time series of prices from initial time 0 to final time T

$$P_0, P_1, P_2, \dots, P_T$$

From this time series we derive the time series of returns

$$X_1, X_2, \dots X_T$$

using logarithmic returns, thus we set

$$X_t = \ln P_t - \ln P_t$$

We use logarithmic returns due to the fact that they are time-additive, see (Ruppert and Matteson, 2015), and more importantly due to their symmetry. One of our main questions concerns symmetric properties of EVT estimates. Logarithmic returns are symmetric in the following sense: A price change by a factor $a > 1$, i.e. $P_1 = a \cdot P_0$, leads to a return

$$X_1 = \ln a$$

While a price change by a factor $1/a$, i.e. $P_1 = \frac{1}{a} \cdot P_0$, leads to a return

$$X_1 = -\ln a$$

For example, if the price doubles the return is $\ln 2 = 0.693$ while if the prices halves the return is $-\ln 2 = -0.693$.

For any subperiod $i = t_1, \dots, t_n$ the mean return is given by the average

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_{t_i}$$

and the volatility is given by the standard deviation

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(X_{t_i} - \overline{X}\right)^2}$$

which is to square root of the average quadratic deviation from the mean.

4.2. Formal definitions of Value-at-Risk and Expected Shortfall

Formally, the risk measure VaR is defined to be a quantile of the return distribution. Given a distribution function F_X for the distribution of returns X and a level $0 < \alpha < 1$, then the Value at Risk at level α (VaR_α) is defined to be

$$VaR_{\alpha} = \inf\{x \in \mathbb{R} \mid F_X(x) \geq 1 - \alpha\}$$

That means the probability that the return is less than VaR_α is less or equal than $1 - \alpha$. In other words a loss that is in absolute value greater than VaR_α only happens with a probability $1 - \alpha$.

The risk measure Expected Shortfall is closely related to the Value at Risk. It is defined to be the expected value of the return given that the return is less or equal than the VaR. Using conditional expected value, the formal definition given a level $0 < \alpha < 1$ is

$$ES_\alpha = \mathrm{E}[X \mid X \leq VaR_\alpha]$$

The ES can be calculated analytically from the VaR

$$ES_\alpha = \frac{1}{1-\alpha} \cdot \int_0^{1-\alpha} VaR_z \, dz$$

Thus, the expected shortfall is the expected loss in case the VaR is breached.

In our analysis we are interested in symmetric properties of VaR and ES estimates. Typically, the definitions above are applied for levels α greater than 0.9. Then the definitions above measure the left tail of the return distribution and thus concern losses.

Formally, the definition of the VaR can also be applied for small levels α. If α is taken to be less than 0.1, then the VaR measures the right tail and thus concerns gains. Thus, if we are interested in the positive tail for a given level α, in order to analyze symmetric properties of the VaR, we apply the same definition but switch from α to $1 - \alpha$

$$VaR_{\alpha} = \inf\{x \in \mathbb{R} \mid F_X(x) \geq \alpha\}$$

In order to define a measure analogous to ES for the right tail, i.e. for gains we put

$$ES_\alpha = \mathrm{E}[X \mid X \geq VaR_\alpha]$$

and in this case, for gains, we have

$$ES_\alpha = \frac{1}{1-\alpha} \cdot \int_\alpha^1 VaR_z \, dz$$

Thus, for gains, we let ES be the expected gain if the return is larger than the VaR. Note that the definitions above are symmetric. If applied to the return series directly, the defined measures give extreme values for the gains. However, if we multiply the return series with -1 in order to change the sign, then these measures are equivalent to the originally defined ones and measure extreme values for the losses.

4.3. Historical Value at Risk and Expected Shortfall

Given a return series

$$X_1, X_2, \dots, X_T$$

the VaR and ES can be estimated directly from the data. This is called the historical (or sample) VaR and ES and is based only on quantiles. The historical VaR is defined to be the $(1-\alpha)$ sample quantile of the return series: Let

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(T)}$$

be the order statistics of the return series. Then for a given level α

$$VaR_\alpha = X_{(\lceil T \cdot (1-\alpha) \rceil)}$$

and

$$ES_\alpha = \frac{1}{\lceil T \cdot (1-\alpha) \rceil} \sum_{i=1}^{\lceil T \cdot (1-\alpha) \rceil} X_{(i)}$$

The historical VaR and ES estimates are non-parametric and use only statistical methods.

4.4. The normal model

The standard way of estimating the VaR and ES is based on the assumption that the return distribution is close to a normal distribution. This normal model is a simple parametric method using only the parameters

$$\mu = \overline{X}$$

the mean of the returns and

$$\sigma = \sqrt{\overline{(x-\mu)^2}}$$

the volatility of returns. Then for a given level α

$$VaR_\alpha = \mu + \sigma \cdot \Phi^{-1}(1-\alpha)$$

and

$$ES_\alpha = \mu + \frac{\sigma}{\alpha} \cdot \phi(\Phi^{-1}(1-\alpha))$$

where Φ is the distribution function and ϕ is the density of the normal distribution.

4.5. The Block-Maxima method

The Block-Maxima method is based on the idea to partition the return series in blocks of a certain length take the maximum for each block and model the distribution of these. The VaR and ES is then calculated from this extreme value distribution. Here we follow (Embrechts et al., 1997) to summarize the theoretical background.

In this section we analyze a general time series

$$X_1, X_2, \dots X_T$$

This can be either a return series or a time series of losses where each return is multiplied by -1 to change the sign.

4.5.1. The iid approach

Let $X_1, X_2, \dots X_T$ be the time series and assume $T = m \cdot n$, where $m \in \mathbb{N}$ is the number of blocks and $n \in \mathbb{N}$ is the length of each block. Note that we shorten the time series from the beginning, if necessary, in order to have only blocks of length n. Then for $i = 1, \dots, m$ and $j = 1, \dots, n$, let $X_j^{(i)} = X_{n(i-1)+j}$ and define

$$M_n^{(i)} = \max\left\{X_1^{(i)}, \dots, X_n^{(i)}\right\}$$

to be the maximum of each block.

For the remainder of this section, let us assume that the time series $X_1, X_2, \dots X_T$ consist of independent, identically random variables with distribution function F. Then the distribution function of $M_n^{(i)}$ is given by

$$\Pr\left(M_n^{(i)} \leq z\right) = \Pr\left(X_1^{(i)} \leq z, \dots, X_n^{(i)} \leq z\right) = F(z)^n$$

for all $z \in \mathbb{R}$ and all $i = 1, \dots, m$. Under the assumption of independent and identically distributed random variables the distribution F can be estimated asymptotically in the limit $n \to \infty$. This result is known as the Fisher-Tippet-Gnedenko Theorem (Embrechts et al., 1997, p. 122). If there are $a_n > 0$, $b_n \in \mathbb{R}$, and a non-degenerate distribution G, then for all $z \in \mathbb{R}$ and all $i = 1, \dots, m$

$$\lim_{n\to\infty} \Pr\left(\frac{M_n^{(i)} - b_n}{a_n} < z\right) = G(z)$$

with

$$G(z) = \exp\left(-\left(1 + \frac{\xi(z-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right)$$

and

$$1 + \frac{\xi(z-\mu)}{\sigma} > 0$$

where $\mu \in \mathbb{R}$, $\xi \in \mathbb{R}$, and $\sigma > 0$. The distribution G is called generalized extreme value distribution (GED), ξ is called shape parameter, μ location parameter, and σ scale parameter.

These parameters can be estimated, e.g. applying the maximum likelihood method. This parameter fitting is well-suited for $\xi > 0$ (Longin, 1996; Smith, 1985).

4.5.2 The extreme index

To apply the Block-Maxima method to return series $X_1, X_2, \dots X_T$ that are stationary but not necessarily independent, identically distributed, a so-called extreme index is used to account for clustering of the extreme values (McNeil et al., 2015).

Let F be the marginal distribution function of X_t, $t = 1, \dots, T$, and $Y_1, Y_2, \dots, Y_T$ be a series of independent, identically distributed random variables with distribution function F. For $i = 1, \dots, m$ define

$$\widetilde{M_n^{(i)}} = \max\left(Y_1^{(i)}, \dots, Y_T^{(i)}\right)$$

to be the block maxima of $Y_1, Y_2, \dots, Y_T$. Then for each $0 < \theta < 1$^and each $\tau > 0$ there is a series $(u_n)_{n=1}^{\infty} \in \mathbb{R}^{\infty}$ such that

$$\lim_{n\to\infty} n\big(1 - F(u_n)\big) = \tau\,, \lim_{n\to\infty} \Pr(\widetilde{M_n^{(i)}} \le u_n) = \exp(-\tau)\,, \lim_{n\to\infty} \Pr(M_n^{(i)} \le u_n) = \exp(-\theta\tau)$$

The exponent θ is the extreme index and independent of $(u_n)_{n=1}^{\infty}$ (McNeil, 1998).

By the definition of the extreme index we have asymptotically

$$\Pr(M_n^{(i)} \le u_n) \approx \Pr\left(\widetilde{M_n^{(i)}} \le u_n\right)^{\theta} = F(u_n)^{n\cdot\theta}$$

Based on these results the VaR for a given level α can be estimated. Let z_α be the α-quantile of the GED, that means $G(z_\alpha) = \alpha$, hence

$$z_\alpha = G^{-1}(\alpha) = \mu - \frac{\sigma}{\xi}(1 - (1 - \log(\alpha))^{-\xi})$$

Asymptotically for $n \to \infty$ we have

$$G(z_\alpha) = \Pr\left(M_n^{(i)} \le z_\alpha\right) = F(z_\alpha)^{n\cdot\theta}$$

We estimate

$$VaR_\alpha = G^{-1}\left(\alpha^{n\cdot\theta}\right)$$

then asymptotically the marginal distribution yields

$$G(VaR_\alpha) = F^{n\cdot\theta}(VaR_\alpha) = \alpha^{n\cdot\theta}$$

thus $F(VaR_\alpha) = \alpha$ as claimed by the definition of Value-at-Risk.

