AIMS Mathematics

AIMS Mathematics, 11(5): 14096–14120.
DOI: 10.3934/math.2026579
Received: 19 March 2026
Revised: 26 April 2026
Accepted: 28 April 2026
Published: 19 May 2026

https://www.aimspress.com/journal/Math

Research article

An enhanced Monte Carlo simulation framework with Backtest-Driven GMM and deterministic initialization for Portfolios risk estimation

Ülkü Erisoglu[1], Selim Gunduz[2,*] and Mert Yaman[1]

[1] Department of Statistics, Faculty of Science, Necmettin Erbakan University Konya 42090, Turkey
[2] Department of Business Administration, Faculty of Economics, Administrative and Social Sciences, Adana Alparslan Türkeş Science and Technology University, Adana 01250, Turkey

* **Correspondence:** Email: sgunduz@atu.edu.tr.

Abstract: This study proposes an enhanced Gaussian mixture model (GMM)-based Monte Carlo simulation design to improve value-at-risk (VaR) estimation performance, specifically tailored for high-volatility environments such as global cryptocurrency markets. The proposed method deals with the critical problems of parameter instability due to random initialization and the limitations of determining the optimal number of components using standard information criteria like AIC or BIC, which are frequent challenges in traditional GMM-based VaR estimation approaches. To overcome these problems, a deterministic clustering algorithm was developed to ensure a stable initialization of GMM, and a simulation design like the parametric bootstrap was applied using the model fitted from observational data. Furthermore, a direct backtest performance was considered in determining the number of GMM components. The empirical application of the study was conducted through the VaR modeling of daily returns for a diversified portfolio consisting of BTC, ETH, BNB, and SOL crypto assets. A dynamic VaR model, utilizing 250-day rolling windows, was developed, and its performance was compared with traditional VaR estimation methods. VaR estimates, calculated at various confidence levels, consistently demonstrated that the proposed approach outperformed traditional methods. The results indicate that the proposed method markedly improves tail risk estimation accuracy, achieving higher success in satisfying both coverage and independence criteria.

Keywords: value-at-risk; backtest; cryptocurrencies; Monte Carlo simulation; Gaussian mixture model
Mathematics Subject Classification: 62H10, 62P20

1. Introduction

With the introduction of Bitcoin by Nakamoto in 2008 [24], cryptocurrencies have become a rapidly growing investment vehicle in financial markets. These assets, which attract investors due to their high return potential, have reached a significant market volume over time. However, when examining past returns, it is evident that cryptocurrencies are highly volatile, characterized by sharp rises and falls. Due to this characteristic, cryptocurrencies are generally classified as high-risk investment instruments, and effective risk management is necessary.

One of the key components of financial risk management is the quantitative measurement of risk; one of the most widely used measures is value-at-risk (VAR). VaR provides a single quantitative estimate of the maximum expected loss over a certain period and at a specified confidence level [11]. Many different methods have been developed in the literature for VAR estimation, and these methods are generally grouped into three categories: parametric methods, historical simulation methods, and Monte Carlo simulation methods [19].

In the literature, the high volatility of assets in the portfolio poses a challenge to the adequacy of methods that rely on the assumption of normality. Therefore, alternative distributions and flexible modeling approaches have gained prominence [12]. Different distribution approaches have been proposed in the literature for more consistent VaR estimates. Tan and Chu [29], Contreras [5], and Li and Wu [17] suggested using a Gaussian mixture model (GMM) for VaR estimation and stated that VaR estimation using a GMM yielded better results than other classical methods. Erisoglu and Koroglu [8] demonstrated that the mixture distribution method, a combination of normal and log-Dagum distributions, performs well in VaR estimation. For VaR calculations related to cryptocurrencies, Silahli et. al. [28] recommended the two-tailed Weibull distribution, while Hrytsiuk et al. [12] recommended the Cauchy distribution. Gkillas and Longin [9] used a generalized extreme value distribution to model Bitcoin's extreme tail behavior and noted that this method provides more accurate risk estimates compared to traditional VaR methods. Jiang et al. [13] proposed a new statistical model based on mixture distributions with time-varying parameters in order to make more flexible and accurate VaR estimates in cases where traditional models are insufficient in measuring the extreme volatility structure of cryptocurrencies. Backtest results showed that the proposed model provided more accurate VaR estimates than traditional models at 99% and 95% confidence levels. Chen et al. [3] tested the performance of different distribution assumptions under various GARCH-type models and found that heavy-tailed and asymmetric distributions yielded more successful results in VaR estimation, whereas Gaussian distributions performed better in volatility estimation.

In the literature, studies employing GMM-based Monte Carlo simulation estimate VaR by first identifying the best-fitting model, generating random samples, and then calculating the quantile values corresponding to the 5% or 1% significance level. Seyfi et al. [27] tested the performance of their proposed Monte Carlo simulation method based on the GMM for estimating the portfolio's VaR and expected shortfall (ES) with Christoffersen's CCI and CI tests. Similarly, Morkūnaitė et al. [22] demonstrated that a multivariate GMM-based Monte Carlo simulation method for VaR estimation in cryptocurrency portfolios yields more successful results than traditional single-mode normal distribution–based approaches in modeling the multimodal distribution characteristics of cryptocurrency assets and the complex dependency structures between assets.

Backtesting is commonly used for evaluating the reliability of VAR estimates; in most studies comparing the performance of VaR methods, Kupiec's POF, Christoffersen's independence (CCI), and interval tests (CI) are widely used. For example, Seyfi et al. [27], Likitratcharoen et al. [18], Panda and Deb [25], Mallela and Gupta [20], and Morkūnaitė et al. [22] used these test statistics as performance criteria in their studies.

In studies using Monte Carlo simulation methods based on multivariate GMM, the number of components in the model is often determined according to information criteria such as AIC or BIC. However, these criteria focus on the overall fit of the model and do not directly optimize VAR estimation performance, where extreme risks should be considered. Additionally, Monte Carlo simulations are typically conducted using only a single synthetic sample, which makes the estimates overly sensitive to randomness and raises concerns about the reliability of the VAR estimates. Furthermore, measures such as confidence intervals for VAR estimates cannot be produced, which limits the effectiveness of risk measurement in decision support systems. While these simulation-based challenges persist, alternative advanced optimization techniques have also been explored in literature. For instance, Leung and Wang [16] proposed a collaborative neurodynamic optimization approach to address minimax and biobjective portfolio selection problems.

Risk assessment and management are not limited to financial markets; they play a vital role in the decision-making processes of complex systems. Recent studies showed that risk analysis methodologies have evolved across a wide range of fields, from engineering to supply chain management. Advanced mathematical models developed to optimize the performance of systems under uncertainty share similar fundamental theoretical frameworks with financial risk forecasting methods.

For example, in modern production systems and sustainable engineering designs, risk assessment is critical for ensuring operational continuity. In this context, the approach presented by Vahabzadeh et al. [30] offers a current and comprehensive perspective on how risk management can be framed within multi-criteria decision-making processes. This study emphasizes that risk is not only a financial loss but also a multidimensional variable affecting system reliability.

Returning to the financial sphere, risk management for assets with high volatility and "heavy-tail" characteristics, such as cryptocurrencies, requires more dynamic solutions compared to traditional assets. This study is conducted to improve the effectiveness of the GMM-based Monte Carlo simulation method by targeting the limitations described above. First, in Monte Carlo–based VAR estimations, the number of model components is selected based on backtests' performance rather than information criteria. Second, instead of using a single large synthetic sample in the simulation process, a new simulation design based on the portfolio size for each model is proposed. Also, variance values and confidence intervals related to VaR estimations can be generated. Instead of randomly initializing the parameters in GMM estimation, a simple and effective method for determining the initial cluster centers and assigning component memberships is proposed in our method.

The rest of this study is organized as follows: In Section 2, the main methods used in VaR estimations—parametric method, historical simulation method, and Monte Carlo simulation method—are explained with their theoretical background. Section 3 introduces backtesting, focusing on the theoretical basis of the Kupiec Proportion of Failures (POF) test, Christoffersen independence test, and Christoffersen Interval tests. Section 4 describes the modeling approach based on a multivariate GMM and explains how the model parameters are estimated. Section 5 presents the new Monte Carlo simulation-based method proposed in this study. It explains the simulation design, which allows building confidence intervals from the VaR distribution, and selects the number of components based on backtesting performance instead of information criteria. In Section 6, an application to cryptocurrencies is carried out, and the proposed method is compared with other methods in the literature. Finally, Section 7 summarizes the main findings, discusses the strengths of the model, and gives suggestions for future research.

