International Journal of Economics and Financial Issues

ISSN: 2146-4138

available at http: www.econjournals.com

International Journal of Economics and Financial Issues, 2026, 16(5), 82-95.

Value-at-Risk and Expected Shortfall Estimation for the Moroccan Stock Market: A Comparative EVT Approach with GARCH Filtering

Hind Mansour[1], Amina El Bernoussi[2]*, Mohamed Dakkon[1]

[1]Laboratory of Management, Finance, Digitalization and Applied Statistics, Faculty of Legal, Economic and Social Sciences, Abdelmalek Essaadi University, Tetouan, Morocco, [2]Laboratory Analysis, Geometry and Applications, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco. *Email: amina.elbernoussi@uit.ac.ma

Received: 24 May 2026 **Accepted:** 25 August 2026 **DOI:** https://doi.org/10.32479/ijefi.24748

ABSTRACT

Portfolio management and regulatory requirements rely on the ability to accurately measure extreme market risk, and such events are more evident in emerging markets. In this paper, the value-at-risk (VaR) and expected shortfall (ES) for the Moroccan all shares index (MASI) are estimated between 2006 and 2026 using 4 estimation methodologies: Historical simulation, normal model, block-maxima (BM) and peaks-over-threshold (POT) with GARCH filtering. Backtesting shows that the Normal model underestimates tail risk systematically and the BM method provides the most conservative tail risk estimates ($V_aR_{99.9\%} = 6.07\%$, $ES_{99.9\%} = 11.45\%$), thus making it appropriate for regulatory stress- testing. The only method that is not rejected by the Kupiec test for all the confidence levels is the POT-GARCH which infers more moderate estimates ($V\ aR_{99.9\%} = 4.61\%$ and $ES_{99.9\%} = 5.61\%$), the most appropriate one for risk monitoring daily over time. The findings are useful in extreme risk management in emerging African markets, and add to the body of literature relating to EVT applications in this context.

Keywords: Moroccan All Shares Index, Extreme Value Theory, Value-at-Risk, Expected Shortfall, Backtesting

JEL Classifications: C22, C58, G15, G32, C49

1. INTRODUCTION

Financial history offers no shortage of reminders that extreme losses occur far more often than a Gaussian world would predict. The 2008 crisis, the 2020 COVID shock, and repeated episodes of currency and commodity stress all point to the same conclusion: Return distributions have tails that conventional models systematically underestimate. Getting this tail right is not a the- oretical nicety; it feeds directly into capital requirements, portfolio decisions, and the stability of the institutions that hold these positions.

The emerging two industry benchmarks are: Value at risk and expected shortfall. By using VaR, the range of where an investor's portfolio will not lose the amount of money is provided at a given confidence level over a given period of time. ES, however, averages the losses beyond that horizon, a distinction which is relevant as the two measures can rank models differently; and therefore lead to different capital charges for the same portfolio (Yamai and Yoshiba, 2005). Basel III and IV incorporate VaR and Expected Shortfall in their market risk capital framework. However, the effectiveness of these measures rests very much on the underlying distributional assumptions. The Normal distribution is a common default assumption because of its mathematical tractability, but the fact that it underestimates extreme losses has been widely documented (McNeil et al., 2015).

Another approach is extreme value theory (EVT), which focuses solely on the tail behavior of the distribution. Gumbel's preliminary work on statistics of extremes (Gumbel, 1958) for- malized into two operational approaches. Block-maxima divides the sample into blocks and fits a generalized extreme value distribution to the block maxima, correcting for clustering via an extremal index. Peaks-over-threshold, combined here with a GARCH filter to take out volatility clustering, implements modeling of exceedances over a threshold, with a generalized pareto distribution. Gen¸cay and Sel¸cuk pioneered the application of this test, systematically building a framework using nine emerging stock markets, finding that the VaR based on EVT has become increasingly accurate compared to other methods. As the level of confidence grew, standard approaches were used (Gen¸cay and Sel¸cuk, 2004). Assaf (2009) extended this compar- ison to four MENA markets, including Morocco, reaching a similar conclusion. The asymptotic concepts of BM and POT are presented in (Coles, 2001; Embrechts et al., 1997; McNeil et al., 2015). Comparisons with the two approaches in practice, especially outside the large, devel- oped, markets, where most of this literature originates, are fairly sparse.

The Moroccan all shares ındex (MASI) is a convenient testing market. As the reference index of the casablanca stock exchange, it covers January 2006 to January 2026, with 4,978 daily observations spanning the 2008 crisis, the 2020 shock and the 2022 geopolitical turbulence, overlapping information bases and thinner liquidity, which are common characteristics of EM markets. The returns are asymmetric with greater fat tails than typically, developed indices (Bekaert et al., 2005) documented. Work on the MASI itself has so far favored classical volatility models over tail risk (Hamou et al., 2025), and EVT-based studies applied to African markets specifically remain limited, which is reason enough to look closely at this one.

There are two aims of this paper. It is an estimation of VaR and ES for the MASI based on four models: Historical simulation, normal model, block-maxima and peaks-over-threshold, in a direct comparison. Second, it assesses the efficacy of each of the methods in predicting the results using Kupiec's unconditional coverage test, Christoffersen's independence test, and the joint VaR-ES test of McNeil and Frey.

The remainder of the paper follows. Section 2 establishes the methodological foundation for our empirical study. It opens with an overview of the risk measures, details the four estimation methods applied, and concludes with the backtesting procedures implemented to assess model performance. Section 3 compares the methods directly across confidence levels with the ES/VaR ratios. Section 4 deals with the implications of this for risk management in the context of the emerging markets and concludes.

2. METHODOLOGY

This section presents the methodological framework underlying our empirical analysis. We first introduce the risk measures considered, then describe the four estimation methods employed, and finally present the backtesting procedures used to validate the models.

2.1. Risk Measures: Value-at-Risk and Expected Shortfall

2.1.1. Definitions

Value-at-Risk is defined as a quantile of the return distribution. Let F_X denote the distribution function of the random returns X, and let $\alpha \in (0, 1)$ be a confidence level. The value at risk at level α, denoted VaR_α, is then given by the negative of the α-quantile of F_X:

$$VaR_\alpha = \inf\{x \in \mathrm{R}: F_X(x) \geq 1 - \alpha\} \tag{1}$$

That is, $\mathrm{P}(X \leq -VaR_\alpha) \leq \alpha$. In other words, a loss exceeding VaR_α in absolute value materializes only with probability at most α.

Expected shortfall is defined as the conditional expectation of the return, given that the return is less than or equal to the negative of VaR. Formally:

$$ES_\alpha = \mathrm{E}[X \mid X \leq VaR_\alpha] \tag{2}$$

The ES can be calculated analytically from the VaR:

$$\mathrm{ES}_a = \frac{1}{1-\alpha}\int_0^{1-a} V_a R_z d_z \tag{3}$$

Thus, expected shortfall represents the expected loss conditional on the VaR being exceeded. This study focuses exclusively on the left tail (losses), which is of primary interest for risk management purposes.

2.1.2. Motivation for the methodological choice

The selection of estimation methods is based on the statistical characteristics of financial returns. The asset returns tend to have some stylized facts as shown in the literature, namely volatility clustering, heavy tails and negative skewness (Cont, 2001). Such properties make parametric models with normality assumption unsuitable for modelling extreme risks (McNeil et al., 2015). In this study, 4 approaches are used, each of which is a paradigm:

1. Historical simulation: A non-parametric benchmark relying solely on empirical quantiles.
2. Normal model: A parametric approach based on the Gaussian assumption, included to illustrate the limitations of standard practice
3. Block-maxima (BM): An EVT method modeling the distribution of block maxima via the generalized extreme value (GEV) distribution
4. Peaks-over-threshold (POT): An EVT method modeling exceedances above a high threshold via the generalized pareto distribution (GPD), combined with a GARCH filter to account for volatility clustering.

The rationale for this comparative design is twofold. First, it allows us to assess the magnitude of the bias induced by the normality assumption in an emerging market context. Second, it enables us to evaluate the relative performance of the two main EVT approaches, which differ fundamentally in their use of data: BM uses only one observation per block, while POT uses all observations above a threshold.

2.1.3. Historical approach

Given a return series $X_1,..., X_T$, the historical VaR and ES rely solely on sample quantiles. Let $X(1) \leq ... \leq X(T)$ denote the order statistics. For a given confidence level α:

$$VaR_\alpha = X_{[T(1-\alpha)]}, ES_\alpha = \frac{1}{[T(1-\alpha)]}\sum_{i=1}^{[T(1-\alpha)]} X_{(i)} \tag{4}$$

These estimates are non-parametric in nature and rely exclusively on the empirical distribution.