The estimate for ES_α is then based on the estimate for VaR_α and the analytical integration, see Section 4.2

4.5.3. Parameter settings

To estimate the extreme index θ that accounts for extreme value clustering, we follow (Embrechts et al., 1997; McNeil, 1998). For a threshold $u > 0$ let N_u be the number of values X_t larger than u, $X_t > u$ for $t = 1, \dots, T$ and let K_u be the number of blocks that contain at least one value larger than u, $M_n^{(i)} > u$ for $i = 1, \dots, m$. Then the estimate for the extreme index is given by

$$\hat{\theta} = \frac{1}{n} \cdot \frac{\log\left(1 - \frac{K_u}{m}\right)}{\log\left(1 - \frac{N_u}{T}\right)}$$

For the threshold u a suitable quantile of the time series is chosen.

Finally, the block length n is chosen as a compromise: on the one hand the GED approximates the block maxima only in the limit $n \to \infty$. On the other hand, the number of blocks m should not be too small to have enough data to fit the distribution.

To test whether the parameter setting is appropriate, the Sherman goodness of fit test (Longin, 2000; Sherman, 1957) is used with

$$\Omega_m = \frac{1}{2} \sum_{j=0}^{m} \left| G(M_{n,i+1}) - G(M_{n,i}) - \frac{1}{m+1} \right|$$

Here G is the fitted GED and $(M_{n,i})_{i=1}^{m}$ is the order statistics of the block maxima $M_n^{(i)}$ with $G(M_{n,0}) = 0$ and $G(M_{n,m+1}) = 1$. The test statistics

$$\frac{\Omega_m - E_m}{\sqrt{D_m}}$$

is approximately normally distributed with $E_m = \left(\frac{m}{m+1}\right)^{m+1}$ and $D_m \approx \frac{1}{m} \cdot \frac{2e-5}{e^2}$ (Sherman, 1957). This test statistics can be used to test the hypotheses that the fitted GED models the distribution of the block maxima (Longin, 2000). Only if the test statistics yields large values the hypothesis is rejected.

4.6. The Peaks over Threshold method

The idea of the Peaks-over-Threshold method is to model the distribution of extreme values that excess a certain threshold. This has the advantage that all these extreme values are used to fit the distribution, while the Block-Maxima method only uses one extreme value for each block. Again, we analyze a time series

$$X_1, X_2, \dots X_T$$

that can be either a return series or a time series of losses.

4.6.1. The iid approach

Firstly, we consider a time series

$$Z_1, Z_2, \dots, Z_T$$

and for this section we assume that this time series consists of independent and identically distributed random variables. For a given threshold $u > 0$ let $N_u \in \mathbb{N}$ be the number of values Z_t, $t = 1, \dots, T$, that are larger than the threshold, $Z_t > u$, and let $Y_1, \dots, Y_{N_u}$ be the excesses so that

$$Z_j = Y_j + u$$

for $j = 1, \dots, N_u$.

Let F be the distribution function of the random variables Z_t. The distribution of excesses is then given by the conditional probability

$$F_u(y) = \Pr(Y \leq y \mid X > u) = \frac{F(y+u) - F(u)}{1 - F(u)}$$

In the limit $u \to \infty$ this distribution is given by the Generalized Pareto Distribution. This result is known as the Pickands-Balkema-de Haan Theorem (Embrechts et al., 1997, p. 165)

$$F_u(y) \xrightarrow{u \to \infty} G(y)$$

with

$$G(y) = 1 - \left(1 + \xi \cdot \frac{y}{\beta}\right)^{-\frac{1}{\xi}}$$

for $\xi \neq 0$ and

$$G(y) = 1 - \exp\left(-\frac{y}{\beta}\right)$$

for $\xi = 0$. Here $\xi \in \mathbb{R}$ is the shape parameter, $\beta > 0$ is the scale parameter and $y \geq 0$ for $\xi \geq 0$ and $y \in \left[0, -\frac{\beta}{\xi}\right]$ for $\xi < 0$. Under the assumption of independent identically distributed random variables these two parameters can be estimated using maximum likelihood estimation.

To estimate the VaR we rewrite the definition of F_u

$$F(x) = F(y + u) = 1 - (1 - F(u)) \cdot (1 - F_u(y))$$

The expression $1 - F(u)$ is the probability $\Pr(X > u)$ and can be estimated with $\frac{N_u}{T}$. Thus, we can approximate

$$F(x) \approx 1 - \frac{N_u}{T} \cdot (1 - G(y))$$

Inserting the definition of the GPD $G(y)$ we can estimate VaR_α for a given level α as the α-quantile of F, namely

$$VaR_\alpha = u + \frac{\beta}{\xi}\left(\left(\frac{T \cdot (1 - \alpha)}{N_u}\right)^{-\xi} - 1\right)$$

for $\xi \neq 0$ and

$$VaR_\alpha = u - \beta \cdot \log\left(\frac{T \cdot (1 - \alpha)}{N_u}\right)$$

for $\xi = 0$.

The estimate for ES_α is then based on the analytical integration, see Section 4.2, namely

$$ES_\alpha = \frac{VaR_\alpha}{1 - \xi} + \frac{\beta - \xi \cdot u}{1 - \xi}$$

for $\xi < 1$.

This method of estimating VaR and ES depends on the choice of threshold $u > 0$. On the one hand u needs to be large enough to justify the approximation, that holds only in the limit $u \to \infty$. On the other hand u cannot be chosen too large in order to have enough excesses to avoid artificial variance in the excesses.

We use two methods simultaneously to determine suitable thresholds. Firstly, we estimate the mean excess function, the expected value of excesses

$$e(u) = E(Z - u \mid Z > u)$$

by

$$\hat{e}(u) = \frac{1}{N_u} \cdot \sum_{j=1}^{N_u} (z_j - u)$$

where z_j, $j = 1, \ldots, N_u$, are the excesses larger than u. For a well-fitted GPD the mean excess function is linear in u (Embrechts et al., 1997). So, we choose a threshold u^* such that the estimated mean excess function is close to linear for $u > u^*$.

Secondly, we employ the Hill method (Hill, 1975; McNeil and Frey, 2000) that estimates a so-called tail index, by

$$\hat{\iota} = \frac{1}{N_u} \cdot \sum_{j=1}^{N_u} (\log z_{(j)} - \log z_{(j+1)})$$

where $z_{(j)}$ is the order statistics of the excesses z_j, $j = 1, \ldots, N_u$ and z_{N_u+1} is chosen as the threshold. The threshold u^* is then chosen in such a way that the function $\hat{\iota}(N_u)$ is stable in a neighborhood of N_{u^*} (McNeil and Frey, 2000).

4.6.2. The GARCH filter

To account for the fact that the return series $X_1, X_2, \dots X_T$ is not independent and identically distributed the Peaks-over-Threshold method is not directly applied but is used to derive the extreme value measures of residuals of a GARCH filter. We refer to (McNeil and Frey, 2000; Ruppert and Matteson, 2015) for details.

Let $X_1, X_2, \dots X_T$ be a stationary time series, e.g. the return series or the time series of losses. Then to apply a GARCH filter we set for all $t = 1, \dots, T$

$$X_t = \mu_t + \sigma_t \cdot Z_t$$

with $\mu_t \in \mathbb{R}$, $\sigma_t > 0$, and $Z_t, t = 1, \dots, T$, being a series of independent, identically distributed random variables with mean 0 and variance 1. An AR(1)-GARCH(1,1) model is fitted to the data to recursively determine the parameters $\widehat{\mu_t}$ and $\widehat{\sigma_t}$, $t = 1, \dots, T$. Since this is not specific for EVT we don't go into detail here but refer to (McNeil and Frey, 2000).

Finally, the time series of GARCH residuals

$$Z_t = \frac{(X_t - \widehat{\mu_t})}{\widehat{\sigma_t}}, t = 1, \dots, T,$$

is defined and the Peaks-over-Threshold method described in the previous section is applied to estimate $VaR_\alpha(Z)$ and $ES_\alpha(Z)$ for the residuals.

4.6.3. The GARCH scaling

The GARCH filter introduced in the previous section yields predictions $\widehat{\mu_{T+1}}$ and $\widehat{\sigma_{T+1}}$ for the parameters describing expected return and volatility respectively. Given these predictions and the extreme value estimates for the time series of residuals, the estimates for VaR and ES for the original time series $X_1, \dots X_T$ are given by

$$VaR_\alpha = \widehat{\mu_{T+1}} + \widehat{\sigma_{T+1}} \cdot VaR_\alpha(Z)$$
$$ES_\alpha = \widehat{\mu_{T+1}} + \widehat{\sigma_{T+1}} \cdot ES_\alpha(Z)$$

These are the final estimates for the extreme values based on the Peaks-over-Threshold method.

5. Extreme Value Theory applied to the gold price: Value at Risk and Expected Shortfall calculations

In this section we provide the main empirical results: the estimates for the extreme value measures VaR and ES and give an interpretation of the results.

5.1. Data basis and general setup

We analyze the gold price in the period from 02.01.1975 until 30.05.2025, i.e. the time series XAU, see Figure 1. This price data is publicly available for example from yahoo finance or Bloomberg.

Figure 1. Gold price 1975 until 2025.

We are interested in estimating the risk measures VaR and ES for various time periods using different methods from the EVT. Our goal is to compare the results from different time periods and use this to contrast the different methods. We refer to Section 4 for the formal definition of the measures and the theoretical background of EVT.

In order to calculate the measures, we initially determine the logarithmic returns from the prices, see Section 4.1. The time series of returns is our fundamental data, see Figure 2. We are mainly interested in the statistical properties of the VaR and ES estimates and the statistical dependencies of the methods of EVT. Hence, we use daily returns as our basic data.