2. Value-at-risk (VaR) and estimation methods

VaR is a statistical measure of risk that expresses the maximum expected loss to which a portfolio may be exposed within a specific time under normal market conditions and at a specific confidence level.

Let X be the random variable representing the return distribution of the portfolio. If $X > 0$, it indicates a gain; if $X < 0$, it indicates a loss. VaR is expressed as follows [23]:

$$VaR_\alpha(X) = -F_X^{-1}(\alpha). \tag{1}$$

The negative sign in the equation is used to express the loss as positive. This measure indicates the magnitude of the loss in the tail region with a probability of occurrence $\alpha \in (0,1)$.

In this study, VaR estimation methods are examined under three subheadings: the parametric method, also known as the variance-covariance approach, the historical simulation method, and the Monte Carlo simulation method.

2.1. Parametric method

In this method, VaR is calculated under the assumption that the returns in the portfolio are normally distributed. For a portfolio, VaR is calculated using the following equation:

$$VaR_\alpha = -\left(\mu + z_\alpha \sigma \sqrt{t}\right)A, \tag{2}$$

where μ is the mean of the portfolio returns, and σ is the standard deviation. t denotes the holding period, A denotes the investment amount, and z_α denotes the critical table value from the standard normal distribution at the α significance level.

In portfolios containing multiple financial instruments, the assumption of a multivariate normal distribution is naturally assumed. The probability density function of the multivariate normal distribution is given by the mean vector μ, the variance-covariance matrix Σ, and the number of variables p, expressed as:

$$f(\mathrm{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(\mathrm{x}-\mu)\Sigma^{-1}(\mathrm{x}-\boldsymbol{\mu})'\right). \tag{3}$$

The average return and volatility of the portfolio are calculated using the investment weights (w) of the assets in the portfolio:

$$\mu_{\text{portfolio}} = \mathrm{w}\boldsymbol{\mu}', \tag{4}$$

$$\sigma_{\text{portfolio}} = (\mathrm{w}\Sigma\mathrm{w}')^{1/2}. \tag{5}$$

Equations (4) and (5) are used to calculate the VaR of the portfolio in a manner similar to Eq (2):

$$VaR_{\alpha,portfolio} = \left(\mu_{portfolio} + z_\alpha \sigma_{portfolio}\sqrt{t}\right)A. \tag{6}$$

2.2. Historical simulation method

The historical simulation method is a nonparametric method used in VaR estimation. In this method, an empirical cumulative distribution function is created based on past returns, assuming the return behavior of the portfolio will continue in a similar way in the future, and the relevant quantile value at the α significance level is used as the VaR estimate.

Given the distribution R of the portfolio's past returns, VaR_α is calculated using the following equation at the α significance level:

$$\mathrm{HVaR}_\alpha(\mathrm{X}) = -\mathrm{S}_\mathrm{R}^{-1}(\alpha), \tag{7}$$

$S_R^{-1}(.)$ represents the inverse of the empirical distribution function, i.e., the empirical quantile function.

2.3. Monte Carlo simulation method

The basic assumption in the Monte Carlo simulation method is that the return behavior in the portfolio will continue in the future in a similar way as in the historical simulation method. Therefore, in this method, the optimum return distribution is determined according to the distribution of past returns, and VaR is calculated with the scenario generated from this distribution. Since the correlation structure between the assets in the portfolio is also considered, it is mostly based on the multivariate normal distribution [10]. In the Monte Carlo simulation based on the multivariate normal distribution, the mean vector and covariance matrix are first calculated from the historical returns of the assets in the portfolio. Using the mean vector and covariance matrix, a sufficiently large random sample of 10,000 units is generated from the multivariate normal distribution. Synthetic returns are calculated by taking into account the weight of the assets in the portfolio in the random sample generated. Synthetic returns are ranked from smallest to largest, and the quantile value corresponding to the significance level α is used as the VaR estimate.

3. Backtesting methodology

The performance of VaR models is analyzed by backtesting, which evaluates the extent to which the estimated risk level is consistent with realized losses. For this purpose, the binomial test, Kupiec's POF test, and Christoffersen's interval and independence test are the most frequently used tests in terms of computational simplicity and efficiency.

3.1. Binomial test

The most basic test method to evaluate the success of the VaR model is the binomial test, which compares the number of realized exceptions with the number of expected exceptions [14]. Given the number of realized exceptions x, the confidence level used in VaR estimation 1-α, and the number of observations n, the binomial test statistic under the normal distribution approach is calculated as follows:

$$Zbin = \frac{x-np}{\sqrt{np\,(1-p)}} \sim N(0,1), \tag{8}$$

where nα denotes the expected number of exceptions. The binomial test only tests the coverage of the VaR model. A $p - value < \alpha$ indicates that the VaR model is not significant in determining the risk threshold.

3.2. Kupiec's POF test

The POF test developed by Kupiec [15] evaluates whether the actual coverage ratio is in line with the expected coverage ratio using the likelihood ratio test and is calculated as follows:

$$LR_{POF} = -2\log\left(\frac{(1-\alpha)^{n-x}\alpha^{n}}{\left(1-\frac{x}{n}\right)^{n-x}\left(\frac{x}{n}\right)^{x}}\right) \sim \chi^2_{(1)}. \quad (9)$$

The equation can also be represented as

$$LR_{POF} = -2 \times \{(n-x)\log(1-\alpha) + n\log(\alpha) - (n-x)\log(1-\hat{\alpha}) - n\log(\hat{\alpha})\},$$

with the realized coverage ratio $\hat{\alpha} = \frac{x}{n}$ and logarithmic function properties.

The calculated likelihood value follows a Chi-squared distribution with one degree of freedom. This test evaluates the number of realized exceptions as in the binomial test, i.e., it only tests the coverage of the VaR model.

3.3. Christoffersen's interval forecast tests

Christoffersen [4] focused on the dependency between exceptions. For this reason, this test is also known as the independence test in the literature. If there is dependency between exceptions, the model is rejected based on the conclusion that the VaR cannot capture periodic volatility in the market. Christoffersen's method defines VaR exceptions as a time series:

$$I_t = \begin{cases} 1, & r_t < VaR_\alpha \\ 0, & r_t \geq VaR_\alpha \end{cases},$$

and defines a transition count matrix between I_{t-1} and I_t as follows:

$$\begin{bmatrix} N_{00} & N_{01} \\ N_{10} & N_{11} \end{bmatrix},$$

where N_{00} is the number of periods with no exceptions followed by a period with no exceptions. N_{10} is the number of periods with exceptions followed by a period with no exceptions. N_{01} is the number of periods with no exceptions followed by a period with exceptions. N_{11} is the number of periods with exceptions followed by a period with exceptions. With the help of this matrix, transition probabilities $\pi_{01} = \frac{N_{01}}{N_{00}+N_{01}}$ and $\pi_{11} = \frac{N_{11}}{N_{11}+N_{10}}$ are obtained.

Under the independence assumption,

$$\pi_{01} = \pi_{11} = \pi = \frac{N_{01} + N_{11}}{N_{00} + N_{01} + N_{10} + N_{11}}$$

are obtained. According to these definitions, Christoffersen's CCI test is calculated as follows:

$$LR_{CCI} = -2\log\left(\frac{(1-\pi)^{N_{00}+N_{10}}\pi^{N_{01}+N_{11}}}{(1-\pi_{01})^{N_{00}}(\pi_{01})^{N_{01}}(1-\pi_{11})^{N_{10}}(\pi_{11})^{N_{11}}}\right) \sim \chi^2_{(1)}. \quad (10)$$

The calculated statistical value shows a chi-square distribution with a single degree of freedom. Under the Ho hypothesis, $\pi_0 = \pi_1$. If these ratios differ significantly from each other, the model is considered unreliable [31].

3.4. Conditional coverage mixed test (joint test)

This test is a combination of Kupiec's POF test and Christoffersen's interval forecast test. While Kupiec's test focuses on the number of exceptions, Christoffersen's test focuses on the dependency of exceptions. By combining the two tests, a more consistent test is obtained by considering both the number and dependency of exceptions. However, it has a limited capability to represent only one of the two features. For this reason, it would be more accurate to join Kupiec's POF and Christoffersen's interval forecast test. The test statistics are obtained as follows:

$$LR_{CC} = LR_{POF} + LR_{CCI}. \quad (11)$$

The calculated statistical value shows a chi-square distribution with two degrees of freedom.