2.1.4. The normal model

Under the normal assumption, for a given confidence level α, the VaR and ES are given by:

$$VaR_\alpha = \mu + \sigma\Phi^{-1}(1-\alpha), \mathrm{ES}_a = \mu + \frac{\sigma}{\alpha}\phi(\Phi^{-1}(1-\alpha) \tag{5}$$

Where μ and σ denote the mean and standard deviation of returns, and Φ and ϕ denote the distribution and density functions of the standard normal distribution.

2.1.5. Extreme value theory methods

Extreme value theory (EVT) provides a statistical framework for modeling the asymptotic behavior of distribution tails without imposing a complete parametric form on the data (Embrechts et al., 1997). Two complementary approaches are developed: The block-maxima (BM) method and the peaks-over-threshold (POT) method (Coles, 2001).

The Block-Maxima Method Let $X_1,..., X_T$ be a return series partitioned into m blocks of size n ($T = m \cdot n$). For each block i, the block maximum is defined as:

$$M_n^{(i)} = \max\{x_n^{(i)},...,x_n^{(i)}\} \tag{6}$$

The Fisher-Tippett-Gnedenko theorem establishes that the distribution of normalized block maxima converges to the generalized extreme value (GEV) distribution as $n \to \infty$ (Fisher and Tippett, 1928; Gnedenko, 1943):

$$\mathrm{G(z)} = \exp\left(-\left(1+\varepsilon\frac{z-\mu}{\sigma}\right)\right)^{-1/\varepsilon} \quad 1+\varepsilon\frac{z-\mu}{\sigma} > 0 \tag{7}$$

where $\mu \in \mathrm{R}$ is the location parameter, $\sigma > 0$ the scale parameter, and $\xi \in \mathrm{R}$ the shape parameter (Coles, 2001). Parameters are estimated via maximum likelihood (Hosking et al., 1985; Hosking, 1985; Macleod, 1989; Prescott and Walden, 1980, 1983; Smith, 1985).

Note that the block maxima method is classically formulated within the theoretical framework of independent and identically distributed (i.i.d.) variables. However, this assumption is rarely satisfied in practice, particularly in the field of finance, where time series typically exhibit temporal dependence, conditional heteroskedasticity, and heavy-tailed distributions. In this context, the direct application of the block approach requires specific precautions, or even methodological adaptations, such as using blocks of sufficiently large size to mitigate the effect of dependence, or employing declustering techniques to select approximately independent peaks. Thus, while the block maxima method has strong theoretical support for i.i.d. variables, using it in finance requires careful implementation that considers the unique characteristics of financial data. Therefore, let us use the extremal index that addresses this issue.

The extremal ındex ın the presence of temporal dependence, the extremal index $\theta \in [0, 1]$ corrects for the clustering of extremes (McNeil, 1998). Following Embrechts et al. (Embrechts et al., 1997), the estimator is:

$$\hat{\theta} = \frac{1}{n}\frac{\log(1-\frac{K_u}{m})}{\log(1-\frac{N_u}{T})} \tag{8}$$

Where N_u is the number of observations exceeding threshold u, K_u the number of blocks con- taining at least one exceedance, m the number of blocks, T the total number of observations, and n the block size.

The VaR for a confidence level α is derived from the GEV quantiles adjusted by the extremal index:

$$VaR_\alpha = G^{-1}(\alpha^{n\theta}), \tag{9}$$

Where G^{-1} is the GEV quantile function:

$$G^{-1}(p) = \mu - \frac{\sigma}{\xi}(1-(-\log p)^{-\xi}), \xi \neq 0. \tag{10}$$

The goodness-of-fit is assessed using the Sherman test (Sherman, 1957).

The peaks-over-threshold method the POT method models observations exceeding a high threshold u. Here, Z_t denotes the GARCH-filtered standardized residuals used as input to the POT procedure. Let $Y_j = Z_j - u$ denote the excesses. The Pickands-Balkema-de Haan theorem establishes that the conditional distribution of excesses converges to the generalized pareto distribution (GPD) as $u \to \infty$ (Embrechts et al., 1997):

$$\mathrm{G(y)} = \begin{cases} \left\{1-(1+\xi\frac{y}{\beta})\right\}^{-1/\xi}, & \xi \neq 0, \\ \left\{1-\exp+(-\frac{y}{\beta})\right\}, & \xi = 0, \end{cases} \tag{11}$$

With $\xi \in \mathrm{R}$ the shape parameter and $\beta > 0$ the scale parameter.

Threshold selection is important. Two tools are used together: the mean excess plot, which should be linear beyond the threshold (Embrechts et al., 1997), and the Hill estimator (Hill, 1975). The stability of the Hill estimator shows a good choice (McNeil and Frey, 2000).

The VaR and ES for a level α are then:

$$VaR_a = u + \frac{\beta}{\xi}\left(\left(\frac{T(1-\alpha)}{N_u}\right)^{-\xi} - 1\right), \xi \neq 0 \tag{12}$$

$$VaR_a = u - \beta \log\left(\frac{T(1-\alpha)}{N_u}\right), \xi = 0 \tag{13}$$

$$ES_a = \frac{VaR_a}{1-\xi} + \frac{\beta - \xi u}{1-\xi}, \xi = 1 \tag{14}$$

The normalized standardized values are then normalized with the POT method to adjust for the volatility clustering. Residuals from a GARCH (1,1) filter that guarantee the i.i.d. approxi- mation needed by the theory (McNeil and Frey, 2000).

2.2. Backtesting Framework

We perform a comprehensive backtesting framework with a set of complementary tests on the goodness-of-fit of the VaR and ES estimates.

2.2.1. Kupiec Test (Unconditional Coverage)

The Kupiec test (Kupiec, 1995) is used to test if the number of violations is equal to the expected number. The expected violation rate for a level of confidence α is $1-\alpha$. The value of the test statistic is:

$$\mathrm{LR}_{\mathrm{UC}} = -2 \log\left(\frac{\alpha^{T-N}(1-\alpha)^N}{((T-N/T)^{T-N}(N/T)^N}\right) \sim x^2(1). \tag{15}$$

Whereas T represents the total number of observations and N the number of violations.

2.2.2. Christoffersen Test (Independence)

The Christoffersen test (Christoffersen, 1998) evaluates if violations are independent over time. The test is a combination of the unconditional coverage test and the independence test:

$$LR_{CC} = LR_{UC} + LR_{IND} \sim \chi^2(2), \tag{16}$$

Where LR_{IND} is a test for independence between violations, based on a Markov chain model.

2.2.3. McNeil-Frey test (joint VaR-ES evaluation)

The McNeil-Frey test (McNeil and Frey, 2000) is based on the standardized residuals that exceed the VaR threshold jointly evaluating VaR and ES:

$$\mathrm{Zt} = \frac{x_t - VaR_t}{ES_t} 1_{\{x_t < VaR_t\}}. \tag{17}$$

Under correct model specification, the conditional distribution of Z_t should be approximately standard exponential.

2.3. Data and Preliminary Analysis

2.3.1. Data description

We will be examining the MASI price from 4 January 2006 until 2 January 2026. The time series is presented in Figure 1. The price information is from the Casablanca Stock Exchange. This paper seeks to estimate two of the most important risk measures, VaR and ES, at different time horizons, based on a range of extreme value theory (EVT) methods. The aim is twofold: to see how the estimates differ over time and to compare how well each method works. For those looking for the formal definitions and the theoretical foundation of EVT, Section 2.1 covers all of these.

To get started, the price data are converted into log returns. This return series is the raw material for the entire analysis, and it is also plotted in Figure 2 for reference. What is really of interest is, on the one hand, the statistical behavior of VaR and ES estimates, and on the other hand, how the choice of EVT method shapes the dependence structures between risks. That is why we use daily returns in our empirical work, which we think is a good compromise between granularity and reliability.

In this study, a long-term approach is followed, and the daily data available in the public domain is used. We are interested in estimating VaR and ES for rare events that happen less often, i.e. events of interest in long time horizons. Specifically, we compute VaR and ES at the confidence levels α = 0.95, 0.975, 0.99, and 0.999.

The rare events are normally interpreted as the left tail of the daily return distribution, or as extreme daily losses. It is this focus that has resulted from the prominently positioned use of VaR and ES in applications of risk management.