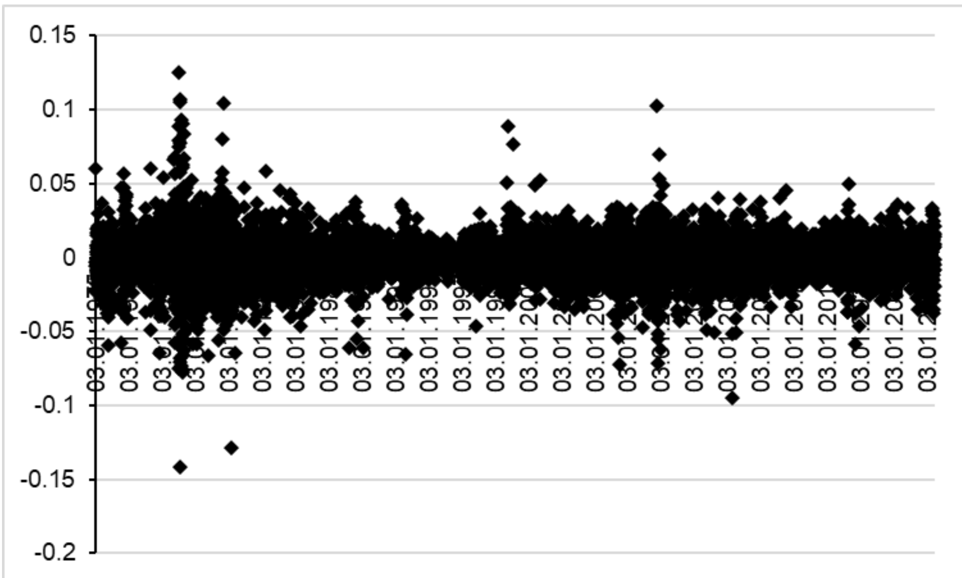

Figure 2. Logarithmic returns 1975 until 2025.

Accordingly, we require daily prices as input data and daily gold prices from the XAU time series are available only form 1975 onwards. Thus, we are working with 50 years of daily return data, there are 12,832 daily returns from 03.01.1975 until 31.12.2024. To investigate the effect of the first months of 2025 on the extreme value estimates, in Section 5.3 we also study the time series extended until 30.05.2025.

In this analysis we concentrate on the long-term view using all publicly available daily data. We are interested in obtaining VaR and ES estimates for rare events that only happen scarcely or, in other words, in view of long time periods. Specifically, we calculate the VaR and ES at the 0.99 and 0.995 level, see Section 4.2 for the formal definition. These levels represent extreme events that on average only happen once in every 100 and 200 days respectively. As one year has about 250 trading days, for the 0.99 level we expect to see about 25 of these rare events within a decade and for the 0.995 level we expect to see about 10 to 15 of the respective rare events.

Typically, these rare events are considered to be extreme daily losses, i.e. often the left tail of the distribution of daily returns is studied. This stems from the important applications of the VaR and ES measures in risk management. As we are mainly interested in the statistical properties, we also investigate the right tail, i.e. the positive tail, of the distribution. That means we also study extreme daily gains. To this end, the measures VaR and ES are applied to gains in a straightforward, symmetric way, see Section 4.2 for a formal definition and technical details. Of course, for gains the notions Value at Risk and Expected Shortfall are not appropriate, since the respective measures here describe potential gains. But we keep the same terminology for simplicity. By comparing the results for the left and the right tail of the distribution of returns we gain insights about the relation of gains and losses and about the symmetry of the return distribution.

To estimate the VaR and ES measures we apply two common methods from EVT, namely the Block-Maxima method with an extreme index (Embrechts et al., 1997; McNeil, 1998; McNeil et al., 2015) and the Peaks over Threshold method applied with a GARCH filter (Embrechts et al., 1997; McNeil and Frey, 2000). In Section 4.5 and 4.6 we provide the theoretical background and the relevant literature. In this section we concentrate on results and interpretation.

As a benchmark for these extreme value estimates we use the in-sample values, i.e. the historical VaR and ES values directly calculated from the time series of daily returns, see Section 4.3. We also provide the standard textbook estimates for VaR and ES that are derived from a normal distribution of daily returns, see e.g. (Ruppert and Matteson, 2015) and Section 4.4. These comparisons yield information about the statistical properties of the extreme value estimates.

In Section 5.3. we discuss the basic statistical properties of the timeseries of daily returns. Then we investigate the time period starting from 1975 as a whole and apply the aforementioned methods to derive and compare VaR and ES measures. In Section 5.4. the time dependence of the measures is studied by analyzing the five decades from 1975 until 2024 separately. Finally in Section 5.5. we explore the predictive quality of the extreme value estimates by conducting an out-of-sample comparison using again the data separated in consecutive decades.

5.2. Basic statistical data

The basic data consists of the 12935 daily returns in the period from 03.01.1975 until 30.05.2025. The basic statistical measures are summarized in Table 1.

Table 1. Statistical properties of the basic time series of returns

mean	median	min	max	stand. dev.	skewness	kurtosis
0.0211%	0.0161%	-14.20%	12.50%	1.19%	-0.0077	14.6

There is a small negative skewness, i.e. asymmetry to the left tail, and the kurtosis of 14.6 shows that the distribution of the daily returns has much fatter tails, i.e. more extreme positive and negative values, than a normal distribution. Both facts will be discussed in more detail below. Using an average of 250 trading days per year, the mean daily return of 0.0211% scales to an average yearly return of $250 \cdot 0.0211\% = 5{,}275\%$ and the daily standard deviation, i.e. the daily volatility, scales to a yearly volatility of $\sqrt{250} \cdot 1.19\% = 18{,}81\%$.

In fact, the standard deviation, i.e. the volatility, of daily returns varies substantially from year to year. Figure 3 shows the standard deviation of daily returns determined for each year separately.

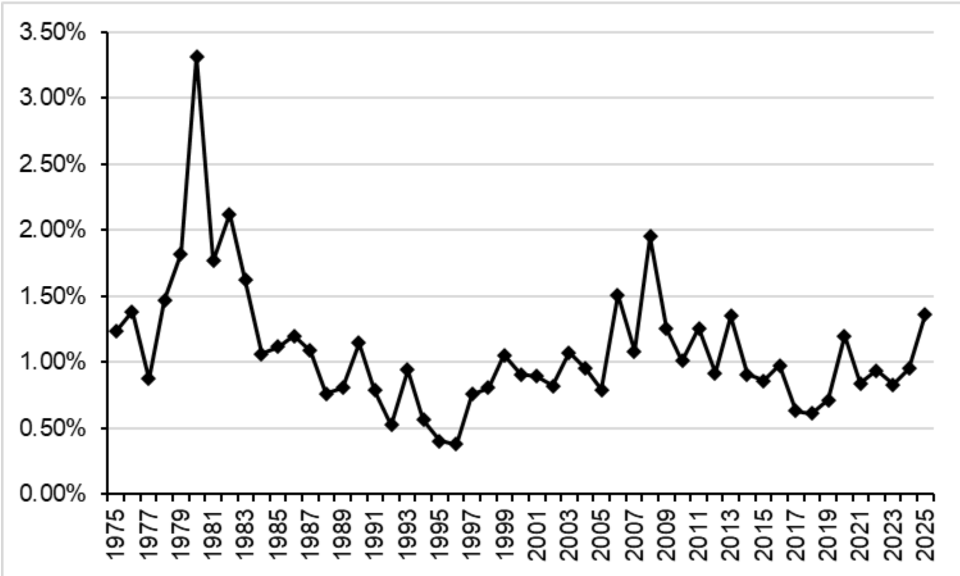

Figure 3. Standard deviation of daily returns for each year from 1975 until 2025.

The highest standard deviation 3.31% stems from the year 1980, followed by a standard deviation of 2.12% in 1982 and 1.95% in 2008. The first 5 months of 2025 exhibit a standard deviation of 1.36% which is not only higher than the overall long-term standard deviation of 1.19% but also the highest since 2008.

The years 1980, 1982 and 2008 also appear, when we look at the highest gains and losses throughout the whole period of 50 years. Table 2 shows the 13 largest daily gains and daily losses and the respective corresponding dates.

Table 2. Largest daily gains and losses from 1975 until 2025

rank	gains		losses	
1	03.01.1980	12.5%	22.01.1980	-14.2%
2	18.01.1980	10.7%	28.02.1983	-12.9%
3	16.01.1980	10.5%	15.04.2013	-9.5%
4	03.09.1982	10.5%	17.03.1980	-7.7%
5	17.09.2008	10.2%	26.03.1980	-7.5%
6	21.02.1980	9.3%	04.01.1980	-7.5%
7	19.03.1980	9.0%	20.02.1980	-7.4%
8	02.01.1980	8.9%	13.06.2006	-7.2%
9	28.09.1999	8.9%	10.10.2008	-7.2%
10	08.04.1980	8.3%	25.01.1980	-7.1%
11	20.08.1982	8.0%	28.01.1980	-6.8%

12	28.12.1979	7.9%	28.09.1981	-6.7%
13	29.01.1980	7.7%	21.03.1980	-6.6%

It is remarkable that the highest and second highest losses, -14.2% and -12.9%, are in absolute value much larger than the highest and second highest gains, 12.5% and 10.7%. But apart from that the extreme gains are larger in absolute value than the corresponding extreme losses. This can be summarized using the concept of the historical VaR and ES at level 0.999, see Section 4.3. The information shown in Table 2 leads to a VaR of -6.6% and an ES of -8.5% at level 0.999 for losses and a VaR of 7.7% and an ES of 9.6% at level 0.999 for gains. Thus, at this level the extreme values of the gains are larger than those of the losses.

5.3. Long-term view: Value at Risk and Expected Shortfall for the whole time period

In this section we calculate, compare, and discuss the VaR and ES values for the whole period from 03.01.1975 until 31.12.2024. To analyze the effect of the first months of 2025 we also show the values for the time series extended until 30.05.2025.

First, we give an overview of the model set-up for the two methods from EVT, the Block-Maxima method and the Peaks-over-Threshold method, and we summarize the relevant parameters. In the following subsections we summarize and discuss the results.

