



# A multiobjective optimization approach for threshold determination in extreme value analysis for financial time series

Carlin C. F. Chu<sup>1</sup> · Simon S. W. Li<sup>2</sup>

Received: 27 October 2022 / Accepted: 31 October 2023  
© The Author(s) 2023

## Abstract

The literature on the extreme value theory threshold optimization problem for multiple time series analysis does not consider determining a single optimal tail probability for all marginal distributions. With multiple tail probabilities, their discrepancy results in a differing number of exceedances, which may favour a particular marginal series. In this study, we propose a single optimal tail probability by integrating trade-offs among multiple time series within an MOO framework. Mathematically, our approach links the peaks-over-threshold technique and goal programming technique by developing a set of regression functions, which represent continuous paths of possible tail areas for multiple time series, and we formulate them at the desired levels within a multiobjective optimization framework. The optimal solution is found as the minimum Chebyshev variant weighted value. Our approach advances the development of the peaks-over-threshold method by considering the characteristics of a group of time series collectively instead of independently. The proposed optimal tail probability can be considered an optimal reference point for practical risk investment portfolio analysis that employs an identical tail size across multiple time series data. The daily log returns of four U.S. stock market indices, namely, S&P 500, NASDAQ Composite, NYSE Composite, and Russell 2000, from 1 July 1992 to 30 June 2022 are studied empirically.

**Keywords** EVT · Generalized pareto distribution · Goal programming model · Multiple criteria decision analysis · Peaks over threshold

---

✉ Carlin C. F. Chu  
carlinchu@hsu.edu.hk

<sup>1</sup> Department of Computing, The Hang Seng University of Hong Kong, Hong Kong, China

<sup>2</sup> Big Data Intelligence Centre, The Hang Seng University of Hong Kong, Hong Kong, China

## 1 Introduction

Extreme value theory (EVT) is a powerful and accurate statistical modelling approach to examine extreme values in financial time series in accordance with their asymptotic nature and the behaviour of the underlying probability distribution (Embrechts et al. 1999; Rocco 2014). Researchers have employed EVT based on the peaks-over-threshold (POT) approach as a risk management tool in financial analysis, such as the investigation of the asymmetry of illiquidity measures of stock prices and trading volumes (Będowska-Sójka et al. 2022); inspection of the tail relationships between returns and volumes for high-frequency cryptocurrencies (Chan et al. 2022); comparison of the extreme potential losses and gains of stock indices in both the short and long terms (Chikobvu and Jakata 2020); examination of the stress testing perspective of a semiparametric copula-GARCH risk model for financial return series (Koliai 2016); and estimation of the potential loss for digital banking transaction risks (Saputra and Chaerani 2022). The procedure to identify an effective threshold for determining the tail of extreme values is crucial, as an inappropriate threshold can lead to the problem of bias or inefficiency in risk analysis (Chan et al. 2022). Choosing a small or large threshold value can result in too many or too few observations of exceedances, respectively, which may cause biased parameter estimates with inappropriate measurement data by including nonextreme observations or may induce inefficient parameter estimates with large standard errors by excluding extreme observations. Hence, the identification of an optimal threshold is a critical issue in the risk management of financial time series.

In previous empirical studies of EVT, various POT approaches have been adopted to choose a threshold in extreme value analysis for financial time series. These approaches include the applications of statistical Pareto QQ, Hill, and mean excess plots (Jakata and Chikobvu 2022); parameter estimates by using moments, probability weighted moments, and elementary percentiles (Jocković 2016); Monte Carlo simulation from a known probability distribution (Longin and Solnik 2001); hyperparameter selection based on machine learning algorithms (Nakamura 2021); statistical simulation from a bivariate probability distribution (Verster and Kwaramba 2022); and Bayesian modelling by deriving a posterior distribution for the unknown generalized parameter (Verster and Raubenheimer 2021). All these POT approaches have merely achieved the single objective of choosing an individual or local threshold and analysing the tail properties of extreme values for each corresponding financial time series. For handling multiple financial time series data, the multivariate generalized Pareto distribution (Rootzén and Tajvidi 2006; Rootzén et al. 2018) and generalized copula method (Falk et al. 2019) have proven to be effective. The two methods aim to find the best distributional fitness, and the values of the optimal thresholds for the involved time series are not restricted to be the same as each other. The discrepancy among threshold values (numbers of exceedances) causes a depute of fairness of the overall modelling process and the values of the optimal tail probabilities for the involved time series are not restricted to be the same as each other (Caeiro and Gomes 2015; Roth et al. 2016). The discrepancy among

optimal values causes a depute of fairness of the overall modelling process. In other words, the unbiasedness is arguable, as the estimated model may favour/overfit a particular marginal series by resulting in a larger number of exceedances than other marginal series.