4. Mixture distribution model of multivariate normal distributions

A mixture distribution model is a probability distribution obtained by weighted sums of G different distributions. Each distribution in the mixture has its own weight and parameters.

$$f(x, \Theta) = \sum_{i=1}^{G} \pi_i f_i(x, \theta_i). \quad (12)$$

The notation $f_i(x, \theta_i)$ in the equation denotes a probability density function with the parameter vector θ_i representing subpopulation i. π_i is the weight of subpopulation i in the mixture distribution model and takes values in the range (0,1). Θ is the parameter vector of the mixed distribution consisting of the mixed weights and the parameter vectors of the subpopulations $\Theta = [\pi_1 \;\; \dots \;\; \pi_G \;\; \theta_1 \;\; \dots \;\; \theta_G]$. Depending on the structure of the data, a mixture distribution can be obtained with different weights and different distributions. In addition, a mixture distribution can also be obtained with a single distribution. Gaussian mixture models can be given as an example of mixture distributions.

The probability density function of the mixture distribution of multivariate normal distributions can be expressed as follows:

$$f(\mathbf{x}, \Theta) = \sum_{i=1}^{G} \pi_i \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)\Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)'\right\}. \quad (13)$$

In the mixture distribution of multivariate normal distributions, the parameter vector is expressed by the equation $\Theta = [\pi_1 \;\; \dots \;\; \pi_G \;\; \boldsymbol{\mu}_1 \;\; \Sigma_1 \;\; \dots \;\; \boldsymbol{\mu}_G \;\; \Sigma_G]$. The maximum likelihood method, which is one of the most well-known methods, is used for parameter estimation. The latent variables are found using the EM algorithm; thus, the probabilities of data belonging to the relevant distributions are obtained. Then, the parameters of the distribution are obtained by calculating the MLE. When the mentioned calculations are made, the parameters of the mixed distribution model of multivariate normal distributions are calculated at the (r + 1) iteration as follows:

$$\hat{\pi}_i^{(r+1)} = \frac{1}{n}\sum_{j=1}^{n} z_{ij}^{(r)}, \tag{14}$$

$$\hat{\boldsymbol{\mu}}_i^{(r+1)} = \frac{1}{\sum_{j=1}^{n} z_{ij}^{(r)}}\sum_{j=1}^{n} z_{ij}^{(r)}\mathbf{x}_j, \tag{15}$$

$$\hat{\Sigma}_i^{(r+1)} = \frac{1}{\sum_{j=1}^{n} z_{ij}^{(r)}-1}\sum_{j=1}^{n} z_{ij}^{(r)}\left(\mathbf{x}_j - \hat{\boldsymbol{\mu}}_i^{(r)}\right)\left(\mathbf{x}_j - \hat{\boldsymbol{\mu}}_i^{(r)}\right)'. \tag{16}$$

The notation $z_{ij}^{(r)}$ in the equations represents the estimated probability that the j-th unit belongs to the i-th subpopulation at the r-th iteration and is obtained using the following equation:

$$z_{ij}^{(r)} = \frac{\hat{\pi}_i^{(r)} f_i(\mathbf{x}_j;\hat{\boldsymbol{\mu}}_i^{(r)},\hat{\Sigma}_i^{(r)})}{\sum_{l=1}^{G}\hat{\pi}_l^{(r)} f_l(\mathbf{x}_j;\hat{\boldsymbol{\mu}}_l^{(r)},\hat{\Sigma}_l^{(r)})}. \tag{17}$$

The algorithm is repeated until the relevant convergence criteria are obtained. Although the EM algorithm provides satisfying solutions, it is sensitive to the initial points.

The Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC), widely used in estimating the appropriate number of components in the mixture of a multivariate normal distribution, are calculated as follows:

$$AIC = -2logL(\widehat{\Theta}) + 2d, \tag{18}$$

$$BIC = -2logL(\widehat{\Theta}) + \log(n)d, \tag{19}$$

where $logL(\widehat{\Theta})$ is the maximum log likelihood value for the MLE obtained with the EM algorithm. In the equations, d is the total number of free parameters in the model. While AIC prefers more flexible models, i.e., models with a higher number of components, BIC generally recommends models with a lower number of components, since it uses a larger number of parameters used as a penalty term multiplied by $log(n)$ [1].

5. Proposed method

In this study, we propose an enhanced and novel Monte Carlo simulation method based on the Gaussian mixture model (GMM) that addresses several limitations of traditional approaches. The proposed method aims to enhance the accuracy and stability of VaR estimates by restructuring three critical steps: (i) selecting the optimal number of mixture components, (ii) initializing the GMM parameters effectively, and (iii) designing a more robust simulation procedure. The systematic workflow of the proposed enhanced methodology is shown in Figure 1, providing a visual representation of the integrated estimation and simulation framework.

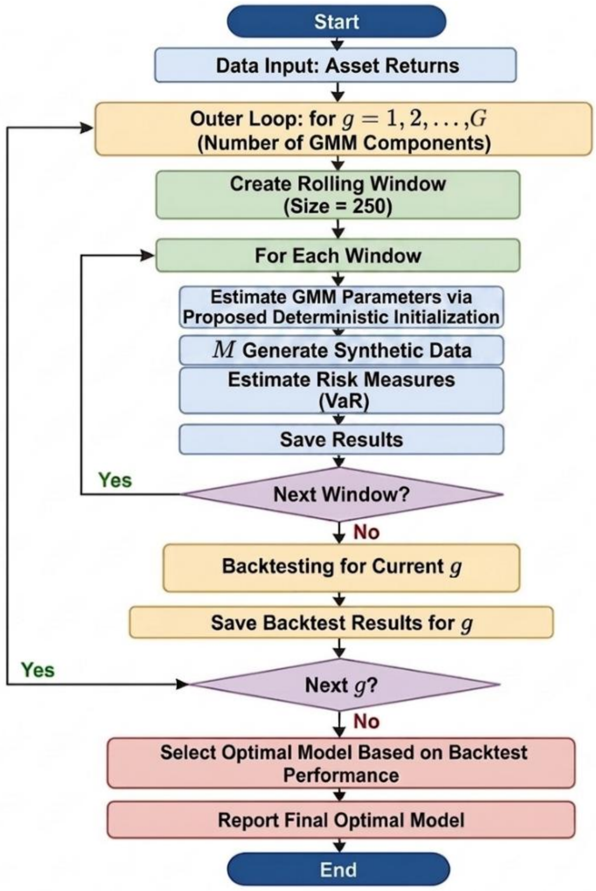

Figure 1. Flowchart of the proposed enhanced Monte Carlo simulation framework.

5.1. Limitations of traditional Monte Carlo GMM-based VaR estimation

Traditional Monte Carlo–based VaR estimation typically involves generating 10,000 synthetic samples from a GMM fitted to historical data and estimating VaR by selecting the α-quantile (e.g., 1% or 5%) of the simulated return distribution. However, despite the seemingly large sample size, this approach remains vulnerable to statistical fluctuations arising from the inherent randomness of the simulation process. Moreover, the number of mixture components is commonly determined using information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) [21]. These criteria prioritize overall model fit but often lack the sensitivity needed to capture tail behavior accurately, especially in the context of extreme risk modeling [6]. As a result, models selected based on these criteria may be inadequate for reliably estimating extreme losses.

5.2. Proposed improved Monte Carlo simulation design for estimating VaR distribution

While traditional approaches generate a fixed number of synthetic data (typically 10,000) regardless of sample size, our proposed method employs GMM to repeatedly generate synthetic samples of varying data size and structure, similar to the parametric bootstrap approach.

M synthetic samples, each with size n, are generated based on GMM parameter estimates obtained from past returns in the portfolio, where n is the number of days of past returns in the portfolio.

For each synthetic sample, an approximate sampling distribution for $V\hat{a}R_{\alpha}^{(1)}, V\hat{a}R_{\alpha}^{(2)}, \dots, V\hat{a}R_{\alpha}^{(M)}$ is generated determining the quantile α, with $m = 1,2, \dots, M$. In our new method, the final VaR estimate at the significance level α is:

$$V\hat{a}R_{\alpha}^{new} = \frac{1}{M}\sum_{m=1}^{M} V\hat{a}R_{\alpha}^{(m)}. \quad (20)$$

This simulation structure mitigates the sensitivity of VaR estimates to random variation and enables the construction of confidence intervals based on the approximate sampling distribution. As a result, it offers a more stable and robust framework for risk measurement compared to traditional Monte Carlo methods. Two different confidence intervals can be created using the normal distribution assumption and the percentage approach based on M VaR estimates.