Figure 1: Evolution of Moroccan all shares index Price between 2006 and 2026

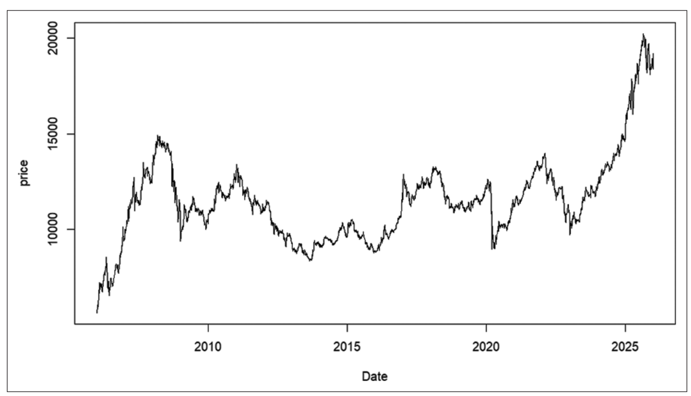
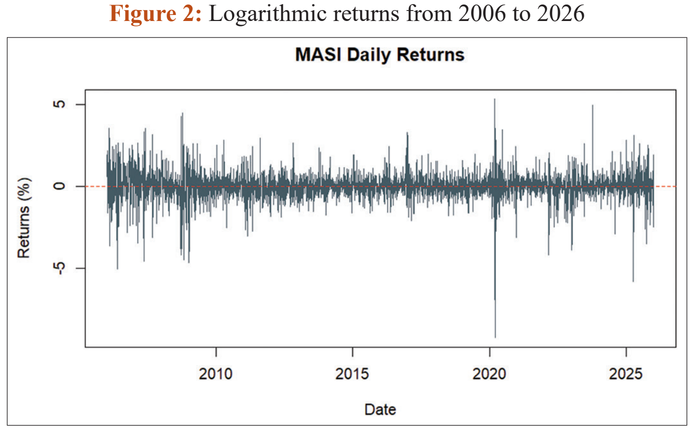


Figure 2: Logarithmic returns from 2006 to 2026



The estimation of VaR and ES is based on two widely used methods from extreme value theory (EVT): The block-maxima method with an extremal index and the peaks-over-threshold (POT) method along with a GARCH filter.

In order to provide a reference for these extreme value estimates, we use the in-sample historical VaR and ES, calculated directly from the daily return series.

2.3.2. Descriptive statistics

The data underlying the data set is 4,978 daily returns between 4 January 2006 and 2 January 2026. Table 1 displays the summary statistics for the return series.

The descriptive statistics of the MASI's daily returns show typical traits of financial assets, including an average daily return of near zero (0.023%) and an average daily volatility of 0.79%. The negative skewness of −0.75 suggests that the losses are heavier than gains, consistent with the extreme loss value of −9.23% and the extreme gain value of +5.31% observed. Above all, the kurtosis of 10.80 —which is much higher than that of a normal distribution (3) —confirms the presence of fat tails and a leptokurtic distribution. These results indicate that parametric models based on the normality assumption are not applicable in the management of extreme risks. Unlike parametric approaches, extreme value theory (EVT) offers a sound theoretical basis for modelling the tail behaviour of a distribution without making any assumptions on the tail shape.

Two basic properties of the time series are confirmed by the stationarity and normality tests performed on the daily returns of the MASI. Table 2 presents the results of these tests.

The ADF test (−15.8687; p-value = 0.01) rejects the unit root hypothesis, while the KPSS test (0.2383; p-value = 0.10) does not reject the stationarity hypothesis, which is very robust and makes it clear that the series is stationary (necessary for any econometric modelling). The Jarque-Bera test (24, 662.7273, p-value = 0.00) on the other hand strongly rejects normality, which further points to the presence of fat tails and substantial asymmetry in the distribution of returns that was also indicated by the descriptive statistics. The overall results confirm that extreme value theory methods not based on the normality assumption are suitable for estimating VaR and ES; and that a GARCH filter is appropriate for capturing the persistence of volatility shown by the ACF of the squares of the returns.

The exploratory analysis of the daily return of the MASI (Figure 3) reveals a set of basic features that make the use of extreme value theory appropriate. Over time, the time series shows volatility clustering, where the period of high volatility is followed by a period of low volatility. The empirical distribution is leptokurtic and negatively skewed, with greater tail thickness and peak, as shown by the QQ-plot, which has points that are significantly off the line at the extremes indicating greater leptokurticity and negative skewness, reflecting the fact that severe losses happen more often than extraordinary gains. Last, the squared returns are strongly persistent: the autocorrelation decays slowly and is significant for a number of lags. Overall, these results validate the hypothesis of fat tails, demonstrate the presence of temporal dependence in the shocks and fully support the application of the extreme value theory and a GARCH filter to provide reliable estimates of VaR and ES.

2.4. Estimation Results

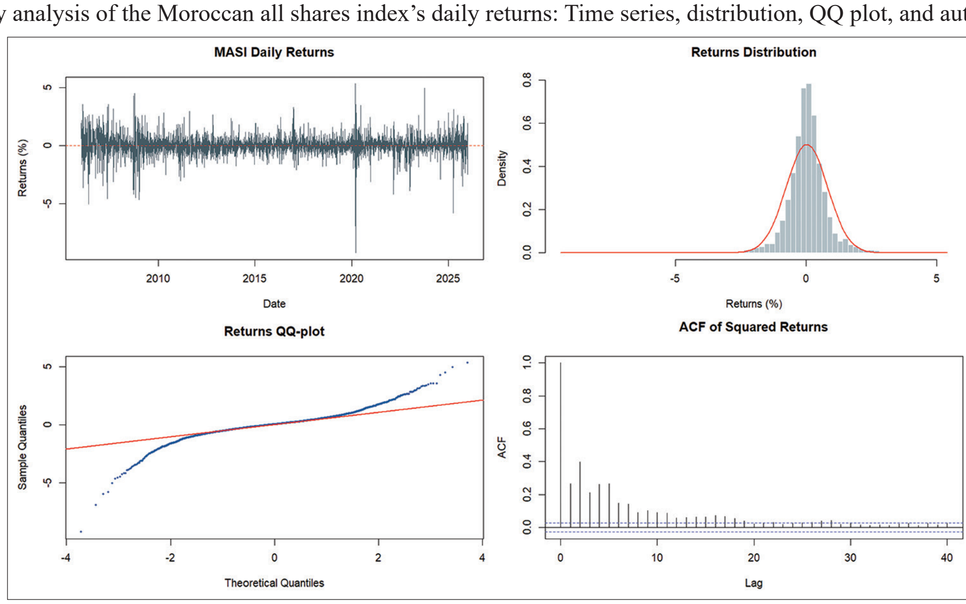
2.4.1. Block-maxima method: VaR and ES

To implement the block maxima approach, four main steps are taken, which are done in a sequence. First, the optimal block size is chosen and then the extremal index is estimated. Then,

Table 1: Descriptive statistics of MASI returns

Mean (%)	Median (%)	Standard deviation (%)	Min (%)	Max (%)	Skewness	Kurtosis	Mean (%)
0.02	0.03	0.79	−9.23	5.31	−0.75	10.8	0.02

Figure 3: Exploratory analysis of the Moroccan all shares index's daily returns: Time series, distribution, QQ plot, and autocorrelation of squares



the GEV distribution is fitted to the block maxima and a number of diagnostic checks are performed to make sure the model fits well. Finally, the VaR and ES estimates are computed. Finally, the estimates are run through a backtesting validation. Each of these stages is described more thoroughly in the subsections that follow.

Step 1: The block-maxima approach is based on a trade-off between the block size and the quality of the asymptotic approximation. On one hand, extreme value theory requires very long blocks for the distribution of maxima to converge to the generalized extreme value (GEV) distribution. In contrast, too large blocks reduce the number of observations for parameter estimation, making the risk measures less precise. To make this choice, the Sherman goodness-of-fit statistic (Sherman, 1957) is considered, which calculates the quality of the GEV fit to the block maxima. Table 3 and Figure 4 present the evolution of this statistic for block sizes ranging from 10 to 100 days.

The analysis reveals an expected pattern: for short blocks (10-20 days), the number of blocks is large but the maxima remain moderate and the GEV fit is less satisfactory (lower Sherman statistic). As block size increases (30-60 days), the maxima become more extreme and the fit improves. For very large blocks (70-100 days), the fit quality stabilizes, but the number of blocks decreases sharply, reducing the statistical power of the estimators.