Multiple criteria decision analysis (MCDA) techniques are widely used in the process of financial decision-making as in the areas of business management and accounting, economics and econometrics, computing, decision science, and financial mathematics and engineering (Zopounidis et al. 2015; Spronk et al. 2016). The multiobjective optimization (MOO) approach is one of the core methods of MCDA to systematically evaluate a set of financial actions or criteria that are usually conflicting in nature and produce an optimal decision in accordance with an overall or a global perspective (Steuer and Na 2003; Steuer et al. 2007; Durbach and Stewart 2012). The MOO approach can be employed to aggregate the results of local thresholds, which are identified by using the POT approach, of individual financial time series into a holistic evaluation framework.

In this study, we fill this gap by proposing a single value representing the optimal tail probability that considers the trade-off among the involved multiple time series within an MOO framework. The optimality of the threshold is assessed by its ability to achieve minimum deviations with reference to the tail probabilities of individual financial time series within the group. Mathematically, our proposed approach links up the POT method and MOO approach by using a set of data-driven regression functions that model the trade-off among various tail probabilities. The regression function set serves as the desired levels for the formulation of MOO. The optimality solution is found as the minimum Chebyshev variant weighted value. Our approach advances the development of the peaks-over-threshold method by considering the characteristics of a group of time series collectively instead of independently.

In practice, the assessment of the tail risk level of investment portfolios carried out by investors is based on an identical tail probability across multiple time series data (Dupuis and Jones 2006). The literature does not directly address the need to obtain a single optimal tail probability for achieving the best overall distributional fitness. Our proposed MOO framework fills this gap, and the single optimal value can be considered an optimal reference point. Furthermore, the assessment of the risk level beyond the MOO optimal tail probability can be accurately estimated by a single derived set of GPDs. It is not necessary to re-estimate a distinct set of GPDs for each user-specified probability separately (Dupuis and Jones 2006). Those portfolio tail risk levels can be examined via the truncated distributions conditioned on the user-specified value. In addition, the derived set of GPDs provided by our model provide the best overall distributional fitness under the restriction of having the same number of exceedances in all marginal time series.

The paper is organized as follows: Sect. 2 covers the background of the POT approach and the corresponding local threshold determination method. In Sect. 3, we introduce the formulation of the proposed multiobjective optimization approach for optimal threshold determination with the use of a data-driven regression function. The new formulation is capable of considering the influence from multiple time series obtained from related financial markets effectively. In Sect. 4, we describe the empirical data and address the simulation results. The proposed model is tested on

four U.S. stock market index data for assessing its effectiveness in determining the MOO tail probability. Concluding remarks are provided in the last section.

## 2 POT local threshold approach

### 2.1 Generalized Pareto distribution (GPD)

The POT approach in EVT is based on the formulation of a probability distribution, a generalized Pareto distribution (GPD), depicting the allocations of the extreme values beyond a certain threshold in terms of cut-off value, percentile, or tail probability (1-percentile). This approach extracts extreme values that exceed a sufficiently large threshold value or location parameter  $\varphi$  then converges to a limiting distribution function GPD with shape parameter  $\beta$  and scale parameter  $\omega$ . The probability density function (PDF)  $h_{\varphi,\beta,\omega}(x)$  and cumulative distribution function (CDF)  $H_{\varphi,\beta,\omega}(x)$  associated with the random variable  $X$  over  $\varphi$  (the extracted extreme values) following  $GPD(\varphi, \beta, \omega)$  are given as follows (Chikobvu and Jakata 2020; Chan et al. 2022):

$$h_{\varphi,\beta,\omega}(x) = \frac{1}{\omega} \left( 1 + \beta \frac{(x - \varphi)}{\omega} \right)^{\left(-\frac{1+\beta}{\beta}\right)}$$

$$H_{\varphi,\beta,\omega}(x) = \begin{cases} 1 - \left( 1 + \beta \frac{x - \varphi}{\omega} \right)^{-\frac{1}{\beta}} & \text{for } \beta \neq 0 \\ 1 - e^{-\left(\frac{x - \varphi}{\omega}\right)} & \text{for } \beta = 0 \end{cases}$$

where  $\varphi, \beta \in \mathcal{R}$ ,  $\omega \in \mathcal{R}^+$ , for  $x \geq \varphi$  when  $\beta \geq 0$ , and  $\varphi \leq x \leq \varphi - \frac{\omega}{\beta}$  when  $\beta < 0$ .

### 2.2 Local threshold

The POT threshold estimation method (Longin and Solnik 2001; Chan et al. 2022) assumes that the distribution of extreme values exceeding the local threshold can be modelled by a Monte Carlo simulation of Student's t-distribution with  $k$  degrees of freedom. In accordance with this assumption, the algorithm for finding the local threshold of a single time series can be summarized as the following algorithm:

The variable  $k$  denotes the degree of freedom of the t-distribution. The choice of  $k=1$  to 10 is based on the suggested procedure (Longin and Solnik 2001), which considers the adequacy of capturing different degrees of tail fatness. The objective of the algorithm is to find the optimal degree of freedom  $k^*$  and to locate the corresponding optimal tail probability  $\alpha_k^*$ . Thereafter, the corresponding optimal distributional parameters of GDP can be found accordingly. Additional details can be found in Appendix A of the paper (Longin and Solnik 2001).