The standard error of VaR estimates, obtained from M simulations $V\hat{a}R_{\alpha}^{(1)}, V\hat{a}R_{\alpha}^{(2)}, \dots, V\hat{a}R_{\alpha}^{(M)}$, is estimated as follows:

$$SE_{V\hat{a}R_{\alpha}^{new}} = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(V\hat{a}R_{\alpha}^{(m)} - V\hat{a}R_{\alpha}^{new}\right)^2}. \quad (21)$$

The 1-γ confidence interval under the assumption of normal distribution is calculated as follows:

$$\left[V\hat{a}R_{\alpha}^{new} - z_{1-\frac{\gamma}{2}} SE_{V\hat{a}R_{\alpha}^{new}}, V\hat{a}R_{\alpha}^{new} + z_{1-\frac{\gamma}{2}} SE_{V\hat{a}R_{\alpha}^{new}}\right]. \quad (22)$$

In the percentile approach, the confidence interval is determined directly using cutoff points from the ordinal distribution of VaR estimates. Since this method does not require any distributional assumptions, it is more flexible and robust to out-of-distribution deviations.

First, the VaR estimates are sorted in ascending order:

$$V\hat{a}R_{\alpha,(1)} \leq V\hat{a}R_{\alpha,(2)} \leq \cdots \leq V\hat{a}R_{\alpha,(M)}. \quad (23)$$

Then, the lower and upper bound indices for the confidence level $1 - \gamma$ are determined as $l = \left\lfloor \frac{\gamma}{2} M \right\rfloor$ and$u = \left\lfloor \left(1 - \frac{\gamma}{2}\right) M \right\rfloor$. The 1-γ confidence interval of VaR is constructed as $\left[V\hat{a}R_{\alpha,(l)}, V\hat{a}R_{\alpha,(u)}\right]$. For example, at M = 1000 and 1-γ = 0.95 confidence level, the 1-γ confidence interval of VaR is $\left[V\hat{a}R_{\alpha,(25)}, V\hat{a}R_{\alpha,(975)}\right]$.

5.3. Tail-focused component selection based on backtesting

To determine the optimal number of components in the GMM, this study proposes a new selection strategy that directly serves extreme risk modeling. Instead of information criteria such as AIC and BIC, which evaluate overall fit, this approach focuses on predictive performance in extreme regions.

Let the set of candidate component numbers be $g \in \{2, 3, \dots, G\}$. For each value of g, binomial and Kupiec's tests, which assess the conformity of VaR estimates obtained from our newly designed GMM-based Monte Carlo simulation with the expected number of exceptions, are applied. Additionally, Christoffersen's conditional coverage independence (CCI) test is used to evaluate the independence of violations, while Christoffersen's conditional coverage (CI) test jointly assesses both coverage ratio and independence. In the proposed approach, the optimal number of GMM components is selected as the smallest value of g that passes all selected backtests.

5.4. Stable initialization for GMM: A novel deterministic membership algorithm

One of the major problems encountered in GMM estimations is the instability that arises from randomly initializing cluster memberships [2]. This is particularly common when the k-means algorithm is used to determine initial memberships and leads to non-reproducible results in VaR calculations. To overcome this problem, a deterministic initial cluster membership algorithm is developed, which determines robust and well-separated initial cluster centers for use in the k-means clustering algorithm.

The data matrix $X \in \mathbb{R}^{n \times p}$ contains n observations and p features. The following algorithm steps are applied to determine the initial cluster memberships in the estimation of the GMM parameters of the X data matrix with G components:

Step 1: The smallest and largest values among the observations for each variable are determined. These values are used to create the 1st and kth candidate initial centers:

$$\boldsymbol{C}_1 = \left[min(X_1) \quad min(X_2) \quad \ldots \quad min(X_p)\right],$$

$$\boldsymbol{C}_{\boldsymbol{G}} = \left[max(X_1) \quad max(X_2) \quad \ldots \quad max(X_p)\right].$$

Step 2: The range of each variable is calculated as:

$$Range_j = max(X_j) - min(X_j),$$

where $j = 1, \ldots, p$.

Step 3: The ranges are divided by $G - 1$ to obtain the increment vector h:

$$\boldsymbol{h} = \frac{1}{G-1}[Range_1 \quad Range_2 \quad \ldots \quad Range_p].$$

Step 4: For cluster number $G > 2$ and $g = 3, \ldots, G - 1$, other candidate initial centers are estimated as:

$$\boldsymbol{C}_g = \boldsymbol{C}_{g-1} + \boldsymbol{h}.$$

Step 5: Candidate initial centers are updated.

First, calculate the Euclidean distances of the $\boldsymbol{Y}_i$ vector in the Y data matrix to the candidate initial center C_1 where $Y = X$:

$$d_{\boldsymbol{Y}_i, \boldsymbol{C}_1} = \sqrt{(\boldsymbol{Y}_i - \boldsymbol{C}_1)(\boldsymbol{Y}_i - \boldsymbol{C}_1)'}.$$

By sorting the Euclidean distances from smallest to largest, the q units closest to the first candidate initial center are determined, and the center is updated with the average of these units. After this update, the q units used to determine this candidate initial center are removed from the Y data set to obtain $(n - q) \times p$ dimensional Y* data matrix. This process is done to prevent overlapping and the close proximity of the centers. To update the second candidate initial center, the Euclidean distances of the units in the reduced Y* data matrix to the candidate initial center $\boldsymbol{C}_2$ are calculated. The second candidate initial center is updated with the average of the q units closest to the $\boldsymbol{C}_2$. The size of the data matrix is reduced by removing q units from the Y* matrix to be used in updating the next center. This step is completed by updating all candidate starting centers in a similar manner.

Step 6: The updated candidate initial centers are used as initial cluster centers in clustering the X data matrix using the k-means. Initial cluster memberships in the GMM model are determined by clustering the X data matrix using k-means.

6. Data analysis

In this section, the performance of our proposed Monte Carlo simulation approach based on the GMM model for VaR estimation is compared with historical simulation, parametric method, and the Monte Carlo simulation approach based on multivariate normal distribution. The dataset consisting of daily returns of BTC, ETH, BNB, and SOL cryptocurrencies between January 1, 2021, and March 31, 2025, is used in the study. Binomial, Kupiec's POF, Christoffersen independence (CCI), and Christoffersen interval (CI) tests are performed to test the effectiveness of the proposed method. Descriptive statistics and normality test results for the daily returns of the equally weighted portfolio comprising four cryptocurrencies are presented in Table 1.

Table 1. Descriptive statistics and normality test results for cryptocurrencies and daily portfolio returns.

	BTC	ETH	BNB	SOL	Portfolio
Mean	0.00119	0.00147	0.00282	0.00472	0.00255
Std. deviation	0.03223	0.04182	0.04660	0.06379	0.03904
Skewness	0.14364	0.14063	2.65045	0.84528	-0.03382
Kurtosis	6.42625	8.06823	42.87196	13.44392	9.23386
Minimum	−0.15629	−0.27886	−0.34068	−0.42354	−0.28175
Maximum	0.19414	0.25957	0.69991	0.64858	0.26716
Range	0.35043	0.53843	1.04059	1.07212	0.54891
Q1	−0.01376	−0.01818	−0.01495	−0.02988	−0.01618
Median (Q2)	−0.00018	0.00093	0.00104	0.00040	0.00193
Q3	0.01563	0.02102	0.01769	0.03361	0.02287
IQR	0.02939	0.03921	0.03264	0.06349	0.03905
Shapiro–Wilk	0.95168	0.94314	0.79659	0.91548	0.93878
p- value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Jaqua–Bera	763.49	1664.06	104487.63	7229.04	2510.07
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table 1 shows that all cryptocurrencies have positive daily average returns, with BNB having the highest average return with 0.00282, followed by SOL, ETH, and BTC. The average daily return of the portfolio is 0.00255. This value indicates the more stable returns achieved through diversification.

Volatility is measured by standard deviation, where SOL and BNB have the highest volatility, with standard deviations of 0.06379 and 0.04660, respectively. BTC has the lowest volatility with a standard deviation of 0.03223, while the portfolio has lower volatility than most of the individual assets, with a standard deviation of 0.03904. This demonstrates the risk-reducing effect of portfolio diversification.

The kurtosis and skewness measures represent the distributional characteristics of the returns. The skewness values of BTC and ETH are 0.14364 and 0.14063, respectively, demonstrating that the distribution of the assets' daily returns is close to a symmetric distribution. In contrast, the returns of BNB and SOL have positive skewness, indicating a higher probability of extreme positive returns. The skewness of the portfolio is close to zero and negative, showing that the distribution of portfolio returns is close to symmetry.

Kurtosis values are greater than 3 for the portfolio, and the crypto assets that make up the portfolio indicate that the distributions are leptokurtic with heavy tails. In particular, the kurtosis value of BNB is quite high, pointing out that the frequency of outliers is much higher than a normal distribution. The portfolio has a very high kurtosis value of 9.23386, showing that extreme outliers are a significant risk factor.