Figure 4: Selecting the optimal block size using Sherman's statistic

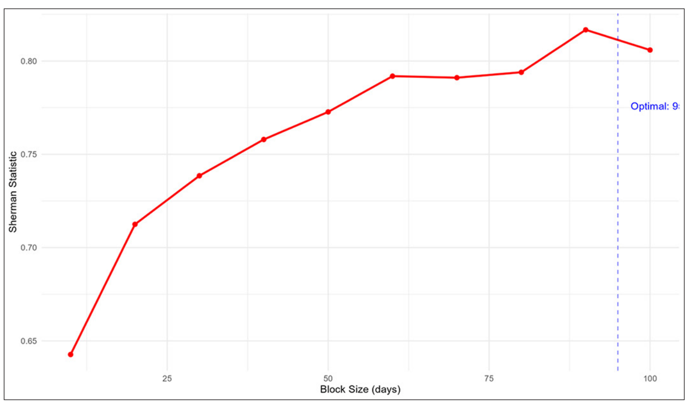


Table 2: Stationarity and normality tests

Test	Statistic	P-value
ADF (stationarity)	−15.8687	0.01
KPSS (stationarity)	0.2383	0.10
Jarque-Bera (normality)	24662.7273	0.00

Table 3: Block-maxima statistics for different block sizes

Block size	Sherman statistic	N blocks	Mean max	Var max
10	0.6427	497	1.0125	0.5699
20	0.7125	248	1.3064	0.6887
30	0.7385	165	1.5140	0.8116
40	0.7579	124	1.6571	0.8770
50	0.7727	99	1.7621	0.9134
60	0.7919	82	1.8866	0.9355
70	0.7911	71	1.9324	0.9864
80	0.7940	62	2.0330	1.0726
90	0.8167	55	2.1316	1.0197
100	0.8059	49	2.1472	1.1106

Table 4 reports the detailed Sherman test results for the selected block size of 90 days. The standardized statistic of −0.0093 lies well below the critical value of 1.6449, leading to non- rejection of the null hypothesis of a good GEV fit.

Consequently, we select a block size of 90 days (approximately 3 months of trading days), which offers an optimal compromise: 55 blocks are available for estimation, sufficient to ensure parameter robustness, while guaranteeing a good GEV fit (Sherman statistic ≈ 0.817).

The extremal index θ is estimated following the procedure recommended by (Bloss et al., 2025):

we compute $\hat{\theta}$ for thresholds ranging from the 0.2% to the 3% sample quantiles, then average the individual estimates. The obtained value is θ = 0.4617, which is <1, confirming the presence of temporal clustering of extremes in the MASI return series.

This moderate dependence justifies the use of the extremal index in the block-maxima approach to correct VaR and ES estimates.

Step 2: GEV distribution fitting and diagnostics the GEV distribution is fitted to the block maxima (90 days) using maximum likelihood estimation. The estimated parameters are the location parameter μ = 1.5658, the scale parameter σ = 0.5969 and the shape parameter ξ = 0.4833.

The positive shape parameter indicates a heavy-tailed distribution, characteristic of financial assets exposed to sudden shocks. Figure 5 presents the goodness-of-fit diagnostics for the GEV distribution. The block maxima plot shows a concentration of extreme values around certain periods of high volatility, reflecting the temporal heterogeneity of extreme risks. The QQ- plots indicate excellent agreement between theoretical GEV quantiles and empirical quantiles, including in the tails of the distribution, which ensures reliable VaR and ES estimates at high confidence levels (0.95-0.999).

Step 3: VaR and ES estimation Table 5 reports the VaR and ES estimates for losses (negative tail) at the 95%, 97.5%, 99%, and 99.9% confidence levels. These measures incorporate the extremal index θ = 0.4617 in the GEV quantile formula. As expected, both risk measures increase substantially with the confidence level. VaR rises from −1.19% at 95% to −6.07% at 99.9%, while ES increases from −1.99% to −11.45% over the same range. The ES-to-VaR ratio increases from 1.67 at 95% to 1.88 at 99.9%, illustrating the leverage effect of heavy tails: The more extreme the threshold, the more the average conditional loss exceeds the quantile. The particularly sharp jump at the 99.9% level (VaR: −6.07%, ES: −11.45%) highlights the model's sensitivity to extreme tail

Table 4: Sherman goodness-of-fit test for GEV distribution (block size=90 days)

Statistic	Value
Statistic Ω_m	0.006736
Expected value $E[\Omega_m]$	0.017538
Standardized statistic	−0.0093
Critical value (95%)	1.6449
Hypothesis rejection	NO

events. These results provide a conservative view of market risk, consistent with the heavy-tailed nature of financial returns and the clustering of extremes captured by the extremal index.

Step 4: Backtesting validation Table 6 presents the results of the Kupiec (unconditional coverage) and Christoffersen (independence of violations) tests, applied across confidence levels. From a prudential risk management perspective, attention naturally focuses on the highest confidence levels (99% and 99.9%), as they reflect potential losses during severe financial crises. At the 99% level, VaR is −2.21% and ES is −3.98% (ES/VaR ratio = 1.80).

The Kupiec test does not reject the adequacy of the violation count (P = 0.31), but the Christoffersen test rejects independence (P = 0.000), indicating clustering of exceedances over time. At the 99.9% level, the values are more extreme: VaR = −6.07%, ES = −11.45% (ratio = 1.88). The Kupiec test does not reject the hypothesis (P = 0.13) and the Christoffersen test, this time, does not reject independence either (P = 0.968). This result is particularly favorable, as it validates the model for the most demanding confidence level. These results confirm that the block-maxima method with an extremal index is particularly well-suited for estimating extreme risks in a stock market context, especially at the 99.9% level, where the model is validated by all tests. It therefore constitutes a reliable tool for managing market risks during periods of extreme stress.

2.4.2. Peaks-Over-Threshold Method: VaR and ES

For the peaks-over-threshold (POT) method, we first fit a GARCH (1,1) model with a skewed Student-*t* distribution to the logarithmic

Figure 5: Diagnostic plots for generalized extreme value fit to block maxima (block size = 90 days)

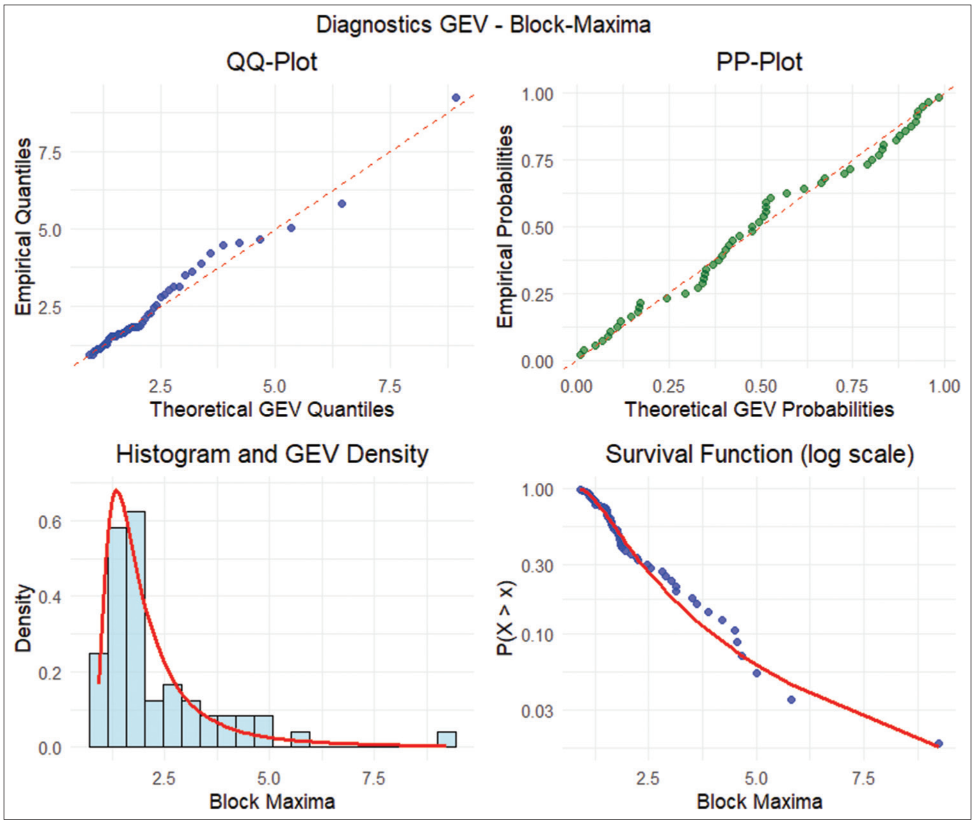


Table 5: Block-maxima VaR and ES estimates at different confidence levels

Level (%)	VaR	ES	Shape	Location	Scale	P	n blocks
95	−1.1875	−1.9888	0.4833	1.5658	0.5969	0.1187	55
97.5	−1.5359	−2.6631	0.4833	1.5658	0.5969	0.3492	55
99	−2.2142	−3.9760	0.4833	1.5658	0.5969	0.6586	55
99.9	−6.0748	−11.4483	0.4833	1.5658	0.5969	0.9593	55

Table 6: Backtesting results: Kupiec and Christoffersen tests

Level (%)	VaR	ES	ES/VaR ratio	Shape GEV	Kupiec violations	Kupiec rate	Kupiec p	Kupiec decision	Christoffersen p	Christoffersen decision
95	−1.1875	−1.9888	1.6748	0.4833	209	4.198	0.0077	Reject	0.000	Reject
97.5	−1.5359	−2.6631	1.7340	0.4833	137	2.752	0.2621	Not reject	0.000	Reject
99	−2.2142	−3.9760	1.7957	0.4833	57	1.145	0.3148	Not reject	0.000	Reject
99.9	−6.0748	−11.4483	1.8846	0.4833	2	0.040	0.1285	Not reject	0.968	Not reject

return series (Table 7). From the resulting GARCH- filtered residuals, we determine a suitable threshold using both the mean excess function and the Hill estimator. The generalized pareto distribution (GPD) is then fitted via maximum likelihood. A sensitivity analysis, not reported in detail here, confirms that neither the specific choice of GARCH specification nor the selected threshold materially affects the main results.