- 
- Step 1. Set  $k = 1$ .
  - Step 2. Generate 1000 simulations of  $N$  samples from the Student's  $t$ -distribution with  $k$  degrees of freedom.
  - Step 3. Consider discrete tail probability  $\alpha = 0.01, 0.02, \dots, 0.20$  for each simulation.  
Determine the corresponding set of extreme values  $S_{k,\alpha}$  with tail probability  $\alpha$ . Compute (for each  $S_{k,\alpha}$ ) the shape parameter estimate  $\hat{\beta}_{S_{k,\alpha}}$  using  $H_{\varphi,\beta,\omega}(\cdot)$ .
  - Step 4. Compute the mean squared error  $MSE$  based on the 1000 simulations for each  $\alpha$ :

$$MSE_{k,\alpha} = (\bar{\hat{\beta}} - \beta)^2 + \frac{1}{1000} \sum_{r=1}^{1000} (\hat{\beta}_r - \beta)^2$$

where  $\hat{\beta}_r$  is the shape parameter estimate for each simulation,  $\bar{\hat{\beta}}$  is the average shape parameter estimate of the 1000 simulations, and  $\beta$  is the shape parameter true value based on the theoretical relationship:  $\beta = \frac{1}{k}$ .

- Step 5. Locate the optimal tail probability  $\alpha_k^*$  at the minimum  $MSE_{k,\alpha}$ .
  - Step 6. Repeat Steps 2 to 5 for  $k = 2, 3, \dots, 10$ .  
Determine the optimal degree of freedom  $k^*$  with the minimum  $MSE$  value.
- 

**Algorithm** Determination of local thresholds for a financial time series

### 3 MOO optimal threshold approach

The MOO approach identifies an optimal solution with respect to several objectives. The goal programming model (GPM) is widely adopted to solve multiobjective problems (Ghandforoush 1993; Tamiz et al. 1998; Jones and Tamiz 2010; Ghufuran et al. 2015). A goal in a GPM refers to the desired level, target, or criterion to be attained. The essence of GPM is to incorporate multiple goals into an MOO formulation that can be solved by using a conventional single objective optimization approach (Steuer et al. 2003). For an identified goal, there are three types of target level: exactly the target level (Type I), at most the target level (less is better, Type II), and at least the target level (more is better, Type III) (Jones and Tamiz 2010). A deviational variable is used to compute the difference between the desired level and a given solution on an identified target. Positive or negative deviational variables  $d^+$  or  $d^-$  represent the difference in the achieved values above or below the target level, respectively. The GPM algorithm minimizes the undesired deviational variables to attain the optimal solution with respect to their identified targets. For Type I goals, both positive and negative deviational variables are undesirable. For Type II and Type III goals, the positive and negative deviational variables are undesirable, respectively. There are four commonly used formulations, namely, zero-one normalization weighted, percentage normalization weighted, Chebyshev variant weighted, and lexicographic GPMs (Ghandforoush 1993; Tamiz et al. 1998). In this study, we adopt the Chebyshev variant weighted GPM (GPM<sub>CVW</sub>), which demands the

optimal solution with minimum value across all weighted undesirable deviations to solve the multiobjective problem (Ghufran et al. 2015).

We extend Longin and Solnik’s algorithm listed in Sect. 2.2 with regression functions to capture the relationships between tail size and MSE of simulated data of the chosen degree of freedom. Based on the discrete simulation procedure, the original algorithm can only consider discrete tail sizes (i.e.,  $\alpha=0.01, 0.02, \dots, 0.20$ ). Our method models the tail size as a continuous variable by using a polynomial linear regression function  $g_i(\alpha)$  to represent a continuous path of all possible tail areas for a particular time series  $i$ . The parameters of  $g_i(\alpha)$  are estimated by using the MSE of the simulated data obtained in step 4 of Algorithm 1 (i.e.,  $MSE_{k^*,\alpha}$ ) with the optimal degree of freedom  $k^*$ .