5.3.1. Model set-up

For the Block-Maxima method, see Section 4.5, in this case of the whole time period, we choose a block length of 63 which leads to about 200 blocks. We use the maximum likelihood estimate to fit the generalized extreme value distribution (GED). The aforementioned choice of block length thereby leads to reasonable results in the Sherman Test for the goodness of fit, see Section 4.5.3, with p-values above 12%. So, the hypotheses that the fitted distribution is suitable cannot be rejected.

To estimate the extreme index, we use the method described in Section 4.5.3. For the threshold we use eight quantiles from the 0.2%-quantile to the 3%-quantile. As an estimator for the extreme index, we then take the average of the respective extreme index values.

Table 3 summarizes the parameters used and obtained for the Block-Maxima method, in particular for the fitted GPD and the extreme index that are subsequently used to derive the VaR and ES estimates.

Table 3. Parameters for the Block-Maxima method

period	**1975 - 2024**	**1975 - 2025**	**1975 - 2024**	**1975 - 2025**
tail	negative	negative	positive	positive
block length	63	63	63	63
location	0.021	0.021	0.019	0.020
scale	0.010	0.010	0.009	0.009
shape	0.207	0.196	0.215	0.178
extreme index	0.587	0.550	0.462	0.469

For the Peaks-over-Threshold method, see Section 4.6, we first use an AR(1)-GARCH(1,1) model fitted to the time series of logarithmic returns. Then we use the mean excess method and the Hill method, see Section 4.6.1, to find a suitable threshold for the series of GARCH-fitted residuals. Then we use the maximum-likelihood estimate to fit the Generalized Pareto Distribution. Let us mention without going into details that a sensitivity analysis shows that the choice of GARCH model and the choice of threshold does not influence the results significantly. For example, we get very similar results using a GARCH(1.1) model and as a threshold the 90% quantile, as suggested in (McNeil and Frey, 2000).

Table 4 shows the relevant parameters that are used for the Peaks-over-Threshold method.

Table 4. Parameters for the Peaks-over-Threshold method

period	**1975 - 2024**	**1975 - 2025**	**1975 - 2024**	**1975 - 2025**
tail	negative	negative	positive	positive
threshold	2	2	2	2
scale	0.614	0.623	0.566	0.557
shape	0.172	0.163	0.192	0.198

5.3.2. Main results and discussion

In this section we show the results of the VaR and ES calculations for the 0.99 and 0.995 levels. We start with the VaR estimates for the negative tail, i.e. for the losses, and then continue the discussion with the results for the positive tail, i.e. for the gains. Finally, we compare the VaR estimates with the ES estimates.

Table 5 and Table 6 summarize the results of the VaR calculation for the period 03.01.1975 until 31.12.2024 and for the period extended until 30.05.2025 for the 0.99 and the 0.995 level respectively.

Table 5. VaR estimates at level 0.99

method	**1975 until 2024**	**1975 until 2025**
BM	-3.27%	-3.33%
PoT	-2.81%	-4.39%
Normal	-2.74%	-2.74%
Historical	-3.45%	-3.45%

Table 6. VaR estimates at level 0.995

method	**1975 until 2024**	**1975 until 2025**
BM	-4.23%	-4.27%
PoT	-3.38%	-5.28%
Normal	-3.03%	-3.04%
Historical	-4.08%	-4.08%

In general, the extent of the daily losses at these levels is rather modest.

For the period 1975 until 2024 the Block-Maxima method yields a VaR of -3.27% and -4.23% respectively. Comparing these results with the VaR estimates from a normal distribution show, that the normal distribution considerably underestimates the VaR and thus extreme losses. This fact will be seen throughout all the considered periods and is consistent with the high kurtosis of the return series, see Section 3.2.

The comparison with the historical VaR estimates, provides a kind of in-sample test. The results from the Block-Maxima method are remarkably close to the historical VaR with a difference of only 0.17%-points and -0.15%-points respectively.

However, the Peaks-over-Threshold estimates seem to underestimate with values of -2.81% for the 0.99 level and -3.38% for the 0.995 level. In absolute value they are substantially lower than the historical VaR estimates.

This picture changes entirely, when we extend the considered period until 30.05.2025. While the results for the Block-Maxima method, the normal method and the historical method are not affected at all by adding the first months of 2025, the results for the Points-Over-Threshold method change significantly: For the 0.99 level the VaR estimates changes by -1.58%-points and for the 0.995 level by -1.90%-points. These estimates are now the highest in absolute value and overestimate the in-sample benchmark by about 1%-point.

This is due to the high volatility of the return series in 2025 in comparison to the previous years. We will analyze this effect in detail in Section 5.3.3.

Now we turn our attention to the positive tail, i.e. the gains. Table 7 and Table 8 summarize the results of the VaR calculation for the period 03.01.1975 until 31.12.2024 and for the period extended until 30.05.2025 for the 0.99 and the 0.995 level respectively.

Table 7. VaR estimates at level 0.99

method	1975 until 2024	1975 until 2025
BM	3.25%	3.30%
PoT	2.75%	4.29%
Normal	2.78%	2.78%
Historical	3.23%	3.23%

Table 8. VaR estimates at level 0.995

method	1975 until 2024	1975 until 2025
BM	4.15%	4.17%
PoT	3.29%	5.11%
Normal	3.08%	3.08%
Historical	3.98%	3.98%

In general, the estimates for the gains show a similar behavior as the estimates for the losses. In particular, the Block-Maxima estimates are again remarkably close to the historical VaR, the normal distribution approach underestimates considerably and the Peaks-over-Threshold method underestimates for the period until 2024 and overestimates for the period extended until 2025.

Comparing the corresponding absolute values of the results for the losses with the results for the gains, we see that the VaR estimates for the losses are consistently larger in absolute value than for the gains. So, this behavior changes from the very extreme values shown in Section 5.2 (corresponding to a 0.999 level) to the 0.995 and in particular to the 0.99 level. Here we encounter a significant difference of the Block-Maxima estimates and the historical VaR: While the Block-Maxima estimates are very close to symmetric, in particular at the 0.99 level (with values of -3.27% and 3.25%) the historical VaR shows a small but substantial asymmetry with larger absolute values for the losses (with values of -3.45% and 3.23%).

The results for the Expected Shortfall estimates are similar in nature and we highlight the most remarkable points in the following. Table 9 and Table 10 summarize the results of the ES calculation for the negative tail for the period 03.01.1975 until 31.12.2024 and for the period extended until 30.05.2025 for the 0.99 and the 0.995 level respectively. Table 11 and Table 12 summarize the respective estimates for the positive tail.

Table 9. ES estimates at level 0.99

method	1975 until 2024	1975 until 2025
BM	-4.90%	-4.92%
PoT	-3.74%	-5.15%
Normal	-3.14%	-3.14%
Historical	-4.80%	-4.79%

Table 10. ES estimates at level 0.995

method	1975 until 2024	1975 until 2025
BM	-6.12%	-6.14%
PoT	-4.43%	-6.08%
Normal	-3.41%	-3.41%
Historical	-5.82%	-5.82%

Table 11. ES estimates at level 0.99

method	1975 until 2024	1975 until 2025
BM	4.78%	4.77%
PoT	3.65%	5.01%
Normal	3.18%	3.19%
Historical	4.72%	4.71%

Table 12. ES estimates at level 0.995

method	1975 until 2024	1975 until 2025
BM	5.93%	5.90%
PoT	4.31%	5.93%
Normal	3.45%	3.46%
Historical	5.93%	5.93%

For the ES the Block-Maxima method overestimates slightly for the negative tail at the 0.995 level in comparison to the in-sample historical ES estimate, but is again remarkably close, in particular for the positive tail.

The Points-over-Threshold ES estimates are in line with the VaR values, underestimating for the period until 2024 and overestimating for the extended period until 2025. This specific behavior is discussed in Section 5.3.3. Let us note, however, that the overestimate is much less significant than for the VaR estimates, in particular for the positive tail at level 0.995.

As expected, the ES estimates based on the normal distribution are much smaller in absolute value than all the other estimates. This effect is even more profound for the ES than for the VaR, further underlining the effect that the distribution of the returns of the gold price has much higher extreme values that a normal distribution.

Comparing the ES results for the losses with the results for the gains, we conclude that the ES estimates are rather symmetric in general. At the 0.995 level the Block-Maxima method yields losses that are in absolute value larger than the gains (with values of -6.14% and 5.90%). This effect, however, is not present in the sample data, i.e. in the historical ES. It is remarkable that the historical ES at 0.995 level is higher in absolute value for the gains than for the losses (with values of -5.82% and 5.93%). This was already mentioned in Section 5.2. and is quite unique for the gold price.

5.3.3. The effect of 2025: Analysis of the GARCH model

In this section we analyze the rather strong effect that the extension of the time period from 31.12.2024 to 30.05.2025 has on the estimates obtained by the Peaks-over-Threshold method. The first months of 2025 show a relatively high volatility compared to the rather low volatility of the years before, see Section 5.2. This influences the predicted volatility in the GARCH model fitted to the data and thus it influences the Peaks-over-Threshold estimates.

Calculating the Peaks-over-Threshold estimates consists of two major steps: Firstly, the Peaks-over-Threshold method is applied to calculate the respective measures of the generalized pareto distribution fitted to the GARCH residuals as described in Section 4.6.1 and 4.6.2. Secondly, these measures are scaled by the predicted volatility and shifted by the predicted return of the GARCH filter as described in Section 4.6.3.

For example, for the VaR at level 0.99 for the period until 31.12.2024 we obtain -2.658 for the quantile of the GARCH residuals and 0.0106 for the predicted volatility and 0.00003 for the predicted return. Thus, we end up with $0.00003 - 2.658 \cdot 0.0106 = -0.0281$ which is the VaR estimate shown in Table 5. For the period extended until 30.05.2025 we also obtain -2.658 for the quantile of the GARCH residuals, so here the extension has no effect. But we get 0.0165 for the predicted volatility and 0.00004 for the predicted return. In this case we obtain $0.00004 - 2.658 \cdot 0.0165 = -0.0439$ which again corresponds to the value shown in Table 5. We see that the increase of the predicted volatility from 1.06% to 1.65% leads to the increases of the VaR value from -2.81% to -4.39%. There is a similar effect for the other estimates obtained by the Peaks-over-Threshold method.