Minimum and maximum daily returns provide information on the magnitude of outliers. BNB and SOL have the widest ranges of 1.04059 and 1.07212, respectively. In terms of the spread of the central tendency, the IQR is highest for SOL, while the IQR of the portfolio is lower due to diversification.

The normality of returns is tested by the Shapiro–Wilk and Jarque–Bera tests. The Shapiro–Wilk and Jarque–Bera test statistics for all series are quite high, and all p-values are <0.0001. These results suggest that crypto asset returns are not normally distributed and have heavy tails and skewness.

The insights from Table 1 show that the crypto assets of the study have the potential for high returns but also carry serious risk factors, such as high volatility, skewness, and kurtosis.

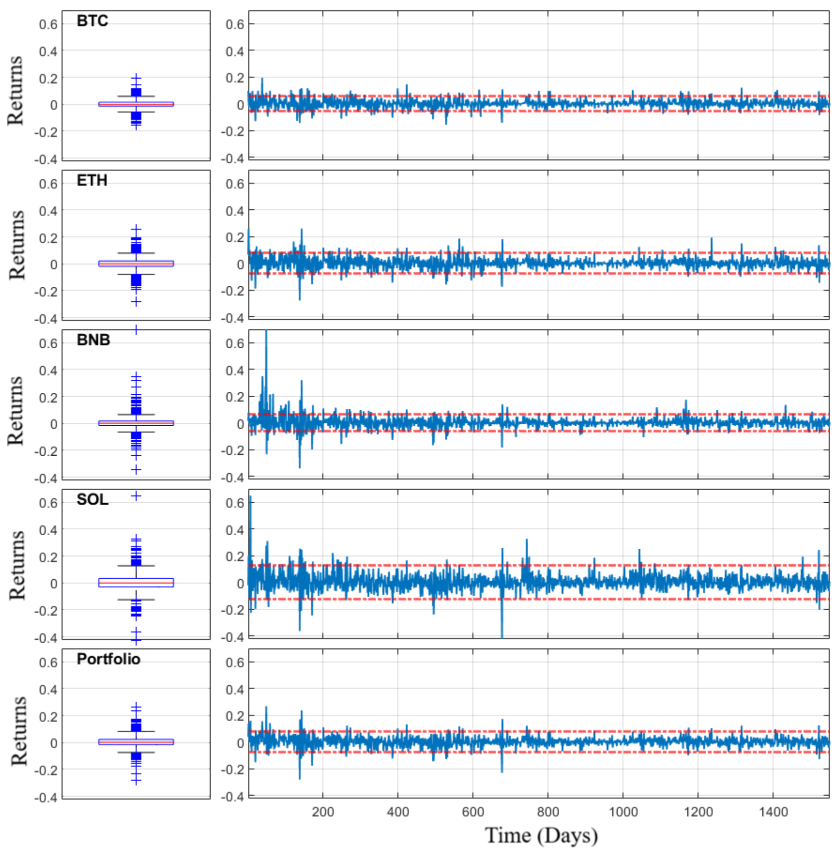

Figure 2. Time series and box plots of crypto assets and portfolio returns.

The time series and box plots of the daily returns of crypto assets and the portfolio are shown in Figure 2. In Figure 2, daily returns of the crypto assets and their portfolio are plotted comparatively with both their distributional and temporal characteristics. For each asset, the left panel shows the boxplots of daily returns, and the right panel shows the time series of the same returns. The red dashed lines added to the time series plots represent the upper and lower bounds of outliers calculated based on the box plots.

Statistical properties of return distributions differ across assets. BTC and ETH exhibit a more stable return behavior with symmetric structures and fewer outlier observations. In contrast, BNB and SOL have longer tails, indicating excessive losses in certain periods. On the other hand, the equal-weighted portfolio has both a narrower return distribution and fewer outliers over time, demonstrating that the diversification strategy reduces both volatility and the frequency of extreme return events.

In the literature, correlation matrices play a critical role in portfolio analysis. According to Markowitz's modern portfolio theory, reducing total risk through portfolio diversification is only possible by using assets with low correlation together. Since highly correlated assets rise and fall together, they may be insufficient to reduce the systematic risk of the portfolio [7].

The correlation matrix between the daily returns of the crypto assets in the portfolio is given in Figure 3. Figure 3 shows both the distributional characteristics of the variables and the correlations between them in a comprehensive way. It shows a high positive correlation of 0.81 between BTC and ETH, suggesting that these two assets are highly sensitive to similar market conditions and often move together. Crypto assets are moderately to highly correlated in general, showing that the portfolio diversification effect may be limited. Considering that correlations may change over time and the tendency to move together may increase during periods of market crises, such analyses should be supported by dynamic models.

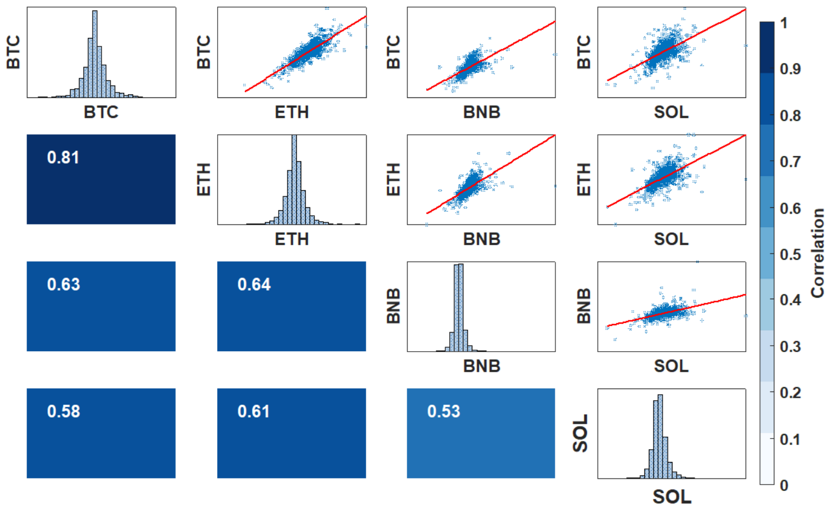

Figure 3. Correlation matrix graph for daily returns of crypto assets.

In the literature, VaR estimates are generally evaluated based on data obtained using a 250-day rolling window approach [26]. Therefore, the comparison of the proposed method with classical VaR estimation approaches is based on rolling windows with 250 observations. In the rolling window approach, the oldest observation of the previous 250-unit data set is removed, and the data is updated by adding the next day's return. In this study, 1300 rolling windows created through this process are utilized for parameter estimation and simulation. Figure 4 illustrates the 250-day rolling mean returns, the overall average return, and the corresponding confidence intervals.

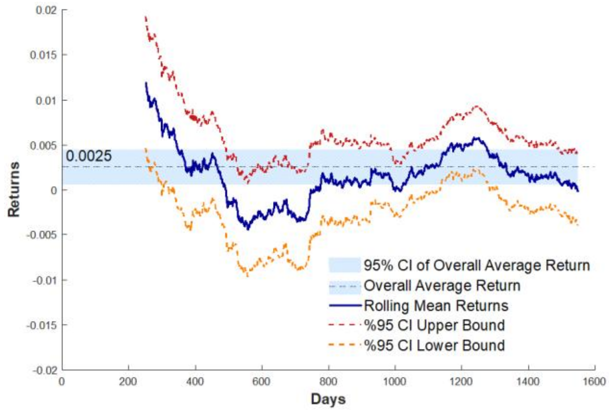

Figure 4. 250-day average daily portfolio returns in the rolling window approach.

Downward movements in the moving average return indicate that the newly added returns are lower than the oldest falling returns. This denotes a weakening in recent performance relative to previous periods when this happens consistently and significantly. Between the 251st and 550th days, the moving average falls steadily. This period covers approximately from 09.2021 to 07.2022. On the other hand, upward movements indicate that recent returns have been stronger than in previous periods, and portfolio performance has improved. This can be seen between the 750th and 1250th days, a period covering approximately from 01.2023 to 06.2024. Most of the moving averages remain within the confidence interval of the overall average, which means that the deviations are not statistically significant.

The assumption of normality is especially critical in parametric risk measurement methods. Therefore, monitoring the change in JB test statistics over time is very important for model fit and reliability analysis. Figure 5 shows the change in JB normality test values in the rolling window approach.