The setup of regression functions is essential to link the POT model with the MOO approach. Each continuous regression function models the characteristics of an individual market, and the regression function set serves as the desired level for the formulation of MOO. The overall formulation of the  $GPM_{CVW}$  is stated below:

Cost function	MinimizeZ
Constraints	$g_1(\alpha)+d_1^- - d_1^+ = g_1(\alpha_1)$ $g_2(\alpha)+d_2^- - d_2^+ = g_2(\alpha_2)$ $\vdots$ $g_m(\alpha)+d_m^- - d_m^+ = g_m(\alpha_m)$ $\frac{(d_1^-ord_1^+)}{g_1(\alpha_1)} \leq Z$ $\frac{(d_2^-ord_2^+)}{g_2(\alpha_2)} \leq Z$ $\vdots$ $\frac{(d_m^-ord_m^+)}{g_m(\alpha_m)} \leq Z$
Boundary conditions	upper bound: $\alpha \leq \alpha_{upper}$ lower bound: $\alpha \geq \alpha_{lower}$
Nonnegativity	$d_1^-, d_1^+, d_2^-, d_2^+, \dots, d_m^-, d_m^+ \geq 0$
Regression function	$g_i(\alpha) = c_6\alpha^6 + c_5\alpha^5 + c_4\alpha^4 + c_3\alpha^3 + c_2\alpha^2 + c_1\alpha + c_0$

where  $g_i(\alpha)$  represents the regression of  $MSE$  on tail probability  $\alpha$ ,  $g_i(\alpha_i)$  represents the target value of minimum  $MSE$ ,  $d_i^-$  represents the negative deviation from  $g_i(\alpha_i)$ , and  $d_i^+$  represents the positive deviation from  $g_i(\alpha_i)$  of the financial time series with index  $i = 1, 2, \dots, m$ . A 6<sup>th</sup> order polynomial linear regression model  $g_i(\alpha_i)$  is used in this study to achieve a satisfactory empirical result. The regression coefficients are estimated by using the  $MSE_{k^*,\alpha}$  obtained in step 4 of Algorithm 1 with the optimal degree of freedom  $k^*$ .  $\alpha_{upper}$  and  $\alpha_{lower}$  represent the upper and lower limits of the tails, respectively. Their values are set to 0.01 and 0.20 with reference to the suggested boundaries in the previous study (Longin and Solnik 2001). The solution of the  $GPM_{CVW}$  represents the optimal tail probability with minimum deviations from the local values among the group of financial time series. Our proposed model is implemented using MATLAB 2019b software (The MathWorks, Inc., Natick, MA, USA) and LINGO 18.0 software (Lindo Systems Inc., Chicago, IL, USA).

## 4 Empirical results

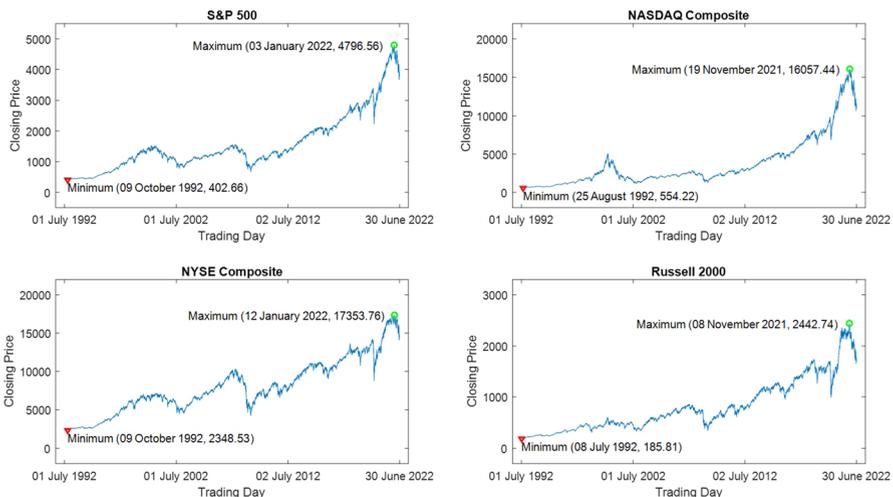
### 4.1 Empirical data

The dataset presented in this study consists of 7556 daily log returns (*DLRs*) of four U.S. stock market indices, namely, the S&P 500, NASDAQ Composite, NYSE Composite, and Russell 2000, from 1 July 1992 to 30 June 2022 covering the periods of dot-com bubble (between April 2001 and December 2001), global financial crisis (between January 2008 and July 2009), and COVID-19 pandemic (between March 2020 and June 2022). Each index consisting of time series data of 7557 trading day information of closing values (from 30 June 1992 to 30 June 2022) is downloaded from the website of Yahoo Finance. The *DLR* at time  $t$  is calculated as follows:

$$DLR(t) = \log\left(\frac{C_t}{C_{t-1}}\right),$$

For  $t = 1, 2, \dots, 7556$  and  $C_0$  is the closing value on 30 June 1992.

The historical closing price time series of the S&P 500, NASDAQ Composite, NYSE Composite, and Russell 2000 indices attained minimum closing prices of 402.66, 554.22, 2348.53, and 185.81 on 09 October 1992, 25 August 1992, 09 October 1992, and 08 July 1992, respectively, and maxima closing prices of 4796.56, 16,057.44, 17,353.76, and 2442.74 on 03 January 2022, 19 November 2021, 12 January 2022, and 08 November 2021, respectively (Fig. 1).