Figure 6 displays the diagnostic plots used for threshold selection in the peaks-over-threshold (POT) method, applied here to the GARCH-filtered standardized residuals rather than to raw returns. The left panel, the mean excess plot, shows the conditional expected excess as a function of the threshold u. For a well-fitted generalized pareto distribution (GPD), this function should be approximately linear beyond a certain threshold.The mean excess plot shows the curve becoming roughly linear starting around the 1.5-2 range, a good sign that a threshold in this neighborhood satisfies the conditions of the Pickands-Balkema-de Haan theorem. Turning to the right panel, the Hill plot traces the shape parameter ξ against the number of exceedances. We see the curve stabilizing nicely over a range of about 150-300 exceedances, suggesting that the shape parameter is fairly robust in this region—which further supports a threshold in this same range.

Taken together, these diagnostics point to a threshold of roughly $u = 2$ as a solid compromise. It's high enough to make sure the asymptotic GPD approximation holds, but low enough to retain a decent number of exceedances—around 200 observations—so that we can estimate our parameters reliably and compute meaningful VaR and ES figures.

According to Table 8 and Figure 7, the diagnostic analysis of the standardized residuals from the GARCH(1, 1) model with skewed Student-t innovations shows that the filter is able to capture the volatility dynamics of the return series. The time series plot of the standardized residuals (left panel) does not show any clustering of volatility fluctuations, and fluctuations are within a stable range around zero, which means that the GARCH model has successfully removed the conditional heteroskedasticity. The density plot (right panel) shows the empirical distribution of the residuals (histogram) with a standard normal density (solid line). It is seen that the empirical distribution has a higher peak and slightly heavier tails than normal, implying some remaining excess kurtosis, although the skewed Student-t specification is already responsible for most of this tail thickness.

Table 7: Summary of GARCH parameters

α_1	β_1	ϕ	v	Skew
0.2549	0.6849	0.9398	4.3678	0.9692

Table 8: Diagnostic test results on residuals

Test	Statistic	P-value
Ljung-box (residuals)	197.1727	0.0000
Ljung-box (squares)	21.3320	0.3778
ARCH-LM	13.2292	0.2111

The formal diagnostic tests in Table 8 can be interpreted to support these visual findings. The Ljung-Box test on residuals gives a statistic of 197.17 with a P = 0.000, which means that the residuals are serially correlated. But this result should be taken care of as the test is sensitive to the large sample size (nearly 5,000 observations). More importantly, the Ljung-Box test on squared residuals gives a statistic of 21.33 with a P = 0.3778 and the ARCH-LM test yields a statistic of 13.23 with a P = 0.2111. Both tests clearly fail to reject the null hypothesis of no remaining ARCH effects, and so the GARCH(1,1) model successfully eliminates volatility clustering. In conclusion, these diagnostic results have shown that the GARCH filter is the right preprocessing step for the Peaks-over-Threshold method as long as the residuals are approximately uncorrelated and homoskedastic and satisfies the i.i.d. assumption for extreme value analysis.

According to Table 9 and Figure 8, the diagnostic plots and parameter estimates confirm the validity of the Peaks-over-threshold (POT) model applied to the GARCH-filtered residuals. The density plot (left panel) shows the empirical distribution of the residuals, with a pronounced peak and moderately heavy tails, consistent with the skewed Student-t specification used in the GARCH filter. The return level plot (right panel) displays the estimated return levels as a function of the return period (in years), with the curve increasing sharply for longer return periods—for example, the 10-year return level reaches approximately 5.5, while the 100-year return level reaches approximately 8.0, reflecting the extreme tail behavior of the residuals.

The parameter estimates reported in the table are useful to support this model. The threshold is set to u = 1.5405 with N_u = 249 exceedances which is exactly 5% of the sample and is sufficient

Figure 6: Diagnostic plots for threshold selection in the peaks over threshold method: Mean excess plot (left) and hill plot (right)

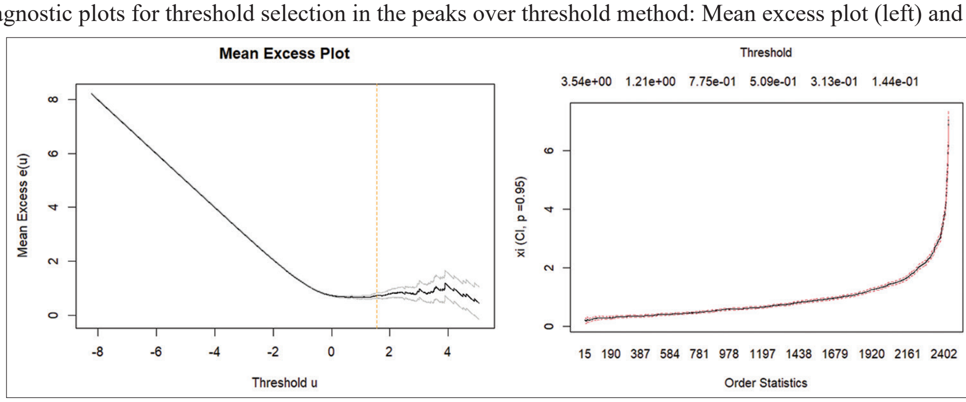


Figure 7: Standardized residuals from the GARCH model and their empirical distribution

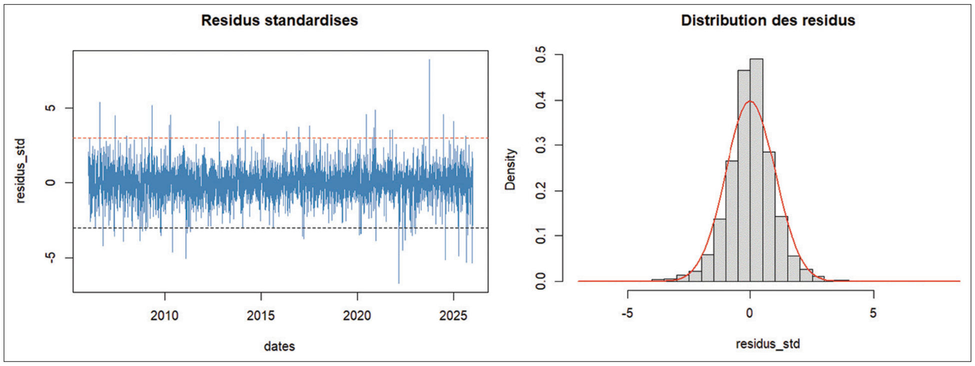


Figure 8: Density plot of GARCH-filtered residuals and corresponding return level plot

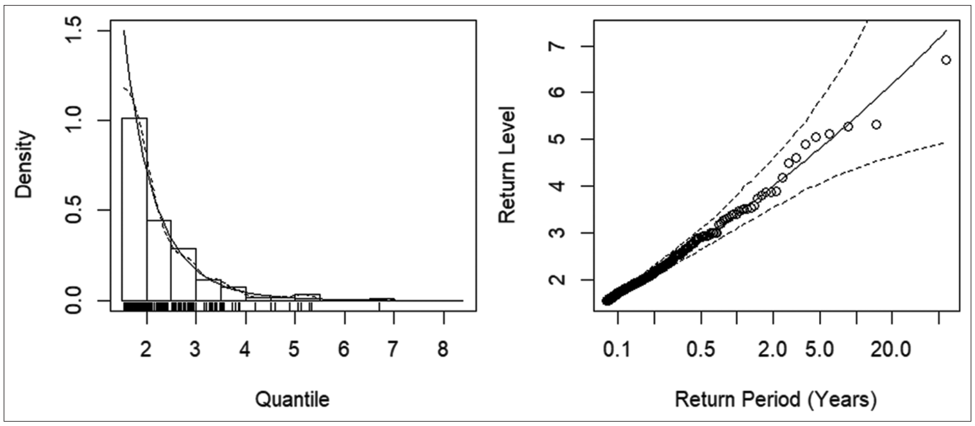


Table 9: Estimated parameters and confidence intervals for the extreme value model