We remark that this effect of the first months of 2025 does not appear if we use the averaged volatility of the GARCH filter in the scaling described in Section 4.6.3 instead of the predicted volatility. If we proceed like this, the estimated Peaks-over-Threshold values are stable, in the sense that we get the same results for the period until 31.12.2024 and for the period extended until 30.05.2025. In Table 13 we show the estimates obtained in this manner for the negative tail.

Table 13. Peaks-over-Threshold estimates using an averaged volatility scaling

Level	VaR	ES
0.99	-2.87%	-3.81%
0.995	-3.44%	-4.51%

Comparing these estimates with the original values, we see that these averages are much closer to the estimates for the period until 31.12.2024 and thus tend to underestimate the in-sample estimates. Thus, from a risk-management point of view, this averaging is not favorable, even though it leads to stable results.

5.4. Time dependence: Value at Risk and Expected Shortfall for the five decades

In this section we analyze how the extreme value estimates for VaR and ES depend on the time period. To this end we split the return data from the period from 03.01.1975 until 31.12.2024 in 5 decades and use the respective data separately to derive the VaR and ES values. In this way we obtain five estimates for each value and each method and we compare the results.

Table 14 summarizes the basic properties of the respective return data. It is remarkable that an investment in gold yielded a strongly positive return during each decade. Note that the minimum and maximum values are in line with the reported data in Section 3.2. Also, the time development of the volatility (the standard deviation of daily returns) fits to the yearly development of the volatility shown in Figure 3 and to an overall volatility of 1.19% with high volatility in the first and the fourth decade and lower volatility in the remaining decades.

Table 14. Basic properties of returns by decade

	1975-1984	1984-1995	1995-2004	2005-2014	2014-2025
number of daily returns	2511	2510	2608	2608	2595
total return	56.9%	21.5%	13.5%	99.4%	79.7%
minimal return	-14.2%	-6.6%	-4.6%	-9.5%	-5.9%
maximal return	12.5%	5.8%	8.9%	10.2%	5.0%
std. dev. of daily returns	1.79%	0.92%	0.84%	1.25%	0.87%

In the following section we summarize the model set-up for the two methods from EVT, the Block-Maxima method and the Peaks-over-Threshold method, and the relevant parameters. In Sections 3.4.2. we present and discuss the results, how the EVT estimates depend on the decade.

5.4.1 Model set-up

For the Block-Maxima method, see Section 4.5, for each decade we choose a block length of 21 which leads to about 120 blocks. Again, we use the maximum likelihood estimate to fit the GPD. The choice of block length is again justified by the Sherman Test for the goodness of fit with p-values above 23%. So, the hypotheses that the fitted distribution is suitable cannot be rejected.

To estimate the extreme index, we again use the method described in Section 4.5.3 with eight quantiles from the 0.2%-quantile to the 3%-quantile. As an estimator for the extreme index, we then take the average of the respective extreme index values.

Table 15 summarizes the parameters used and obtained for the Block-Maxima method, in particular for the fitted generalized pareto distribution and the extreme index that are subsequently used to derive the VaR and ES estimates.

Table 15. Parameters for the Block-Maxima method

period	1975 - 1984	1985 - 1994	1995 - 2004	2005 – 2014	2015 - 2024
tail	negative	negative	negative	Negative	negative
block length	21	21	21	21	21
location	0.0207	0.0130	0.0117	0.0184	0.0133
scale	0.0120	0.0072	0.0054	0.0091	0.0055
shape	0.2333	0.2071	0.0793	0.1321	0.1150
extreme index	0.6806	0.9698	0.8581	0.8411	0.8412

period	1975 - 1984	1985 - 1994	1995 - 2004	2005 - 2014	2015 - 2024
tail	positive	positive	positive	positive	positive
block length	21	21	21	21	21
location	0.0199	0.0119	0.0108	0.0171	0.0130
scale	0.0106	0.0063	0.0060	0.0074	0.0050
shape	0.3137	0.1763	0.1814	0.0941	0.0732
extreme index	0.6036	0.8239	0.7851	0.8603	0.7868

For the Peaks-over-Threshold method, see Section 4.6, we again use an AR(1)-GARCH(1,1) model fitted to the time series of logarithmic returns. Then we use the mean excess method and the Hill method to find a suitable threshold for the series of GARCH-fitted residuals. Then we use the maximum-likelihood estimate to fit the GPD. Similar as for the results in Section 5.3. a sensitivity analysis shows that the choice of GARCH model and the choice of threshold does not influence the results significantly.

Table 16 shows the relevant parameters that are used for the Peaks-over-Threshold method.

Table 16. Parameters for the Peaks-over-Threshold method

period	1975 - 1984	1985 - 1994	1995 - 2004	2005 - 2014	2015 - 2024
tail	negative	negative	negative	negative	negative
threshold	2	1	1	1	1
scale	0.695	0.675	0.609	0.725	0.658
shape	0.078	0.096	0.028	-0.001	0.024

period	1975 - 1984	1985 - 1994	1995 - 2004	2005 - 2014	2015 - 2024
tail	positive	positive	positive	positive	positive
threshold	1.2	1	1	1	1
scale	0.579	0.588	0.600	0.579	0.658
shape	0.095	0.152	0.145	-0.005	-0.061

5.4.2. Main results and discussion

Here we present and discuss the results of the VaR and ES calculations for the individual decades at the 0.99 and the 0.995 level. First, we consider the VaR for the negative tail. Tables 17 and 18 and Figures 4 and 5 summarize and illustrate the results.

Table 17. VaR estimates at level 0.99

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	-5.02%	-2.66%	-2.15%	-3.60%	-2.38%
PoT	-2.77%	-1.19%	-2.17%	-3.52%	-2.65%
Normal	-4.15%	-2.13%	-1.94%	-2.86%	-1.99%
Historical	-5.11%	-2.61%	-2.31%	-3.69%	-2.34%

Table 18. VaR estimates at level 0.995

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	-6.45%	-3.41%	-2.59%	-4.43%	-2.86%
PoT	-3.35%	-1.45%	-2.56%	-4.14%	-3.12%
Normal	-4.60%	-2.36%	-2.15%	-3.17%	-2.21%
Historical	-6.28%	-3.53%	-2.72%	-4.26%	-2.92%

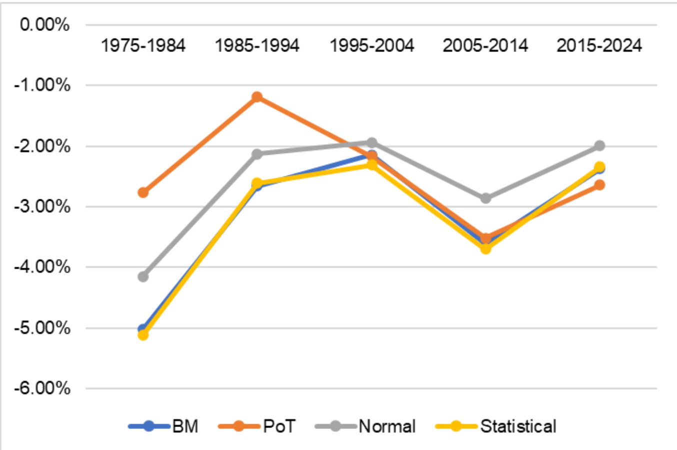

Figure 4. VaR estimates at level 0.99.

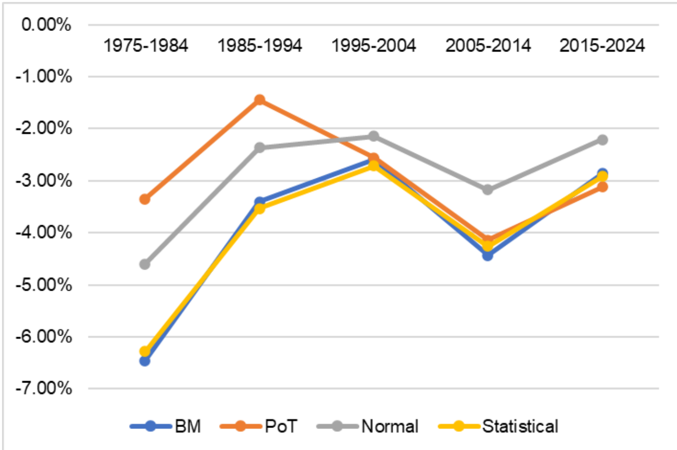

Figure 5. VaR estimates at level 0.995.

The Block-Maxima method yields VaR estimates that vary substantially between -5.02% in the first decade and -2.15% in the third decade at level 0.99 and between -6.45% in the first decade and -2.59% in the third decade at level 0.995. In particular, the values for these two decades differ substantially from the VaR estimates for the whole period from 1975 until 2024: -3.27% and -4.23% respectively, see Section 5.3.

We see that the Block-Maxima estimates are influenced by the volatility of the decade, i.e. the standard deviation of returns: The first decade has the highest and the third decade has the lowest standard deviation of returns, see Section 5.4.1. Actually, the VaR estimates based on the Block-Maxima method show a high negative correlation with the standard deviation of the decade.

Similar as the Block Maxima estimates for the whole period, the Block Maxima estimates for the decades are all remarkably close to the historical VaR estimates and hence show very good in-sample accuracy.

The VaR estimates based on the normal distribution share the strong dependence on standard deviation by definition, see Section 4.4. However, as expected, the normal distribution underestimates extreme values systematically as was already discussed in Section 5.3.