**Fig. 1** Thirty-year historical closing price time series (consisting of 7556 trading days) of the four U.S. stock market indices from 1 July 1992 to 30 June 2022

### 4.2 Daily log returns

The histograms, annualized means, maxima, minimums, annualized SDs, skewness, and kurtoses of the daily log returns depict the distribution patterns of the four U.S. stock market indices over the past thirty years (Fig. 2). The NASDAQ Composite achieved the highest annualized mean *DLR* of 0.0991 compared with 0.0753, 0.0742, and 0.0603 achieved by the Russell 2000, S&P 500, and NYSE Composite, respectively.

### 4.3 Local tail probabilities

The regressions of the mean squared error (*MSE*) on tail probability ( $\alpha$ ) of the S&P 500, NASDAQ Composite, NYSE Composite, and Russell 2000 indices are 6th order polynomials (Fig. 3). The R-squared of each regression is greater than 0.99.

The regression coefficients ( $c_6, c_5, c_4, c_3, c_2, c_1, c_0$ ) for  $i = 1$ (S&P 500), 2 (NASDAQ Composite), 3(NYSE Composite), and 4 (Russell 2000) are given by (30,625, - 21,716, 6118.0, - 872.89, 66.747, - 2.4490, 0.047158), (35,567, - 23,958, 6411.3, - 872.28, 64.160, - 2.2734, 0.046319), (50,697, - 34,794, 9433.8, - 1284.0, 92.319, - 3.1967, 0.052644), and (21,054, - 14,978, 4256.4, - 617.31, 48.791, - 1.8926, 0.040067), respectively. The regressions reached minimum *MSE* values of 0.013690, 0.016288, 0.011089, and 0.012299 at local tail probabilities of 0.03951, 0.03810, 0.03853, and 0.04343, respectively.

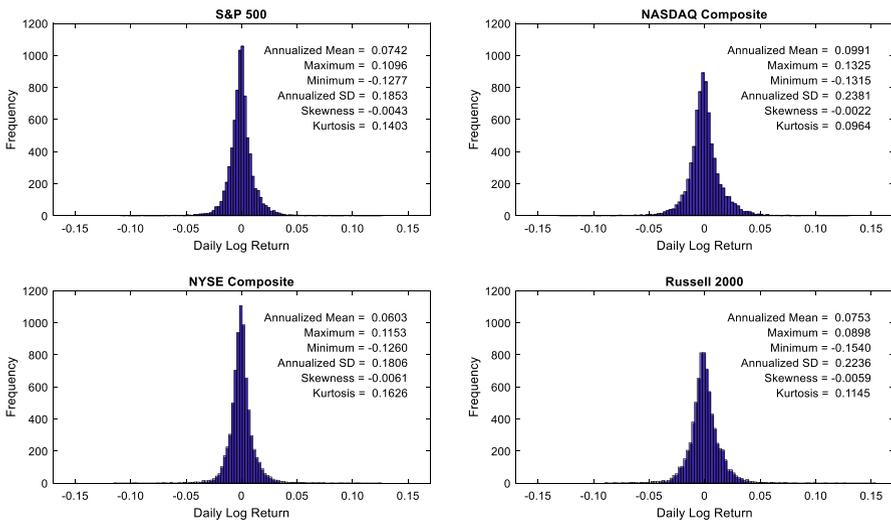
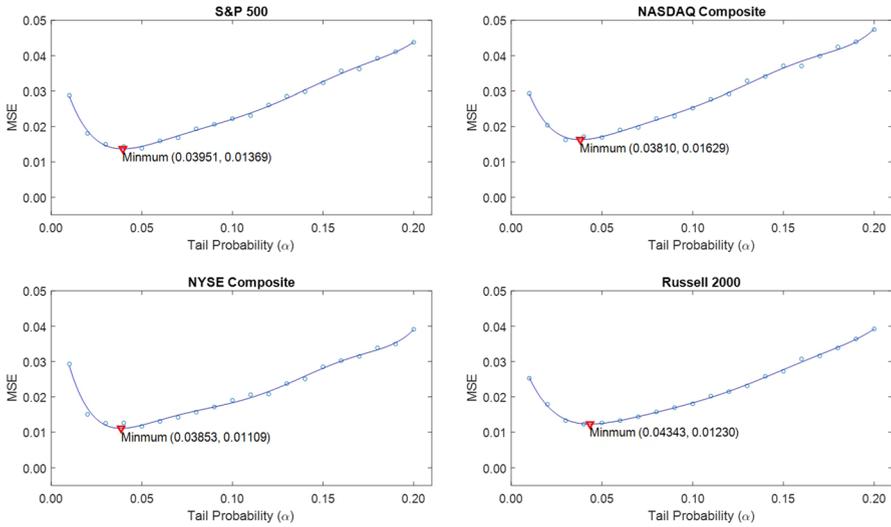