Threshold u	N_u	Proportion (%)	ξ	ξ CI lower	ξ CI upper	σ	σ CI lower	σ CI upper
1.5405	249	5.00	0.0803	−0.0423	0.2366	0.6685	0.5535	0.8007

for GPD estimation. The shape parameter $\xi = 0.0803$ is positive, but near zero with a 95% confidence interval from −0.0423 to 0.2366, indicating that the tail is rather light and is consistent with the Gumbel domain of attraction (exponential-like tails). The scale parameter $\sigma = 0.6685$ has a confidence range of (0.5535, 0.8007) and shows that the exceedances are scattered. The fact that the confidence interval for ξ includes zero implies that the tails are not heavy-tailed, which is very reassuring for risk management and data analysis. These results suggest that the POT model with GARCH filtering is a stable and well-calibrated model for estimating extreme quantiles and return levels and that both parameter estimates and extreme projections are realistic.

Table 10 reports the VaR and ES estimates for the GARCH-filtered residuals at four confidence levels, obtained using the Peaks-over-Threshold (POT) method with a threshold of $u = 1.5405$ and 249 exceedances.

As expected, both measures increase in magnitude with the confidence level: VaR falls from −1.54% at 95% to −4.61% at 99.9%, while ES falls from −2.27% to −5.61% over the same range. The ES estimates are consistently larger in magnitude than the corresponding VaR values, with the ES-to-VaR ratio ranging from 1.47 at 95% to 1.22 at 99.9%, reflecting that ES captures the

Table 10: Value-at-risk (VaR) and expected shortfall (ES) estimates using the peaks-over- threshold method

Level (%)	VaR_POT	ES_POT
95	−1.5408	−2.2677
97.5	−2.0173	−2.7858
99	−2.6894	−3.5166
99.9	−4.6136	−5.6089

average severity of losses beyond the VaR threshold. The relatively moderate magnitudes of these estimates are consistent with the low positive shape parameter $\xi = 0.0803$ estimated for the GPD, indicating that the tail of the residual distribution is relatively light and that extreme losses are less pronounced than they would be under a heavy-tailed distribution. These results provide a robust baseline for risk management, particularly for moderate confidence levels (95-99%), while the 99.9% level offers a conservative estimate for extreme stress scenarios. Backtesting results for the POT VaR and ES estimates are summarised in Table 11 and provide a good view of the POT model's performance in 3 different complementary tests. The Kupiec test (Panel a) evaluates the unconditional coverage by comparing the observed violation rate with the expected one. For all confidence levels (95%, 97.5%, 99% and 99.9% in the analysis), the number of violations is very close to the expected value and P-values are well above the usual significance level (0.657-0.995), with the decision of Non reject holding throughout.

The Christoffersen test for the independence of violations over time is different. For the 95%, 97.5% and 99% levels, the test rejects the null hypothesis of independent violations (P = 0.0001, 0.0096 and 0.0182), indicating that exceedances tend to cluster in certain periods of time, which is common in financial returns. At the 99.9% level, the test does not reject independence (P = 0.9042) so it is clear that the most extreme violations occur randomly.

The McNeil-Frey test (that jointly assesses VaR and ES) is reported for the 95% and 99% levels only. All P-values over 0.10 clearly show that the model can estimate both risk measures accurately.

Overall, these results indicate that the POT model with GARCH filtering performs satisfactorily for extreme confidence levels (99% and 99.9%), where it captures both the frequency and the clustering of extreme losses, while for moderate levels, the violation clustering suggests that dynamic risk updates would be beneficial in practice.

3. RESULTS

This section presents and analyzes the VaR and ES estimates obtained from the four methods at confidence levels of 95%, 97.5%, 99%, and 99.9%. We first provide a comparative overview of the estimates, then examine the ES/VaR ratios, and finally discuss the implications for risk management in the Moroccan market context.

3.1. Comparative Analysis of VaR and ES Estimates

Table 12 reports the VaR and ES estimates across the four methods, along with the shape parameters for the EVT-based approaches.

3.1.1. Key findings

The comparative analysis reveals substantial differences in risk assessment across the four methods, particularly at extreme confidence levels.

3.1.1.1. The normal model: Systematic underestimation

The Normal method is known to give the most optimistic (smallest in magnitude) risk estimates at all levels. At the 99.9% confidence level, the Normal model is showing only VaR of −2.43% and

Table 11: Backtesting results: Kupiec test, Christoffersen test and McNeil-Frey test

a. Kupiec test results					
Level (%)	**Violations**	**Rate (%)**	**LR**	**P-value**	**Decision**
95	249	5.002	0.0000	0.9948	Non reject
97.5	119	2.391	0.2483	0.6182	Non reject
99	52	1.045	0.0986	0.7536	Non reject
99.9	6	0.121	0.1970	0.6572	Non reject

b. Christoffersen independence test			
Level (%)	**LR**	**P-value**	**Decision**
95	14.5577	0.0001	Reject
97.5	6.7010	0.0096	Reject
99	5.5785	0.0182	Reject
99.9	0.0145	0.9042	Non reject

c. McNeil-Frey test				
Level (%)	**P_simple bilateral**	**P_simple unilateral**	**P_standardized bilateral**	**P_standardized unilateral**
95	0.596	0.717	0.877	0.464
99	0.393	0.215	0.227	0.109

Table 12: Comparative backtesting results: VaR, ES and Shape parameters across methods

a. Value-at-Risk estimates (%)				
Level (%)	**VaR_Hist**	**VaR_Normal**	**VaR_BM**	**VaR_POT**
95	−1.0922	−1.2801	−1.1875	−1.5408
97.5	−1.5824	−1.5301	−1.5359	−2.0173
99	−2.3308	−1.8208	−2.2142	−2.6894
99.9	−5.0167	−2.4268	−6.0748	−4.6136

b. Expected shortfall estimates (%)				
Level (%)	**ES_Hist**	**ES_Normal**	**ES_BM**	**ES_POT**
95	−1.9136	−1.6116	−1.9888	−2.2677
97.5	−2.5242	−1.8299	−2.6631	−2.7858
99	−3.5033	−2.0897	−3.9760	−3.5166
99.9	−6.5969	−2.6465	−11.4483	−5.6089

c. Shape parameters (GEV/GPD)		
Level (%)	**Shape BM**	**Shape POT**
95	0.4833	0.0803
97.5	0.4833	0.0803
99	0.4833	0.0803
99.9	0.4833	0.0803

ES of −2.65%, which is the result of the poor tail thickness measurements. This is not surprising in light of the well-known limitations of the Gaussian model in financial applications (Cont, 2001). The MASI's return distribution has a kurtosis of 10.80, which is nearly four times that of a normal distribution. This is why the Gaussian model is not suitable for this market. As mentioned in (McNeil et al., 2015), the use of the Normal model for tail risk estimation may lead to dangerously optimistic capital requirements at the moment when it is most critical to be able to measure the tail thickness properly.

3.1.1.2. The historical method: Robust but sample-limited

The historical method provides intermediate estimates (VaR 99.9%: −5.02% and ES: −6.60%) while showing only empirical quantiles and not parametric extrapolation. This is robust in the sense that it does not make any distributional assumptions but is constrained by the sample size (Embrechts et al., 1997). With 4,978 daily returns, the 99.9% VaR is given by fewer than 5 observations. This is an estimate that is very sensitive to individual extreme events and is not well adapted to extrapolation outside of the observed sample range.

3.1.1.3. The block-maxima method: Maximum conservatism

The Block-Maxima (BM) method gives the most conservative estimates, particularly at extreme values (VaR 99.9%: −6.07% and ES: −11.45%). This is due to its high shape parameter $\xi = 0.4833$, which indicates a heavy tail and a slow decay in the extreme probability. The BM method also takes extreme clustering into account (by its extremal index $\theta = 0.4617$), which also makes it more conservative as it is more likely that extreme events will occur in clusters and can increase the effective risk over any given horizon.