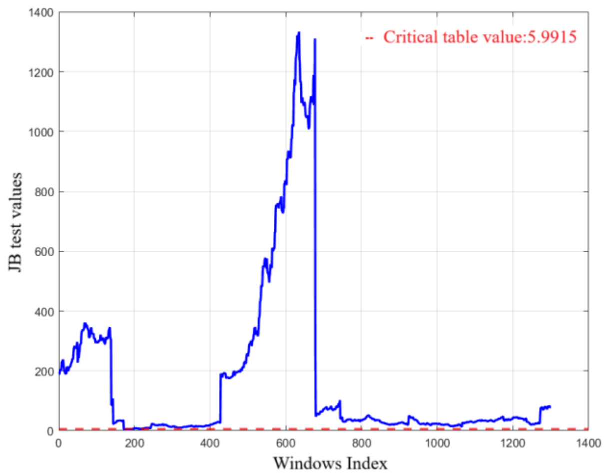

Figure 5. Changes in Jarque–Bera normality test values in the rolling window approach.

JB test statistics are above the critical value for many windows. This demonstrates that the assumption of normality is frequently violated, and the data distributions are heavily skewed and/or kurtotic. There is a significant increase in the JB test values between 13.10.2022 and 09.08.2023, which corresponds to the windows in the 400–700 range, indicating significant distortions in the return distribution. Around the 700th window, JB test statistics approach the critical value level and remain relatively stable in the following periods. This suggests that data distribution becomes closer to normality, and more reliable results can be obtained for statistical modeling in these windows.

Table 2 presents the backtest statistics obtained for the number of components $g = 2,3,\ldots,G$ with the proposed approach at a 0.95 confidence level on the daily returns of crypto assets in the portfolio. $q = 20$ is taken in determining the initial cluster centers, and the number of repetitions in the simulation design is $M = 1000$. The values in parentheses are the p-values of the test statistics.

Considering that the expected number of exceptions at the 95% confidence level is theoretically 65, Table 2 shows that the number of exceptions observed in the 8-component model corresponds exactly to this value. This situation reveals that the model predictions are in line with the statistical significance level at a first glance.

Table 2. Selected backtest statistics for VaR95 estimates at different component numbers in the proposed method.

g	Exception	Binomial	Kupiec's POF	N00	N01	N10	N11	CCI	CI
2	69	0.5090	0.2542	1170	60	60	9	6.4211	6.6753
		(0.6107)	(0.6141)					(0.0113)	(0.0355)
3	70	0.6363	0.3954	1169	60	60	10	8.2863	8.6817
		(0.5246)	(0.5295)					(0.0040)	(0.0130)
4	70	0.6363	0.3954	1168	61	61	9	6.0682	6.4636
		(0.5246)	(0.5295)					(0.0138)	(0.0395)
5	66	0.1273	0.0161	1176	57	57	9	7.5546	7.5707
		(0.8987)	(0.8990)					(0.0060)	(0.0227)
6	67	0.2545	0.0642	1173	59	59	8	5.0256	5.0898
		(0.7991)	(0.8000)					(0.0250)	(0.0785)
7	67	0.2545	0.0642	1173	59	59	8	5.0256	5.0898
		(0.7991)	(0.8000)					(0.0250)	(0.0785)
8	**65**	**0.0000**	**0.0000**	**1176**	**58**	**58**	**7**	**3.7121**	**3.7121**
		(1.0000)	**(1.0000)**					**(0.0540)**	**(0.1563)**
9	66	0.1273	0.0161	1174	59	59	7	3.4544	3.4705
		(0.8987)	(0.8990)					(0.0631)	(0.1764)

The p-values of all test statistics for the 8-component model are greater than 0.05. In particular, Kupiec's POF and binomial test p-values strongly support the validity of the model, both proportionally and probabilistically. Moreover, in the independence test, the CCI value is 3.7121 with a p-value of 0.0540, which is at the significance threshold but still not rejected at the 5% level. This indicates that VaR exceptions satisfy the independence assumption. Since the CCI p-value is close to the significance threshold of 0.054, the 9-component model was also tested, but no significant improvement in model performance was observed despite using a higher number of components.

According to the proposed method, the GMM model with $g = 8$ is accepted as a valid and reliable VaR estimation model, since all p-values for backtest results are greater than the 5% significance level. Therefore, this model structure, which provides the highest accuracy with the lowest number of components, is recommended, with 8 as the optimal number of components.

In order to compare the performance of the proposed method, historical simulation, parametric method, and Monte Carlo simulation method based on multivariate normal distribution, which are widely used in the literature, are considered. To evaluate the effectiveness of the proposed initial cluster membership algorithm, its performance is also compared with the widely recognized k-means++ algorithm. The statistical results of the backtest performance of the VaR estimates of the proposed and selected methods at 95% confidence level are given in Table 3.

Table 3. Comparison of VaR95 backtest results of the proposed method and selected methods.

	Exception	Binomial	Kupiec's POF	N00	N01	N10	N11	CCI	CI
Historical simulation	68	0.3818 (0.7026)	0.1437 (0.7047)	1172	59	59	9	6.7863 (0.0092)	6.9300 (0.0313)
Parametric	62	−0.3818 (0.7026)	0.1479 (0.7005)	1183	54	54	8	6.7387 (0.0094)	6.8866 (0.0320)
Monte Carlo simulation	63	−0.2545 (0.7991)	0.0654 (0.7981)	1181	55	55	8	6.3711 (0.0116)	6.4365 (0.0400)
Proposed method with k-means++ (g = 8)	67	0.2545 (0.7991)	0.0642 (0.8000)	1173	59	59	8	5.0257 (0.0250)	5.0898 (0.0785)
Proposed method (g = 8)	**65**	**0.0000 (1.0000)**	**0.0000 (1.0000)**	**1176**	**58**	**58**	**7**	**3.7121 (0.0540)**	**3.7121 (0.1563)**

It is seen that the historical simulation approach remains above the expected number of exceptions (65) with 68. A higher-than-expected number of exceptions indicates that the model systematically underestimates the financial risk; in other words, it is overly optimistic. On the other hand, parametric and Monte Carlo simulation methods were below the expected number of exceptions with 62 and 63, respectively. This suggests that these methods overestimate the risk, exhibiting an overly conservative approach.

Although the selected methods produce either overly optimistic or overly cautious estimates, the binomial and Kupiec's POF tests reveal that the coverage ratios are statistically valid, as the p-values are above 0.05. However, the CCI and CI statistics for the independence tests indicate that the selected methods fail to satisfy the independence assumption for VaR exceptions. This points out that the methods do not adequately take into account the temporary structural changes in the market or temporal dependencies in the data.

In this perspective, the proposed method stands out as the only method that performs well not only in terms of the number of exceptions but also in all key backtest criteria, such as coverage ratio and independence of exceptions. The p-values of all test statistics are above the 5% significance level, indicating that the proposed model generates both valid and reliable VaR estimates and satisfies the independence assumption. Also, the proposed initial cluster membership algorithm outperforms the conventional k-means++ algorithm. In this regard, the proposed model exhibits more stable and robust performance than traditional models and algorithms and enables more reliable risk predictions, especially in dynamic and volatile market conditions like cryptocurrency markets.

Backtest statistics for VaR estimates corresponding to the 99% confidence level for different component numbers for the proposed method are presented in Table 4. The theoretically expected number of exceptions at this confidence level is 13.

Table 4. Selected backtest statistics for VaR99 estimates at different component numbers in the proposed method.

g	Exception	Binomial	Kupiec's POF	N00	N01	N10	N11	CCI	CI
2	24	3.0662	7.5233	1253	22	22	2	3.1247	10.6480
		(0.0022)	(0.0061)					(0.0771)	(0.0049)
3	**13**	**0.0000**	**0.0000**	**1273**	**13**	**13**	**0**	**0.2628**	**0.2628**
		(1.0000)	**(1.0000)**					**(0.6082)**	**(0.8769)**
4	11	−0.5575	0.3279	1277	11	11	0	0.1879	0.5158
		(0.5772)	(0.5669)					(0.6647)	(0.7727)

Table 4 shows that the number of exceptions in the 3-component model is exactly 13. Both the binomial and Kupiec's POF tests statistics confirm the coverage ratio of the model. The p-values of the CCI and CI statistics are 0.6082 and 0.8769, respectively, which are above the 0.05 significance level. This strengthens the independence of VaR exceptions and the reliability of VaR estimations. As a result, it is observed that 3 is the optimal number of components for VaR estimation at a 99% confidence level. This model not only satisfies the theoretically expected value for the number of exceptions but also meets the validity and independence criteria in all statistical tests.

VaR estimations at the 99% confidence level show that the most appropriate number of components in the GMM model is determined to be 3 with our proposed method. To investigate the common assumption that increasing the number of components in the model would improve the prediction performance automatically, results obtained for the 4-component model are shown. The backtest performance of the 4-component model is lower than that of the 3-component model, especially in terms of the number of exceptions and coverage ratios. This shows that increasing the number of components does not always result in more reliable VaR estimations. Thus, the proposed method fills an important gap in the literature on how excessive use of components can reduce model efficiency and emphasizes the need to take empirical performance into account when choosing the number of components.