Fig. 2 Histograms of 7556 daily log returns of the four U.S. stock market indices from 1 July 1992 to 30 June 2022



**Fig. 3** The 6th-order polynomial regressions of mean squared error (MSE) on tail probability ( $\alpha$ ) of the four U.S. stock market indices

### 4.4 MOO optimal tail probability

The formulation of the  $GPM_{CVW}$  is given by:

*Minimize*  $k$

*subject to*

$$30625\alpha^6 - 21716\alpha^5 + 6118.0\alpha^4 - 872.89\alpha^3 + 66.747\alpha^2 - 2.4490\alpha + d_1^- - d_1^+ = -0.033469$$

$$35567\alpha^6 - 23958\alpha^5 + 6441.3\alpha^4 - 872.28\alpha^3 + 64.160\alpha^2 - 2.2734\alpha + d_2^- - d_2^+ = -0.030031$$

$$50697\alpha^6 - 34794\alpha^5 + 9433.8\alpha^4 - 1284.0\alpha^3 + 92.319\alpha^2 - 3.1967\alpha + d_3^- - d_3^+ = -0.041555$$

$$21054\alpha^6 - 14978\alpha^5 + 4256.4\alpha^4 - 617.31\alpha^3 + 48.791\alpha^2 - 1.8926\alpha + d_4^- - d_4^+ = -0.027769$$

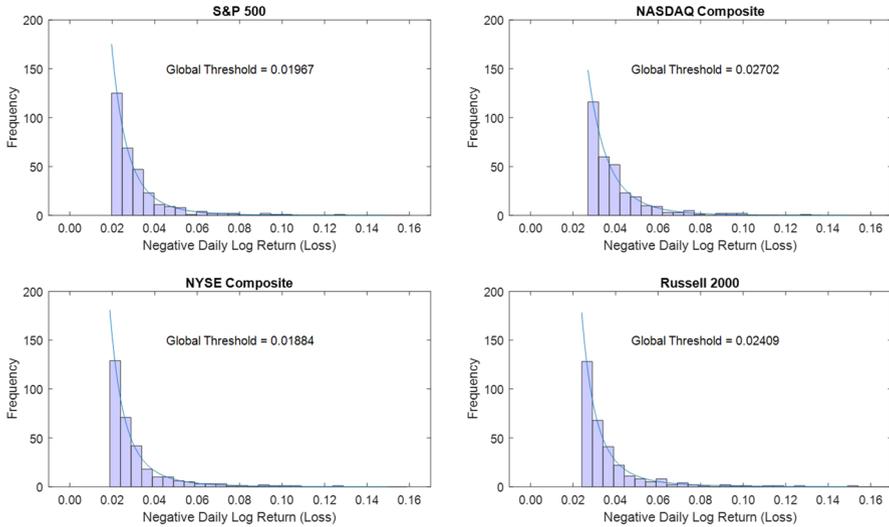
$$\frac{d_1^+}{0.013690}, \frac{d_2^+}{0.016288}, \frac{d_3^+}{0.011089}, \frac{d_4^+}{0.012299} \leq k$$

$$\alpha \geq 0.01, \alpha \leq 0.20, d_1^-, d_1^+, d_2^-, d_2^+, d_3^-, d_3^+, d_4^-, d_4^+ \geq 0$$

The solution of the  $GPM_{CVW}$  representing the MOO optimal tail probability is 0.04073. The distributions and curves of the GPD of the extreme values based on the optimal tail probability of 0.04073 attain global threshold values of 0.01967, 0.02702, 0.01884, and 0.02409 negative *DLRs* for the S&P 500, NASDAQ Composite, NYSE Composite, and Russel 2000 indices, respectively (Fig. 4).

## 5 Summary of GDP parameter estimates

The tail probabilities and number of exceedances, together with the corresponding GPD parameter estimates of the four U.S. stock market indices, are tabulated in Table 1.



**Fig. 4** The distributions and curves generated by the GPD of the 308 extreme negative daily log returns extracted from 7556 daily log returns of the four U.S. stock market indices

The optimal tail probability value (i.e., 0.04073) estimated by the MOO model provides the best overall distributional fitness under the restriction of having the same number of exceedances (i.e., 308) in all marginal time series. Slight differences among the GPD parameters can be observed. For the case in which the optimal tail probability is larger than the locally optimized tail probability, the GPD location parameter  $\varphi$  is adjusted to a smaller value to accommodate the inclusion of additional exceedances, and vice versa. No indicative pattern is observed for the values of shape  $\beta$  and scale  $\omega$  parameters.