3.1.1.4. The POT-GARCH method: Balanced and dynamic

The POT method applied to GARCH-filtered residuals leads to very moderate estimates (VaR 99.9%: −4.61%, ES: −5.61%). The low shape parameter ($\xi = 0.0803$) suggests rather light tails, which is likely as the GARCH filter is effective at removing volatility clustering. It is a well- balanced result between the extremely optimistic Normal model and the very conservative BM method, which is most suitable for day-to-day risk monitoring problems when accuracy and responsiveness are all important.

3.2. ES/VaR Ratio Analysis

Table 13 presents the ES/VaR ratios across the four methods and confidence levels.

The ES/VaR ratios provide valuable insight into the tail behavior captured by each method. The Normal method has the lowest and most stable ratio (1.09-1.26) because the Gaussian distribution is thin-tailed. In the normality case the expected loss beyond VaR is only slightly larger than the VaR itself. This is not consistent with the heavy tails of financial returns (Embrechts et al., 1997).

Table 13: Ratio ES/VaR across methods (Historical, Normal, BM, POT)

Level (%)	Historical	Normal	BM	POT
95	1.7520	1.2590	1.6748	1.4718
97.5	1.5952	1.1959	1.7340	1.3810
99	1.5031	1.1477	1.7957	1.3076
99.9	1.3150	1.0905	1.8846	1.2157

The Historical method shows decreasing ratios from 1.75 at 95% to 1.32 at 99.9%, since the finite sample size makes extreme observations more sparse.

The BM method has the highest and increasing ratios (1.67 at 95-1.88 at 99.9%), which indicates that the average loss beyond the VaR threshold becomes disproportionately large at extreme levels. This is a feature of heavy-tailed distributions (Coles, 2001) and hence the BM method is well suited for stress testing scenarios.

The POT method falls in between with decreasing ratios (from 1.47 to 1.22), indicating that the tail is lighter after GARCH filtering. The decreasing pattern suggests that while extreme losses are captured, their severity in relation to VaR drops at very high confidence levels.

These ratios are consistent with the theory of EVT: for heavy-tailed distributions (positive shape parameter) the ratio of ES/VaR increases with confidence level, whereas for distributions in the Gumbel domain of attraction (shape parameter close to zero) the ratio tends to decrease or remain stable (McNeil et al., 2015).

3.3. Visual Comparison

Figure 9 provides a visual comparison of VaR and ES estimates across the four methods, along with the evolution of the ES/VaR ratio.

The figure confirms the key findings discussed above. While all methods produce increasing risk estimates as the confidence level rises, the divergence between them becomes substantial at extreme levels. The Normal method gives the lowest estimates and is well-known for its poor performance in measuring tail thickness. The Historical method gives different estimates and is based on real-world observations not on parametric extrapolation. The BM method gives the most conservative estimates (VaR 99.9 = 6.1%, ES 99.9 = 11.4%) due to its large shape parameter ($\xi = 0.48$), implying very heavy tails. The POT method applied to GARCH-filtered residuals gives more moderate estimates ($\xi = 0.08$, VaR 99.9 = 4.6%, ES 99.9 = 5.6%), reflecting the lighter tail in the final analysis.

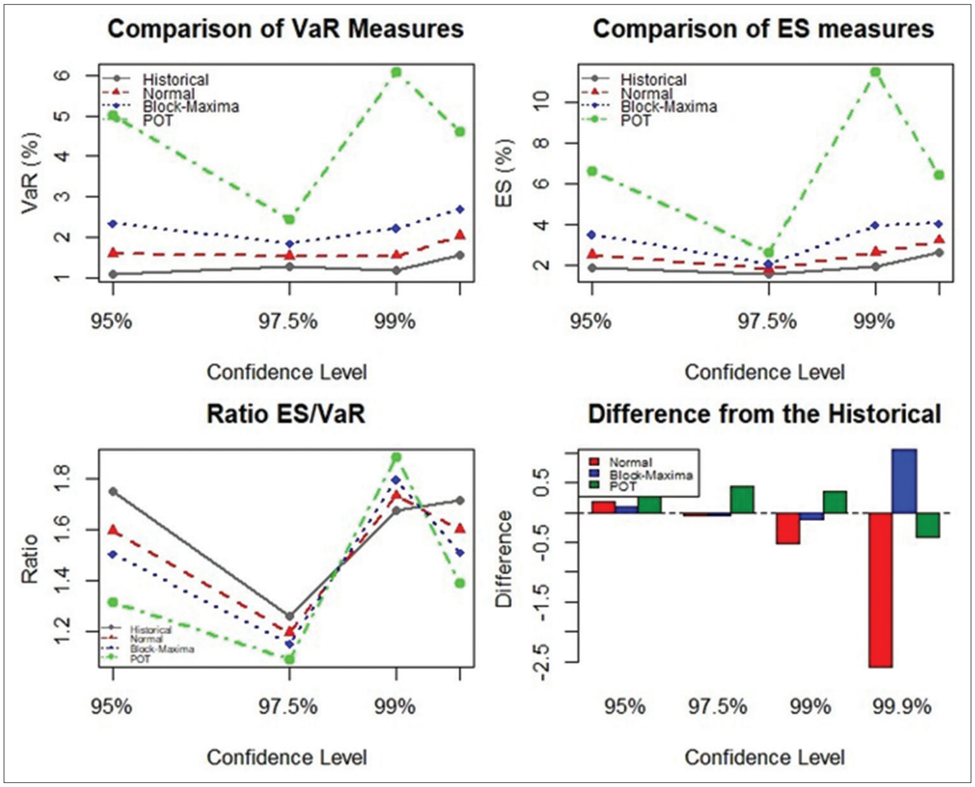
3.4. Implications for Risk Management

The results have clear implications for risk management practice in the Moroccan market context:

3.4.1. Regulatory capital and stress testing

BM is the best approach to regulatory capital calculations and stress testing (Longin, 2000). Its conservative estimates, verified by backtesting at the 99.9% level, offer a suitable buffer against extreme market events. The method's explicit accounting for extreme clustering via the extremal index ($\theta = 0.4617$) is especially important in the Moroccan context, where market liquidity limits may lead to stronger effects on extreme events.

Figure 9: Comparison of value-at-risk (VaR) and expected shortfall (ES) estimates across the four methods (Historical, Normal, Block-Maxima, peaks over threshold) and evolution of the ES/VaR ratio



3.4.2. Daily risk monitoring

The POT-GARCH method offers a more dynamic and responsive framework for day-to-day risk monitoring. Its moderate estimates, combined with strong backtesting performance (Kupiec test never rejecting adequacy across any confidence level), make it suitable for daily VaR and ES calculations where both accuracy and responsiveness are valued.

3.4.3. Model selection guidelines

Based on these findings, the following practical guidelines are proposed for risk management practitioners. For worst-case regulatory scenarios and stress testing purposes, the BM method is recommended as it provides a safety margin that is crucial for capital adequacy. For daily risk monitoring, the POT-GARCH method is preferred for its dynamic nature and excellent backtesting performance that make it reliable for operational risk management. The Historical method is a good non-parametric approach to risk assessment without distributional assumptions. The Normal method should not be used for tail risk assessment as it systematically underestimates extreme risks, especially at high confidence levels.

3.5. Comparison with Existing Literature

Our findings are broadly consistent with the existing EVT literature on emerging markets. The shape parameter estimated for the BM method (ξ = 0.4833) is comparable to those reported for other emerging market indices. For instance, (Longin, 1996) found shape parameters between between 0.3 and 0.5 for developed markets, while studies on emerging markets have reported values up to 0.6, reflecting the higher tail risk associated with these markets (Bekaert et al., 2005). The lower shape parameter obtained with the POT-GARCH method (ξ = 0.0803) is consistent with the findings of (McNeil and Frey, 2000), who showed that GARCH filtering tends to reduce the estimated tail thickness.

The ES/VaR ratios observed for the BM method (1.67-1.88) are also consistent with the theoretical predictions of EVT for heavy-tailed distributions. (Embrechts et al., 1997) show that the ES/VaR ratio for a GPD with shape parameter ξ is given by $1/(1 - \xi)$, which for ξ = 0.48 yields a ratio of approximately 1.92, close to our observed maximum of 1.88.

3.6. Summary of Key Results

In summary, the Normal model systematically underestimates tail risk, with VaR99.9% of only −2.43% compared to −6.07% for the BM method. The BM method provides the most conservative estimates, and is confirmed in backtesting at the 99.9% level, while the POT-GARCH method offers a more balanced approach with good backtesting results, which makes it suitable for daily risk monitoring. The ES/VaR ratios show that the MASI returns are very heavy- tailed; the BM method shows increasing ratios up to 1.88 at 99.9%. The decision between BM and POT then comes down to the risk management objective: BM for worst-case regulatory scenarios, and POT for daily volatility-adjusted risk monitoring.