In order to evaluate the effectiveness of the proposed method at a 99% confidence level, backtest analysis is performed for VaR estimates obtained by historical simulation, the parametric method, and the Monte Carlo simulation method based on a multivariate normal distribution. To evaluate the effectiveness of the proposed initial cluster membership algorithm, its performance is also compared with the widely recognized k-means++ algorithm. The results of the backtest performances of the methods are given in Table 5.

According to Table 5, although the historical simulation method is slightly below the theoretical expectation with 11 exceptions, the p-values of the binomial and Kupiec's POF tests are above the 5% significance level. Therefore, this method is considered valid in terms of coverage ratio. Moreover, the CCI and CI p-values for the independence tests also exceed the significance level, confirming that the historical simulation method also satisfies the independence assumption. On the other hand, the parametric method and the Monte Carlo simulation method based on the multivariate normal distribution fail with 23 exceptions and p-values below the 5% significance level in both binomial and Kupiec's tests. This clearly demonstrates that both methods fail to achieve the coverage ratio. The VaR estimates of the proposed approach fully satisfy the theoretical expectation with 13 exceptions and perform statistically strongest in terms of both coverage ratio and independence, with p-values greater than 0.05 in all test statistics (binomial, Kupiec's POF, CCI, CI). Also, the proposed initial cluster membership algorithm outperforms the conventional k-means++ algorithm. These results show that the proposed approach is not only valid but also produces the most consistent and reliable VaR estimates. This suggests that the proposed method can be considered as a strong alternative in financial applications where the accuracy of risk estimates is critical, especially at a high confidence level.

Table 5. Comparison of VaR99 backtest results of the proposed method and selected methods.

	Exception	Binomial	Kupiec's POF	N00	N01	N10	N11	CCI	CI
Historical simulation	11	−0.5575 (0.5772)	0.3279 (0.5669)	1277	11	11	0	0.1879 (0.6647)	0.5158 (0.7727)
Parametric	23	2.7875 (0.0053)	6.3230 (0.0119)	1255	21	21	2	3.4126 (0.0647)	9.7355 (0.0077)
Monte Carlo simulation	23	2.7875 (0.0053)	6.3230 (0.0119)	1255	21	21	2	3.4126 (0.0647)	9.7355 (0.0077)
Proposed method with k means++ (g = 3)	15	0.5575 (0.5772)	0.2961 (0.5863)	1269	15	15	0	0.3508 (0.5538)	0.6466 (0.7238)
Proposed method (g = 3)	**13**	**0.0000 (1.0000)**	**0.0000 (1.0000)**	**1273**	**13**	**13**	**0**	**0.2628 (0.6082)**	**0.2628 (0.8769)**

Figure 6 graphically evaluates the performance of the VaR estimates obtained by the proposed method in terms of both the frequency and independence of the exceptions. The graph shows the daily returns of the portfolio (blue line), the VaR values estimated by the proposed method at the relevant confidence level (red line), and the exceptions (black "x" sign).

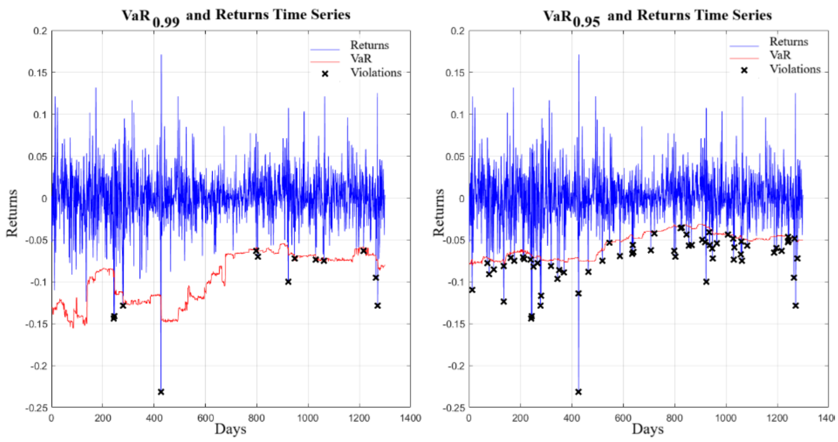

Figure 6. Daily portfolio returns and VaR estimates obtained by the proposed method at different confidence levels and exception points.

The results obtained at the 99% confidence level in the left panel show that the VaR bound is quite cautious and located at the lower bound of the portfolio returns. At this level, exceptions are rare and periodic, suggesting that the proposed method works well at a high confidence level and is able to accurately isolate risky days.

In the right panel, more exceptions are observed, and the distribution of exceptions over time is statistically close to random at the 95% confidence level. This indicates that the proposed method performs well not only in terms of coverage ratio but also in terms of independence of exceptions.

When both panels are considered together, they show that the proposed method generates stable and reliable VaR estimates in terms of both coverage ratio and independence at different confidence levels. The models can control the number of exceptions in accordance with the desired confidence level and effectively flag extremely risky situations. These results support that the method can be used as a powerful tool for managing market risk.

Another important advantage of the proposed method is that it provides not only point estimates but also confidence intervals for VaR, offering a more comprehensive assessment by accounting for the uncertainty inherent in risk estimation. Figure 7 presents the VaR estimates, their 95% confidence intervals, and the portfolio returns obtained using the proposed method.

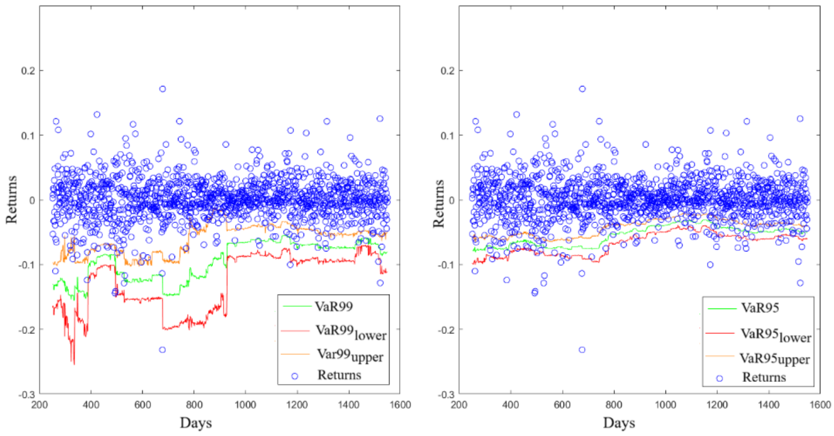

Figure 7. VaR estimates at different confidence levels and 95% confidence intervals of VaR estimates.

Figure 7 shows that the VaR estimates generally adequately cover negative portfolio returns. VaR estimates at the 99% confidence level cover more extreme negative values of portfolio returns, while estimates at the 95% confidence level provide a narrower band and tend to explain more frequent moderate losses. This suggests that higher confidence levels provide more conservative risk estimates with wider confidence intervals.

When both panels are evaluated together, the proposed GMM-based Monte Carlo approach not only produces VaR estimates that are appropriate for the confidence level but also expresses the uncertainty of these estimates statistically. This indicates that the method is well-suited for risk management applications, enabling more informed decision-making. The confidence intervals increase the effectiveness of decision support systems by providing not only a single probabilistic boundary but also the potential deviation area.

In our proposed method, the appropriate number of components for VaR estimation at 95% confidence level is determined to be 8, while the appropriate number of components at 99% confidence level is determined to be 3. Figure 8 shows the appropriate number of components if AIC and BIC information criteria are used in the model selection of the number of components in the GMM model in accordance with the common usage in the literature.

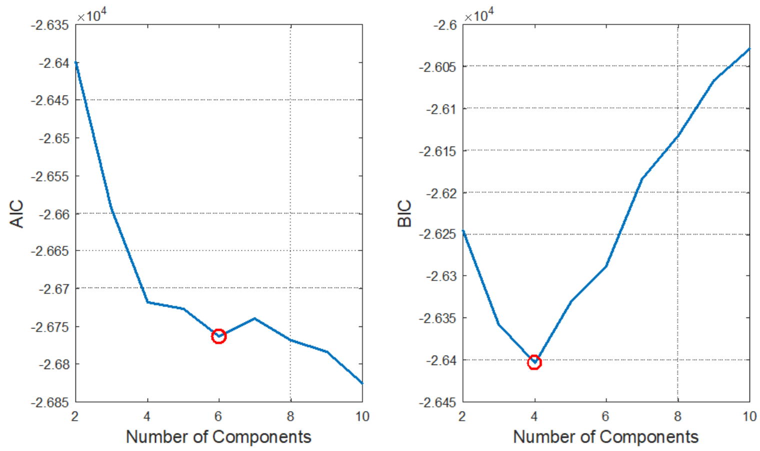

Figure 8. Optimal number of components based on AIC and BIC.