The Peaks-over-Threshold method yields VaR estimates that are more stable with a range of -3.52% to -1.19% at level 0.99 and -4.14$ to -1.45% at level 0.995. Most of the estimates are reasonably close to the values for the whole period form 1975 until 2024, namely -2.81% and -3.38% respectively. Here the second decade stands out with rather small VaR values in absolute value.

This can be explained by the properties of the GARCH models, see Section 4.6 and 5.3.3. The end of the second decade falls into a period of very low volatility with a standard deviation of 0.56% in 1994. So, we see that the Peaks-over-Threshold estimates are rather influenced by the volatility at the end of the considered period.

Now we turn our attention to the VaR estimates for the positive tail. Tables 19 and 20 and Figures 6 and 7 summarize and illustrate the results.

Table 19. VaR estimates at level 0.99

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	5.06%	2.48%	2.35%	3.08%	2.25%
PoT	2.76%	1.15%	2.38%	3.10%	2.53%
Normal	4.19%	2.15%	1.95%	2.94%	2.06%
Historical	5.20%	2.65%	2.08%	3.05%	2.17%

Table 20. VaR estimates at level 0.995

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	6.64%	3.12%	2.96%	3.70%	2.66%
PoT	3.30%	1.41%	2.91%	3.60%	2.90%
Normal	4.64%	2.38%	2.16%	3.25%	2.27%
Historical	6.81%	3.09%	2.65%	3.29%	2.46%

Generally speaking, the properties described for the negative tail can also be observed for the positive tail: The Block Maxima estimates are rather varying but are close to the historical VaR estimates and have high correlation with the standard deviation of the decade. The Peaks-over-Threshold estimates are more stable with dependence on the standard deviation at the end of the decade.

Having a closer look at the historical VaR estimates, thus at the data itself, shows remarkable asymmetries, in particular in the first and the fourth decade at level 0.995. In the first decade the VaR estimate for the positive tail, 6.81%, is by 0.53%-points higher in absolute value that the estimate for the negative tail. So here we have the rather uncommon behavior that extreme gains outweigh extreme losses, as was already observed in Section 5.2 and 5.3.2. In the fourth decade the historical VaR for the positive tail, 3.29%, is by 0.97%-points smaller in absolute value than the estimate for the negative tail. Both asymmetries are not that pronounced for the Block-Maxima and the Peaks-over-Threshold estimates. Note that the fourth decade somehow stands out due to a rather large difference between the Block-Maxima estimate and the historical VaR estimate. Hence, the methods from EVT seem to underestimate asymmetries at this level.

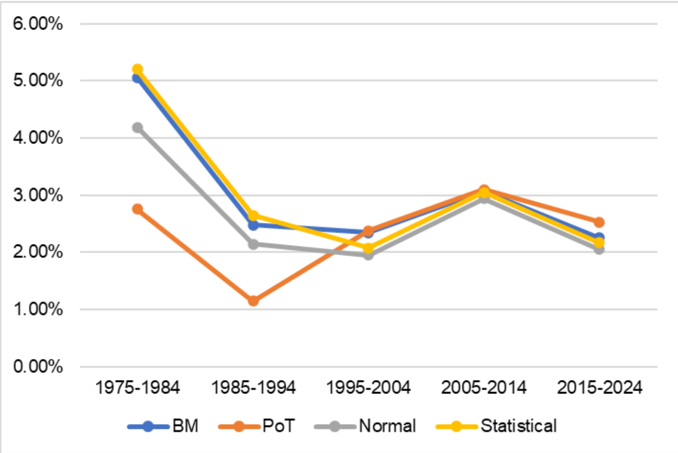

Figure 6. VaR estimates at level 0.99.

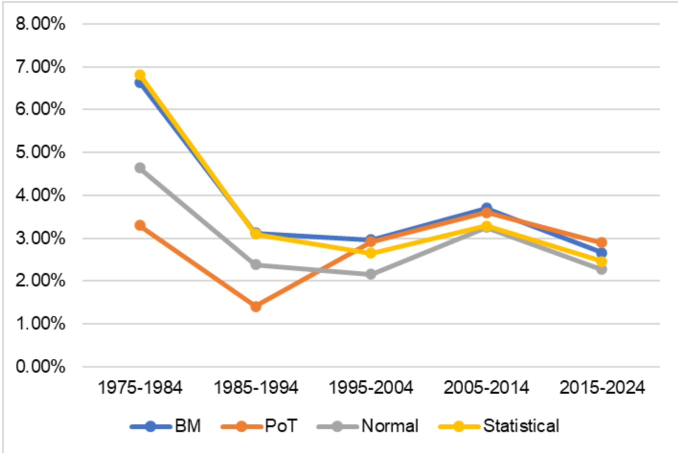

Figure 7. VaR estimates at level 0.995

To complete this section, in Tables 21, 22, 23, and 24 we present the results of the ES calculations. Overall, the ES estimates show similar behavior as the VaR estimates and we highlight some specific points in the following.

Table 21. ES estimates at level 0.99

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	-7.49%	-3.92%	-2.81%	-4.91%	-3.13%
PoT	-3.71%	-1.59%	-2.74%	-4.42%	-3.34%
Normal	-4.76%	-2.44%	-2.22%	-3.28%	-2.29%
Historical	-6.83%	-3.79%	-2.83%	-4.84%	-3.13%

Table 22. ES estimates at level 0.995

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	-9.36%	-4.86%	-3.29%	-5.87%	-3.67%
PoT	-4.40%	-1.87%	-3.14%	-5.04%	-3.83%
Normal	-5.16%	-2.65%	-2.41%	-3.57%	-2.49%
Historical	-7.97%	-4.68%	-3.18%	-5.69%	-3.67%

Table 23. ES estimates at level 0.99

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	8.01%	3.52%	3.36%	4.03%	2.72%
PoT	3.59%	1.57%	3.24%	3.81%	3.04%
Normal	4.80%	2.46%	2.23%	3.36%	2.35%
Historical	7.47%	3.35%	3.29%	3.88%	2.85%

Table 24. ES estimates at level 0.995

method	1975-1984	1985-1994	1995-2004	2005-2014	2015-2024
BM	10.29%	4.30%	4.11%	4.72%	3.30%
PoT	4.18%	1.88%	3.86%	4.30%	3.39%
Normal	5.21%	2.67%	2.42%	3.64%	2.55%
Historical	9.06%	3.86%	4.18%	4.54%	3.37%

For the first decade the different methods yield rather differing results, e.g. the Block-Maxima method yields -9.36% at level 0.995 while the Peaks-over-Threshold method yields -4,40%. In comparison to the historical VaR estimate of -7.97% the Block-Maxima method overestimates while the Peaks-over-Threshold method underestimates substantially. The overestimate of the Block-Maxima method seems to be due to the high volatility and the high number of very extreme returns during the first decade. The underestimate of the Peaks-over-Threshold method can again be explained by the GARCH model, see Section 4.6 and 5.3.3. While the volatility in the first decade was high with a standard deviation of 1.79%, towards the end of the decade the volatility was low with a standard deviation of 1.06% in 1984.

Let us also note that the historical ES estimates, similar as the historical VaR estimates, show a rather strong asymmetry towards the positive tail in the first decade and a strong asymmetry towards the negative tail in the fourth decade. For the ES estimate this behavior is also present in the estimates based on the Block-Maxima method.

5.5. Predictive quality of Value at Risk and Expected Shortfall estimates

In this section we explore the predictive qualities of the EVT estimates. As before we are interested in the long-term view. Here we check, how well the EVT estimates for VaR and ES from one decade describe and predict the actual behavior of extreme values in the following decade. To this end, we compare the theoretically expected number of VaR breaches, i.e. the number of returns that are in absolute value larger than the VaR, with the actual number of VaR breaches and we compare the estimated ES values with the actual mean of the respective VaR breaches.

By definition of the VaR, for the 0.99 level we expect 1% of returns to be larger in absolute value than the VaR. E.g. for a decade with 2500 trading days 25 returns are expected to breach the VaR. Given a decade, we take the VaR estimate from the previous decade, and we compare this expected number of VaR breaches with the actual number of VaR breaches. So, we take information that is available at the beginning of a decade to predict the number of extreme events during the decade.

For example, for the decade 2015 until 2024 the VaR estimate from the previous decade 2005 until 2014 is -3.60% for the 0.99 level, see Section 5.4.2. (in this example we use the Block-Maxima values). The actual number of returns from 2015 until 2024 that are more negative than -3.60% is 5. Comparing this with the expected number of 25 leads to the conclusion that the VaR of -3.60% is overestimated for the decade 2015 until 2024.

There is a controversy how to test ES estimates in a meaningful way (Gneiting, 2011; Székely and Acerbi, 2014). Nevertheless, we complement the prediction of VaR estimates with a comparison for the estimates of ES values. To this end we directly take the ES estimate of a decade as a prediction for the following decade and compare this value with the mean of the actual VaR breaches mentioned above. In the aforementioned example the ES estimate from 2005 until 2014 is -4.91% for the 0.99 level. The comparative value, the average of the 5 VaR breaches is -4.31%. So, the ES prediction is also slightly overestimated.

These comparisons should illustrate the predictive qualities of the EVT estimates. For simplicity, we restrict this presentation to the negative tail and the 0.99 level. The findings for the 0.995 level and the positive tail are similar in nature. Tables 25 and 26 summarize the comparisons.

Table 25. Prediction values for the Block-Maxima estimates

	1985-1994	1995-2004	2005-2014	2015-2024

VaR estimate from previous decade	-5.02%	-2.66%	-2.15%	-3.60%
number of returns	2510	2608	2608	2595
expected number of VaR breaches	25	26	26	25
actual number of VaR breaches	4	16	112	5
ES estimate from previous decade	-7.49%	-3.92%	-2.81%	-4.91%
actual mean of VaR breaches	-6.08%	-3.11%	-3.18%	-4.31%

Table 26. Prediction values for the Peaks-over-Threshold estimates

	1985-1994	1995-2004	2005-2014	2015-2024
VaR estimate from previous decade	-2.77%	-1.19%	-2.17%	-3.52%
number of returns	2510	2608	2608	2595
expected number of VaR breaches	25	26	26	25
actual number of VaR breaches	23	168	107	6
ES estimate from previous decade	-3.71%	-1.59%	-2.74%	-4.42%
actual mean of VaR breaches	-3.94%	-1.72%	-3.23%	-4.18%

In Section 3.4.2. it is described that the VaR estimates based on the Block-Maxima method are rather varying from decade to decade. The dependence on volatility leads to substantial over- and underestimates if these values are used as predictions. The actual number of VaR breaches for the Block-Maxima method varies from 4 in 1985-1994 to 112 in 2005-2014. Both are essentially an order of magnitude away from the expected number of 25 or 26.