4. DISCUSSION AND CONCLUSION

The results show that the four methods discussed here have different results and tell different stories of how risk settles in the tails of the MASI return distribution. The gap between these stories increases dramatically if the confidence level is increased.

At 95%, the estimates are similar enough that the choice of method is superfluous: VaR ranges from −1.09% (Historical) to −1.54% (POT). But at 99.9%, the situation drastically changes. The Normal model still has −2.43%, while Block-Maxima has −6.07%. This is not a small difference that can be overlooked. It is directly related to how each method treats the extreme tail behavior, and is just as important when considering solvency in the real crisis.

The Normal model must be taken into account. MASI daily returns have kurtosis of 10.80, almost four times that of a Gaussian distribution. It also has a negative skewness of −0.75, which means big losses occur more often and are more severe than a symmetric bell curve would suggest. When a model that assumes normality is used, tail risk is underestimated. For instance, at the 99.9% level, the Normal model estimates the expected shortfall (ES) at −2.65%, while the Block-Maxima calculates it at −11.45%, which is more than four times higher. In this case, to rely on the Normal model alone to get capital buffers or margin calls, institutions are left with a lack of preparedness in the very scenario they need to be prepared for.

The Historical method does better, as it does not make any distributional assumptions and works from the empirical quantiles of the sample. Its 99.9% estimates (−5.02% for VaR, −6.60% for ES) are much closer to what is found with the EVT-based methods at least for moderate confidence levels. Its weakness is clearly at the very end of the tail: with 4,978 daily returns, the 99.9% quantile can be pinned down by fewer than five observations. At that point, one or two extreme days in the sample can swing the estimate substantially, which makes the method a poor candidate for extrapolating beyond what has actually been observed.

The two EVT methods were created specifically for this problem and are applied in different ways with different results. Block-Maxima, applied to the raw return series with 90-day blocks and accounting for temporal clustering using an extremal index $\theta = 0.4617$, produces the most conservative Expected Shortfall estimates from the 99% confidence level onward, and the most conservative VaR estimates at the most extreme confidence level (99.9%), where its heavy-tailed shape parameter drives a sharp escalation in both VaR and ES. Its GEV shape parameter $\xi = 0.4833$ is accurate for a truly heavy tail and the ES/VaR ratio increases from at 95% to 1.88 at 99.9%, showing that once losses cross the VaR threshold, the average loss above this threshold grows faster than the VaR threshold. Both the Kupiec and Christoffersen tests validate this method at the 99.9% level and show it is a good choice for regulatory capital and worst-case stress testing.

Peaks-over-Threshold fitted on the GARCH-filtered residuals tells a more moderate story, and it answers a different question. By stripping out the time-varying volatility component before fitting the GPD, this approach pulls out two things that are usually tangled together: the clustering of volatility already captured by the GARCH filter, and the intrinsic thickness of the residual tail (which is measured as a shape parameter only at $\xi = 0.0803$). A value this low means that once the volatility dynamics are accounted for, what is left is light-tailed, closer to the Gumbel domain than to a truly heavy-tailed one. In practice, much of the extreme losses seen on the MASI are from short-lived volatility spikes, not some deeper structure within the distribution. It is also the only method for which the Kupiec test never rejects the calibration, at any confidence level, and the McNeil-Frey test confirms it estimates VaR and ES jointly with accuracy, which makes it a natural fit for day-to-day risk monitoring, where responsiveness to changing conditions matters more than worst-case conservatism.

These results are consistent with what has already been reported for developed markets (Coles, 2001; Embrechts et al., 1997; McNeil et al., 2015; Bloss et al., 2025), and extend this to the market that is very different in nature: the MASI is thinner, more concentrated, and prone to sudden volatility surges in a way that many developed exchanges are not. The fact that extreme losses cluster in time, as can be seen both in the extremal index below one and in the Christoffersen test rejections at moderate confidence levels, isn't unique to the MASI either. It is a feature it shares with far more liquid markets, which suggests that its tail behavior is less exotic than one might expect from an emerging, lower liquidity index.

Two limitations are worth mentioning here. First, the backtesting is done in sample: parameters are estimated and evaluated on the same period, which is standard in the EVT literature. But that doesn't substitute for a true out-of-sample or rolling-window test of how well these models would have worked in real time. Second, the analysis is only done for one index and therefore it is not guaranteed to be the same for every stock and also for global markets in general. Both of these points suggest promising directions for future work, such as a rolling window backtest of other African or emerging indices or a multivariate EVT treatment of cross-asset dependence and how these models respond to structural breaks such as COVID-19.

Taken as a whole, the picture is fairly clear-cut. The Normal model should not be relied on for tail risk on the MASI, as the distributional mismatch is too large to ignore. The historical method is a reasonable, assumption-free benchmark and runs out of statistical power right where extreme risk measurement is important. The choice between the two EVT methods is less about which one is better and more about what question is being asked: Block-Maxima for the worst case, regulatory view of risk, and POT-GARCH for day-to-day, dynamically adjusted view. When used together rather than in competition, they give a fairly complete picture of how extreme risk behaves on the Moroccan stock market and more generally provide a working template for extreme risk measurement in emerging markets that share the MASI's combination of low liquidity and episodic volatility.

REFERENCES

Assaf, A. (2009), Extreme observations and risk assessment in the equity markets of MENA region: Tail measures and value-at-risk. International Review of Financial Analysis, 18(3), 109-116.

Bekaert, G., Harvey, C.R., Lundblad, C. (2005), Does financial liberalization spur growth? Journal of Financial Economics, 77(1), 3-55.

Bloss, M., Ernst, D., Geisinger, L. (2025), Extreme value theory and gold price extremes, 1975-2025: Long-term evidence on value-at-risk and expected shortfall. Commodities, 4, 24.

Christoffersen, P.F. (1998), Evaluating interval forecasts. International Economic Review, 39(4), 841-862.

Coles, S. (2001), An İntroduction to Statistical Modeling of Extreme Values. Berlin: Springer.

Cont, R. (2001), Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223-236.

Embrechts, P., Klu¨ppelberg, C., Mikosch, T. (1997), Modelling Extremal Events for insurance and Finance. Berlin: Springer.

Fisher, R.A., Tippett, L.H.C. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society, 24, 180-190.

Gencay, R., Sel¸cuk, F. (2004), Extreme value theory and value-at-risk: Relative performance in emerging markets. International Journal of Forecasting, 20(2), 287-303.

Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une s´erie al´eatoire. Annals of Mathematics, 44, 423-453.

Gumbel, E.J. (1958), Statistics of Extremes. Columbia: Columbia University Press.

Hamou, O., Oudgou, M., Boudhar, A. (2025), Analysis of the effectiveness of classical models in forecasting volatility and market dynamics: Insights from the MASI and MASI ESG indices in Morocco. Journal of Risk and Financial Management, 18(7), 370.

Hill, B.M. (1975), A simple general approach to inference about the tail of a distribution. Annals of Statistics, 3, 1163-1174.

Hosking, J. (1985), Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. Applied Statistics, 34, 301-310.

Hosking, J., Wallis, J., Wood, E. (1985), Estimation of the generalized extreme-value distribution by the method of probability weighted moments. Technometrics, 27, 251-261.

Kupiec, P.H. (1995), Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives, 3, 73-84.

Longin, F.M. (1996), The asymptotic distribution of extreme stock market returns. Journal of Business, 69, 383-408.

Longin, F.M. (2000), From value at risk to stress testing: The extreme value approach. Journal of Banking and Finance, 24, 1097-1130.

Macleod, A. (1989), AS R76 - A remark on algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Applied Statistics, 38, 198-199.

McNeil, A.J. (1998), Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory. Switzerland: ETH Zurich.

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.

McNeil, A.J., Frey, R., Embrechts, P. (2015), Quantitative Risk Management: Concepts, Techniques and Tools. London: Princeton University Press.

Prescott, P., Walden, A. (1980), Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Biometrika, 67, 723-724.

Prescott, P., Walden, A. (1983), Maximum likelihood estimation of the parameters of the three-parameter generalized extreme value distribution from censored samples. Journal of Statistical Computation and Simulation, 16, 241-250.

Sherman, B. (1957), Percentiles of the w_n statistic. Annals of Mathematical Statistics, 28, 259-261.

Smith, R.L. (1985), Maximum likelihood estimation in a class of non-regular cases. Biometrika, 72, 67-90.

Yamai, Y., Yoshiba, T. (2005), Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking and Finance, 29(4), 997-1015.