However, from the empirical analysis we conclude that the location parameter  $\varphi$  estimated with the use of various local and global threshold values is between 0.01884 and 0.02702 (mean=0.02251) for the S&P 500, NASDAQ Composite, NYSE Composite, and Russell 2000 indices from 01 July 1992 to 30 June 2022 (Table 1). The result is comparable to the findings of a previous study that applied the POT approach to compute tail risk measures of daily returns for six global stock market indices (Gilli and K ellezi 2006).

## 6 Conclusion

Investigation of probable extreme or rare incidents plays a key role in the risk analysis of financial portfolio management in terms of loss or gain for investors with a long or short position on the portfolio, respectively. The application of

**Table 1** Summary of the parameter estimates based on the local and global thresholds Time series of negative daily log returns (01 July 1992–30 June 2022)

	S&P 500	NASDAQ composite	NYSE composite	Russell 2000
Number of trading days	7556			
Local threshold	0.03951	0.03810	0.03853	0.04343
Tail probability				
Parameters of GP distribution	Location ( $\rho$ )	0.02783	0.01930	0.02334
	Shape ( $\beta$ )	0.16118	0.26960	0.26742
	Scale ( $\omega$ )	0.01023	0.00872	0.00894
Global threshold	299	288	291	328
Number of exceedances	0.04073			
Standardized tail probability	0.01967	0.02702	0.01884	0.02409
Parameters of GP distribution	Location ( $\rho$ )			
	Shape ( $\beta$ )	0.14787	0.27390	0.29567
	Scale ( $\omega$ )	0.01037	0.00852	0.00865
Standardized number of exceedances	308			

MCDA is considered a useful tool for regulators, policy-makers, and individual, institutional, and corporate investors (Zopounidis et al. 2015). Accurate estimation of tail risk exposure of value at risk (VaR) and expected shortfall is crucial for risk analysis and portfolio management. The POT approach in EVT provides a powerful framework to quantify extreme risk measures that avoid the underestimation or overestimation of risk levels (Jakata and Chikobvu 2022).

The literature on the extreme value theory threshold optimization problem for multiple time series analysis does not consider determining a single optimal tail probability for all marginal distributions. With multiple tail probabilities, the discrepancy results in a differing number of exceedances, which may favour a particular marginal series. In this study, we develop a single optimal tail probability by integrating trade-offs among multiple time series within an MOO framework. Our method unites the POT technique and MOO framework. Here, we develop a set of regression functions, representing continuous paths of possible tail areas for multiple time series. The regression functions are then formulated as the desired levels within an MOO framework. The optimality of MOO tail probability is assessed by its ability to achieve minimum Chebyshev variant weighted deviations with reference to the tail probabilities of individual financial time series within the group.

Apart from proper estimation of the portfolio tail risk level, an identification of extreme returns is vital in formulating the connectedness among financial time series and in identifying contagion during financial turmoil (Bae et al. 2003). Researchers have investigated the linkages and temporal relationships for rare incidents in multivariate time series (Arnold et al. 2007; Chen and Chihying 2007) as well as transmission and comovements of prices in financial markets by using the correlation among stocks that can be employed to signify the dominance of specific stocks in accordance with the correlation-based network under consideration (Ganeshapillai et al. 2013). The optimal tail probability identified in this study can help investors connect a group of financial time series by using the same number of extreme returns and identify the risk that is generated in tandem. Furthermore, we consider the negative *DLRs* of four U.S. stock market indices, adopt the POT approach to identify the local threshold, and employ the Chebyshev variant weighted GPM to find the global threshold. In future studies, researchers can evaluate both positive and negative *DLRs* of stock prices or market indices in other countries (Koliai 2016; Będowska-Sójka et al. 2022), can use other methods, such as the block maxima (BM) method, to find the local threshold (Nakamura 2021; Będowska-Sójka et al. 2022), and can apply other weighted lexicographic GPMs to choose the global threshold (Jones and Tamiz 2010).

To conclude, we propose a novel MOO threshold determination framework that is capable of producing a single optimal tail probability for obtaining the same number of exceedances across multiple time series. The MOO tail probability can be considered an optimal reference point for practical risk investment portfolio analysis that employs an identical tail size across multiple time series data. The assessment of risk level beyond the MOO optimal tail can be accurately estimated via the truncated distributions conditioned on the user-specified value. Moreover, the calculation method is demonstrated empirically with the use

of four U.S. stock market index data covering the period from 1 July 1992 to 30 June 2022.

**Acknowledgements** The work reported in this paper was supported by the Research Matching Grant Scheme (HSU#700029) administered by the University Grants Committee in Hong Kong.