Under the Hypothesis that the distribution of the extreme events given by EVT from the previous decade is still valid for the decade under consideration, both the event of 4 or less VaR breaches and the event of 112 or more VaR breaches have p-values below 1%. Hence, with high significance we can reject the hypothesis that the distribution of extreme events fitted to the return data of a decade can be employed for long-term predictions, i.e. to predict extreme events for the next decade.

Even though the VaR estimates based on the Peaks-over-Threshold method are more stable, see Section 3.4.2., the prediction is again substantially over- und underestimating. In particular, for the decades 1995-2004 and 2005-2014 the actual numbers of VaR breaches 168 and 107 respectively are much higher than the expected 26. Again, this means the hypothesis that the distribution of extreme events of one decade fits to predict the extreme events of the following decade can be rejected with p-values below 1%.

It is remarkable that for both methods, the ES predictions are much closer to the actual mean of VaR breaches. Here predictions and actual values follow each other rather accurately from decade to decade. This, however, is mainly due to the fact that overestimated VaR and ES values lead to only few VaR breaches that are large in absolute value and thus to a large actual mean of VaR breaches, while underestimated VaR and ES leads to a rather large number of VaR breaches that are rather low in absolute value. Hence using the ES in stead of the VaR eases the prediction and leads to better results.

5.6 Summary of empirical results

The previous sections provided a comprehensive overview of VaR and ES values for the return series based on the gold price for the period 03.01.1975 until 31.12.2024. Using different methods for the VaR and ES calculations, analyzing different time periods, and calculating result for losses and gains allowed us to derive statistical properties of the gold price and to compare dependencies of different methods from EVT.

The whole period 1975 until 2024 exhibits rather modest risk for daily returns of the gold price: At level 0.99 the VaR is slightly higher than 3% in absolute value and the ES is slightly smaller than 5%. For the 0.995 level the VaR estimates are close to 4% and the ES is slightly lower than 6%, all in absolute value. These values are rather consistent for the Block-Maxima method and the historical VaR definition, as well as for losses and gains.

Overall, the estimates based on the Block-Maxima method are close to the values based on the historical method and thus show very good in-sample accuracy. The Block-Maxima values for the whole period are also stable: Extending the time series to 30.05.2025 has a negligible effect. However, the Block-Maxima values vary substantially from decade to decade. We see that the VaR and ES estimates based on this method show a strong dependence on the overall volatility, i.e. the standard deviation of daily returns, of the considered period.

The estimates based on the Peaks-over-Threshold method are more stable from decade to decade. However, considering the whole period from 1975 until 2024, extending this period to 30.05.2025 changes the values. This can be explained by the observation that the estimates based on the Peaks-over-Threshold method with a GARCH filter are rather influenced by the volatility at the end of the considered period.

In general, the estimates based on the normal distribution are smaller in absolute value than the other estimates. As usual for financial returns, the distribution of returns of the gold price has much higher extreme values than a normal distribution.

Comparing the results for the losses with the results for the gains, we see some remarkable asymmetries in the return series based on the gold price. At level 0.99 the VaR estimates, in particular the historical VaR, is larger in absolute value for the losses than for the gains. This behavior changes at level 0.995. Especially the historical ES at this level shows larger gains than losses. It is remarkable that, in this sense, for the gold price the extreme gains seem to outweigh the extreme losses.

These asymmetries are most profound in the historical VaR and ES estimates. The methods from EVT seem to balance asymmetries.

Finally, we saw that long-term predictions of the VaR measure is cumbersome. The hypothesis that return data of one decade is suitable to estimate the VaR for the following decade can be rejected with high significance.

6. Conclusions

This study addressed seven research questions about the role of gold as a financial asset. The focus was on its extreme risk characteristics and safe-haven properties. The study applied EVT and the related risk measures VaR and ES. The findings provide several important insights.

The results clearly demonstrate that EVT is a suitable and powerful tool for capturing the distributional properties of extreme gold price movements. Unlike traditional approaches, which often underestimate tail risk, EVT explicitly models the behavior of extreme returns, allowing for a more accurate assessment of the risks associated with rare yet impactful market events. These results confirm the robustness and appropriateness of EVT in the context of precious metal markets.

The aim of research question 1 was to analyze how extreme swings in the gold price evolved between 1975 and 2024 and to what extent these movements can be correlated with geopolitical or macroeconomic events.

The descriptive analysis of the largest daily losses and gains (see Table 2 in Section 5.2) shows a striking accumulation of extreme price movements in the early 1980s, particularly in connection with the dramatic developments in January 1980. Both the highest daily gain (+12.5% on January 3, 1980) and the sharpest daily loss (-14.2% on January 22, 1980) occurred in close temporal proximity. These events can plausibly be linked to the geopolitical uncertainty resulting from the Iranian Revolution, the Soviet invasion of Afghanistan and the escalating inflation expectations in the US at the time.

Another striking cluster of extreme losses in the gold price occurred in 2008 and 2013, with both phases linked to significant macroeconomic developments. In October 2008, at the height of the global financial crisis following the collapse of Lehman Brothers, there were massive short-term price losses on the gold market. Although gold is often considered a safe haven in times of crisis, liquidity shortages and margin calls led to widespread selling at that time, resulting in a daily loss of –7.2% on October 10, 2008. The fall in the price of gold in April 2013 was particularly striking, with a decline of –9.5% on April 15, 2013, representing the third-largest daily loss in the entire period under review. This event is linked to the monetary policy shift in the US. In spring 2013, there were already increasing signs that the US Federal Reserve under Ben Bernanke could scale back its expansionary monetary policy – in particular its government bond purchase program (“quantitative easing”) – in the medium term. The prospect of an end to ultra-loose monetary policy (“tapering”) led to a loss of confidence in gold as a hedge against inflation. This led to speculative selling, exacerbated by technical factors such as margin calls and the reduction of institutional gold holdings.

The quantitative application of the EVT – using both the Block-Maxima method and the Peaks-over-Threshold method – confirms these observations in the context of the decade-by-decade analysis (see section 5.4). The first decade (1975–1984) not only shows the highest volatility (standard deviation of daily returns: 1.79%, see Table 14), but also the highest VaR and ES estimates across all decades (see Tables 17, 18, 21, 22).

The decades 1995–2004 and 2015–2024, on the other hand, show significantly reduced extreme values, both in terms of magnitude and frequency, which is reflected in significantly lower EVT estimates. It is striking that in both periods, the general level of volatility and the extent of geopolitical tensions have also been relativized, which suggests a link between market stress and extreme behavior.

The quantitative research questions 2 to 6 were studied and answered in detail in Chapter 5. In summary, comparing downside risk and upside opportunities showed some remarkable asymmetries in the return series based on the gold price. At level 0.99 the VaR estimates are larger in absolute value for the losses than for the gains. This behavior changes at level 0.995. At this level the risk measures tend to show larger gains than losses. It is remarkable that, in this sense, for the gold price the extreme gains seem to outweigh the extreme losses. This asymmetry is unusual in the context of financial time series and points to a special market structure during speculative exaggerations.

We have also seen that extreme value estimates based on the Block-Maxima method are close to the estimates based on the historical method and thus show very good in-sample accuracy. The Block-Maxima method also reveals a dependence on the overall volatility, i.e. the standard deviation of daily returns, of the considered period. The estimates based on the Peaks-over-Threshold method are more stable from decade to decade. However, they are rather influenced by the volatility at the end of the considered period.

Finally, we saw that the hypothesis that return data of one decade is suitable to estimate extreme values for the following decade can be rejected with high significance.

Research question 7 examined whether the opportunities and risks of gold investments differ in different geopolitical and macroeconomic phases.

The analysis of VaR and ES values over five decades allows for a comparative assessment of the risks and opportunities of the gold market in different geopolitical and economic phases. The analysis shows that gold as an asset class offers both significant risks and exceptional opportunities for gains in times of crisis, albeit to varying degrees and in different directions, depending on the macroeconomic environment of the respective decade.

Particularly striking is the first decade (1975–1984), which was marked by the oil crises of the 1970s, high inflation, the collapse of the Bretton Woods system, and political uncertainties such as the Iranian Revolution and the Soviet invasion of Afghanistan. During this phase, the gold price not only exhibited the highest volatility of the entire analysis, but also by far the most extreme daily swings – both in terms of losses and gains. The extreme risk measures for this decade (e.g., VaR –6.45% and ES –9.36% at the 0.995 level, see Tables 18 and 22) thus directly reflect the uncertain geopolitical climate. At the same time, the high positive swings demonstrate gold's potential as a “crisis winner”.

In the following decades, the risk profile changed significantly. The 1990s, for example, a period of relative geopolitical stability and economic expansion (due in part to globalization and the end of the Cold War), saw the lowest VaR and ES values over the entire period. The risk and profit potential of gold was limited during this era, which is also reflected in the comparatively low volatility (see Table 14).

In contrast, the decade 2005–2014 shows increased extreme values, particularly in the context of the global financial crisis of 2008, the effects of which also reached the gold market. During the crisis, there were significant price swings – both due to selling pressure in the liquidity phase and subsequent sharp price rises in the wake of monetary policy easing. The extreme values of this decade are reflected, for example, in the block maximum-based VaR of –3.60% (0.99 level) (see Table 17) and a high actual number of VaR violations in the following decade (see Table 25).

The decade 2015–2024, which has been marked by the COVID-19 pandemic, the invasion of Ukraine in 2022, and increasing geopolitical tensions between major powers, among other things, again shows increased asymmetry in the extreme values – especially on the positive side. The statistical ES at the 0.995 level is 5.93% for gains, compared with –5.82% for losses (see Tables 10 and 12), which points to gold's role as a safe haven in times of political and economic uncertainty.

In summary, the decade-by-decade analysis shows that the risk and opportunity profiles of gold differ significantly across economic eras. In times of geopolitical instability or monetary policy upheaval, extreme values dominate, while more stable phases with lower volatility also lead to significantly reduced risk measures and upside potential.

The practical implications of our study are twofold. For investors, gold remains an effective diversification tool for portfolios, but it should not be considered a standalone or guaranteed safe haven. Its protective role is situational and should be integrated with other risk management tools. For risk managers, EVT-based VaR and ES offer a more reliable methodological foundation for stress-testing portfolios and managing downside risk than conventional models do. Thus, the findings encourage practitioners to incorporate EVT into their risk management frameworks.

This study provides valuable insights into the extreme risk behavior of gold and its role as a safe haven. However, several limitations should be acknowledged.

From a research perspective, the study highlights the potential of EVT to refine the measurement of extreme risks in financial markets. The analysis relies heavily on EVT to model tail risks. While EVT is well suited to capturing extreme observations, it is sensitive to the choice of thresholds and block sizes. Furthermore, EVT assumes stationarity in return distributions, a premise that may not hold in the presence of structural breaks or regime shifts.

Future research should analyze the dependence on parameters in more detail and explore the balance of asymmetries in a broader context. Moreover, a thorough and extensive backtesting including daily predictions of VaR and ES values could shed more light on the predictive qualities of EVT estimates. Further studies also may extend the analysis to other asset classes or explore nonlinear dependencies between gold and macroeconomic variables under stress.

The study focuses on historical gold price data from a specific time period. While this allows for meaningful statistical inference, it limits the generalizability of the findings. Crisis-specific dynamics, such as the Global Financial Crisis of 2008 versus the COVID-19 pandemic, may produce different safe-haven behaviors that cannot be fully captured within a single dataset. Furthermore, excluding intraday or high-frequency data means that short-term extreme movements in gold prices remain unexplored.

The safe-haven properties of gold may vary across different economies and currencies. The study predominantly examines global and U.S. dollar-denominated gold markets, which might not reflect safe-haven behavior in emerging markets or under different monetary regimes. Additionally, the interplay between gold and monetary policy, exchange rate regimes, and geopolitical shocks is not modeled.

Although VaR and ES based on EVT are theoretically superior measures of extreme risk, their application in practice is not without challenges. The computational complexity and data requirements of EVT may pose barriers to adoption for many investors and practitioners. Furthermore, reliance on historical data inherently limits predictive power, particularly when facing unprecedented or non-recurring crisis events.

In summary, this study provides strong evidence that EVT is a robust framework for analyzing extreme risks in the gold market. Gold continues to play a dual role as a hedge in normal times and a conditional safe haven in times of crisis. While not flawless, gold remains an indispensable asset for academic investigation and practical portfolio management, especially when evaluated using rigorous tail-risk methodologies, such as EVT.

References

Aizenman, J., Inoue, K., 2013. Central banks and gold puzzles. Journal of the Japanese and International Economies, 28(C), 69–90. https://doi.org/10.1016/j.jjie.2013.02.001

Batten, J. A., Ciner, C., Lucey, B. M. 2010. The Macroeconomic Determinants of Volatility in Precious Metals Markets. Resources Policy 35 (2): 65–71. https://doi.10.1016/j.resourpol.2009.12.002

Baur, D. G., Lucey, B. M. 2010. Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold. Financial Review 45 (2): 217–229. https://doi.org/10.1111/j.1540-6288.2010.00244.x

Baur, D. G., McDermott, T. K. 2016. Why Is Gold a Safe Haven? Journal of Behavioral and Experimental Finance 10: 63–71. https://doi.org/10.1016/j.jbef.2016.03.002.

Beckmann, J., Berger, Th., Czudaj, R., 2015. Does Gold Act as a Hedge or a Safe Haven for Stocks? A Smooth Transition Approach. Economic Modelling 48, 16–24. https://doi.org/10.1016/j.econmod.2014.10.044

Capie, F., Mills, T. C., Wood, G., 2005. Gold as a Hedge Against the Dollar. Journal of International Financial Markets, Institutions and Money 15 (4): 343–352. https://doi.org/10.1016/j.intfin.2004.07.002

Chaithep, K., Sriboonchitta, S., Chaiboonsri, C., & Pastpipatkul, P., 2012. Value at Risk Analysis of Gold Price Returns Using Extreme Value Theory. The Empirical Econometrics and Quantitative Economics Letters, 1(4), 151–168

Cheema, M. A., Faff, R., Szulczuk, K., 2020. The 2008 Global Financial Crisis and COVID-19 Pandemic: How Safe Are the Safe Haven Assets?, International Review of Financial Analysis, 83, Article 102316. https://doi.org/10.1016/j.irfa.2022.102316

Chinhamu, K., Huang, C.-K., Huang, C.-S., Chikobvu, D., 2014. Extreme Risk, Value-At-Risk And Expected Shortfall In The Gold Market. International Business & Economics Research Journal, 14(1), 107–122. https://doi.org/10.19030/iber.v14i1.9035

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics. Springer London, London. https://doi.org/10.1007/978-1-4471-3675-0

Eichengreen, B., 2008. Globalizing Capital: A History of the International Monetary System. 2nd ed. Princeton: Princeton University Press

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33483-2

Giles, D. E., Chen, Q., 2014. Risk Analysis for Three Precious Metals: An Application of Extreme Value Theory (Econometrics Working Papers No. 1402). Department of Economics, University of Victoria. Retrieved from RePEc [RePEc handle]

Gilli, M., & Këllezi, E., 2006. An application of extreme value theory for measuring financial risk. Computational Economics, 27(2), 207–228. https://doi.org/10.1007/s10614-006-9025-7

Gkillas, K., Longin, F., 2019. Is Bitcoin the new digital gold? Evidence from extreme price movements in financial markets. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3245571

Gneiting, T., 2011. Making and Evaluating Point Forecasts. Journal of the American Statistical Association 106, 746–762. https://doi.org/10.1198/jasa.2011.r10138

Hill, B.M., 1975. A Simple General Approach to Inference About the Tail of a Distribution. Ann. Statist. 3. https://doi.org/10.1214/aos/1176343247

Ilesanmi, K. D., Tewari, D. D., 2020. Financial Stress Index and Economic Activity in South Africa: New Evidence. Economies, 8(4), 110. https://doi.org/10.3390/economies8040110

Kaewkheaw, M., Leeahtam, P., Chaiboonsri, C., 2014. An analysis of relationship between gold price and U.S. dollar index by using bivariate extreme value copulas. In S. Sanguanbhokai, S. Premasathian, W. Bholsithi (Eds.), Stochastic processes and simulation (pp. 455–462). Springer. https://doi.org/10.1007/978-3-319-03395-2_29

Khan, G. R., Abdulrahman, A. T., Alamri, O., Iqbal, Z., Ahmad, M., 2021. Risk Analysis of Gold Prices in Pakistan Using Extreme Value Theory. Mathematical Problems in Engineering, 2021, Article ID 2704142. https://doi.org/10.1155/2021/2704142

Khemawanit, K., Tansuchat, R., 2016. The Analysis of Value at Risk for Precious Metal Returns by Applying Extreme Value Theory, Copula Model and GARCH Model. International Journal of Applied Business and Economic Research, 24(2), 1011–1025.

Liu, W.-H., 2013. Is Gold a Safe Haven Asset? A Revisit (SSRN Working Paper). SSRN. https://doi.org/10.2139/ssrn.2312117

Liu, W.-H., 2020. Are gold and government bond safe-haven assets? An extremal quantile regression analysis. International Review of Finance, 20(2), 451–483. https://doi.org/10.1111/irfi.12232

Longin, F.M., 2000. From value at risk to stress testing: The extreme value approach. Journal of Banking & Finance 24, 1097–1130. https://doi.org/10.1016/S0378-4266(99)00077-1

Longin, F.M., 1996. The Asymptotic Distribution of Extreme Stock Market Returns. The Journal of Business 69, 383–408.

McNeil, A.J., 1998. Calculating quantile risk measures for financial return series using extreme value theory. ETH Zurich. https://doi.org/10.3929/ETHZ-A-004320029

McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7, 271–300. https://doi.org/10.1016/S0927-5398(00)00012-8

McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative risk management: concepts, techniques and tools, Revised edition. ed, Princeton series in finance. Princeton University Press, Princeton, NJ.

Reboredo, J. C., 2013. Is gold a safe haven or a hedge for the US dollar? Implications for risk management. Journal of Banking & Finance, 37(8), 2665–2676. https://doi.org/10.1016/j.jbankfin.2013.03.020

Ruppert, D., Matteson, D.S., 2015. Statistics and Data Analysis for Financial Engineering: with R examples, Springer Texts in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4939-2614-5

Sherman, B., 1957. Percentiles of the ω_n Statistic. Ann. Math. Statist. 28, 259–261. https://doi.org/10.1214/aoms/1177707052

Smales, L. A., 2021. Investor attention and global market returns during the COVID-19 crisis. International Review of Financial Analysis, 73, Article 101616. https://doi.org/10.1016/j.irfa.2020.101616

Smith, R.L., 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika 72, 67–90. https://doi.org/10.1093/biomet/72.1.67

Székely, B., Acerbi, C., 2014. Back-testing expected shortfall. Risk 27, 76–81.

Tang, Y., Zhong, J., 2023. Predicting gold volatility: Exploring the impact of extreme risk in the international commodity market. Finance Research Letters, 58(PB), Article 104491. https://doi.org/10.1016/j.frl.2023.104491