**Author contributions** The two authors contributed equally to this work.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arnold A, Liu Y, Abe N (2007) Temporal causal modeling with graphical granger methods. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining. pp 66–75
- Bae KH, Karolyi GA, Stulz RM (2003) A new approach to measuring financial contagion. *Rev Financ Stud* 16(3):717–763
- Beđowska-Sójka B, Echaust K, Just M (2022) The asymmetry of the Amihud illiquidity measure on the European markets: the evidence from extreme value theory. *J Int Finan Markets Inst Money* 78:101563
- Caeiro F, Gomes MI (2015) Threshold selection in extreme value analysis. *Extrem Value Model Risk Anal Methods Appl* 1:69–87
- Chan S, Chu J, Zhang Y, Nadarajah S (2022) An extreme value analysis of the tail relationships between returns and volumes for high frequency cryptocurrencies. *Res Int Bus Financ* 59:101541
- Chen P, Chihying H (2007) Learning causal relations in multivariate time series data. *Economics* 1(1):20070011
- Chikobvu D, Jakata O (2020) Analysing extreme risk in the south african financial index (j580) using the generalised extreme value distribution. *Stat Optim Inf Comput* 8(4):915–933
- Dupuis DJ, Jones BL (2006) Multivariate extreme value theory and its usefulness in understanding risk. *North Am Actuar J* 10(4):1–27
- Durbach IN, Stewart TJ (2012) Modeling uncertainty in multi-criteria decision analysis. *Eur J Oper Res* 223(1):1–14
- Embrechts P, Resnick SI, Samorodnitsky G (1999) Extreme value theory as a risk management tool. *North Am Actuar J* 3(2):30–41
- Falk M, Padoan SA, Wishechel F (2019) Generalized pareto copulas: a key to multivariate extremes. *J Multivar Anal* 174:104538
- Ganeshapillai G, Gutttag J, Lo A (2013) Learning connections in financial time series. In: International conference on machine learning. pp 109–117. PMLR
- Ghandforoush P (1993) Optimal allocation of time in a hospital pharmacy using goal programming. *Eur J Oper Res* 70(2):191–198

- Ghufuran S, Khowaja S, Ahsan MJ (2015) Optimum multivariate stratified double sampling design: Chebyshev's goal programming approach. *J Appl Stat* 42(5):1032–1042
- Gilli M, Këllezi E (2006) An application of extreme value theory for measuring financial risk. *Comput Econ* 27(2):207–228
- Jakata O, Chikobvu D (2022) Extreme value modelling of the monthly south african industrial index (J520) returns. *Stat Optim Inf Comput* 10(2):383–400
- Jocković, J. (2016). Quantile estimation for the generalized pareto distribution with application to finance. *Yugoslav Journal of Operations Research*, 22(2).
- Jones D, Tamiz M (2010) Practical goal programming, vol 141. Springer, New York
- Koliai L (2016) Extreme risk modeling: an EVT–pair-copulas approach for financial stress tests. *J Bank Finance* 70:1–22
- Longin F, Solnik B (2001) Extreme correlation of international equity markets. *J Financ* 56(2):649–676
- Nakamura C (2021) On Choice of Hyper-parameter in extreme value theory based on machine learning techniques. arXiv preprint [arXiv:2107.06074](https://arxiv.org/abs/2107.06074)
- Rocco M (2014) Extreme value theory in finance: a survey. *J Econ Surv* 28(1):82–108
- Rootzén H, Tajvidi N (2006) Multivariate generalized Pareto distributions. *Bernoulli* 12(5):917–930
- Rootzén H, Segers J, Wadsworth JL (2018) Multivariate generalized Pareto distributions: parametrizations, representations, and properties. *J Multivar Anal* 165:117–131
- Roth M, Jongbloed G, Buishand TA (2016) Threshold selection for regional peaks-over-threshold data. *J Appl Stat* 43(7):1291–1309
- Saputra MPA, Chaerani D (2022) Estimation of maximum potential losses for digital banking transaction risks using the extreme value-at-risks method. *Risks* 10(1):10
- Spronk J, Steuer RE, Zopounidis C (2016) Multicriteria decision aid/analysis in finance. Multiple criteria decision analysis. Springer, New York, pp 1011–1065
- Steuer RE, Na P (2003) Multiple criteria decision making combined with finance: a categorized bibliographic study. *Eur J Oper Res* 150(3):496–515
- Steuer RE, Qi Y, Hirschberger M (2007) Suitable-portfolio investors, nondominated frontier sensitivity, and the effect of multiple objectives on standard portfolio selection. *Ann Oper Res* 152(1):297–317
- Tamiz M, Jones D, Romero C (1998) Goal programming for decision making: an overview of the current state-of-the-art. *Eur J Oper Res* 111(3):569–581
- Verster A, Kwaramba N (2022) A Different way of choosing a threshold in a bivariate extreme value study. *Stat Optim Inf Comput* 10(2):505–518
- Verster A, Raubenheimer L (2021) A different approach for choosing a threshold in peaks over threshold. *Stat Optim Inf Comput* 9(4):838–848
- Zopounidis C, Galariotis E, Doumpos M, Sarri S, Andriosopoulos K (2015) Multiple criteria decision aiding for finance: an updated bibliographic survey. *Eur J Oper Res* 247(2):339–348

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.