According to the AIC criterion in Figure 8, the number of components with the first local minimum value of 6 is determined as the optimal number of components, while the number of components with the smallest BIC value of 4 is determined as the optimal number of components according to the BIC criterion. The backtest results in Tables 2 and 4 show that the models based on the optimal number of components do not yield the best performance. In Table 2, the backtest results obtained for the numbers of 4 and 6 components for the 95% confidence level do not show statistical validity in terms of both the number of exceptions and independence tests. Table 4 shows that at a 99% confidence level, the appropriate number of components (3) selected by the proposed method outperforms the appropriate number of components (4) suggested by the BIC criterion. These findings suggest that relying solely on information criteria for model selection may be insufficient, particularly in contexts where direct evaluation of predictive performance is crucial, such as financial risk measurement.

Our proposed method, which extends beyond model selection based on traditional information criteria and bases the selection of the number of components on backtest performance, improves the reliability and applicability of the model to real market conditions as it directly focuses on the empirical performance of VaR estimates. This contribution enables more reliable risk prediction, especially in financial systems with high volatility, such as cryptocurrency markets, and emphasizes the importance of empirical validation in model selection.

Another important limitation of model selection based on information criteria in determining the appropriate number of components in the GMM model is that the VaR estimation performance of the same number of components at different confidence levels may be inconsistent. In the analysis performed, when the backtest performances of VaR estimations obtained at two different confidence levels, such as 95% and 99%, are compared, it is clear that the best results are not achieved with the same number of components at both levels. This suggests that although the information criteria are focused on the overall fit of the model, they do not take into account the sensitivity of the prediction to the risk level. In other words, the number of components selected by the information criteria does not guarantee the most accurate and valid VaR estimates corresponding to different confidence levels in all cases. Therefore, information criteria alone may not be sufficient for model selection, especially in financial risk measurement.

7. Conclusions and suggestions

7.1. Conclusions

In this study, a new and improved simulation design is proposed by considering the limitations of the traditional Monte Carlo simulation method based on the Gaussian mixture model (GMM) for VaR estimation. The proposed approach relies on backtesting performance to determine the number of components in the GMM directly instead of the overall model fit, thus aiming to estimate extreme risks more accurately. Moreover, the initial values in both the simulation design and the EM algorithm are determined more stably to reduce the effect of randomness in the estimation of GMM parameters.

The effectiveness of the method is tested in the cryptocurrency market, which is characterized by high volatility and sharp ups and downs. The performance of the proposed method is compared with the historical simulation method, the parametric method, and the Monte Carlo method based on a multivariate normal distribution. The findings show that the proposed method improves the modeling performance, especially in tails, and provides more reliable VaR estimates. Furthermore, the proposed initialization clustering algorithm minimized the impact of stochasticity (randomness) in the GMM parameter estimation process, leading to the most accurate VaR95 and VaR99 calculations. When compared to the widely adopted k-means++ algorithm, the effectiveness of this approach was validated, demonstrating that the proposed method offers a more robust performance.

One of the most important findings of the study is that using backtesting instead of general information criteria, such as AIC or BIC, when determining the number of components is a more accurate approach for modeling extreme risks.

7.2. Limitations and suggestions for future research

Despite the improved accuracy and stability of the proposed GMM-based Monte Carlo framework, this study acknowledges certain limitations that warrant consideration. Primarily, the integration of deterministic initialization and a backtest-driven component selection process introduces an increased computational burden compared to traditional parametric methods, particularly when generating $M = 1000$ synthetic samples for each rolling window. Furthermore, while the 250-day window employed in this analysis aligns with standard literature, the model's predictive performance may exhibit sensitivity to different window sizes, especially in the presence of structural breaks inherent in volatile cryptocurrency markets. Lastly, although the framework effectively captures the distribution of returns, it does not explicitly incorporate long-term memory effects or complex volatility clustering beyond the dynamics inherent in the rolling window approach, suggesting areas where the model's temporal granularity could be further refined.

In the estimation of GMM parameters, robust estimation methods using time-varying weights can be applied instead of classical maximum likelihood approaches to improve the stability of the model. In future studies, the proposed method can be tested under different market conditions to assess its robustness or be extended to different types of risk analysis by integrating alternative risk measures, such as conditional VaR (CVaR). The proposed structure can be integrated into portfolio optimization models to contribute to the risk-return balance of investment strategies.

Author contributions

Ülkü Erişoğlu: Conceptualization, writing-original draft, methodology; Selim Gündüz: Investigation, validation, application, funding acquisition; Mert Yaman: Editing, software, application, formal analysis. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have used Large Language Models (LLMs) in the creation of this article. AI tools were only employed to refine the grammar and structure of the text. The authors checked and agreed with all the changes made by the AI.

Conflict of interest

The authors declare that there is no conflict of interest in this paper.

References

1. S. Akogul, M. Erisoglu, An approach for determining the number of clusters in a model-based cluster analysis, *Entropy*, **19** (2017), 452. https://doi.org/10.3390/e19090452
2. C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models, *Comput. Stat. Data Anal.*, **41** (2003), 561–575. https://doi.org/10.1016/S0167-9473(02)00163-9
3. Q. Chen, Z. Huang, F. Liang, Forecasting volatility and value-at-risk for cryptocurrency using GARCH-type models: the role of probability distribution, *Appl. Econ. Lett.*, **31** (2024), 1907–1914. https://doi.org/10.1080/13504851.2023.2208824
4. P. F. Christoffersen, Evaluating interval forecasts, *Int. Econ. Rev.*, **39** (1998), 841–862. https://doi.org/10.2307/2527341
5. R. J. Contreras, Market risk measures using finite gaussian mixtures, *J. Appl. Financ. Bank.*, **4** (2014), 29–45
6. J. Danielsson, *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*. John Wiley & Sons, 2012.
7. E. J. Elton, M. J. Gruber, S. J. Brown, W. N. Goetzmann, *Modern portfolio theory and investment analysis*, 9 Eds., Wiley, 2014.
8. U. Erisoglu, Y. Koroglu, Value at risk for risk management with normal-logDagum mixture distribution, *J. Stati. Manag. Syst.*, **24** (2021), 1489–1501. https://doi.org/10.1080/09720510.2020.1826630
9. K. Gkillas, F. Longin, Is Bitcoin the new digital gold? Evidence from extreme price movements in financial markets, 2019. http://doi.org/10.2139/ssrn.3245571

10. P. Glasserman, P. Heidelberger, P. Shahabuddin, Efficient Monte Carlo methods for value-at-risk, In: *Mastering risk: Application*, 3Eds., New York: Financial Times/Prentice Hall, 2001.
11. G. A. Holton, *Value-at-risk: Theory and practice*, 2Eds., Academic Press, 2014.
12. P. Hrytsiuk, T. Babych, L. Bachyshyna, Cryptocurrency portfolio optimization using Value-at-Risk measure. In: *Proceedings of the 6th International Conference on Strategies, Models and Technologies of Economic Systems Management (SMTESM 2019)*, 2019, 385–389. https://doi.org/10.2991/smtesm-19.2019.75
13. K. Jiang, L. Zeng, J. Song, Y. Liu, Forecasting Value-at-Risk of cryptocurrencies using the time-varying mixture-accelerating generalized autoregressive score model, *Res. Int. Bus. Financ.*, **61** (2022), 101634. https://doi.org/10.1016/j.ribaf.2022.101634
14. P. Jorion, *Value at Risk*, 2Eds., McGraw-Hill, New York. 2001.
15. P. Kupiec, Techniques for Verifying the Accuracy of Risk Management Models. *J. Deriv.*, **3** (1995), 73–84. https://doi.org/10.3905/jod.1995.407942
16. M. F. Leung, J. Wang, Minimax and biobjective portfolio selection based on collaborative neurodynamic optimization, *IEEE T. Neur. Net. Lear.*, **32** (2021), 2825–2836. https://doi.org/10.1109/TNNLS.2019.2957105
17. Y. Li, J. Wu, The estimation of VaR and ES based on Gaussian mixture model–Take the Shanghai Composite Index as an example, In: *2022 22nd International Conference on Control, Automation and Systems (ICCAS)*, 2022, 1497–1502. https://doi.org/10.23919/ICCAS55662.2022.10003863
18. D. Likitratcharoen, P. Chudasring, C. Pinmanee, K. Wiwattanalamphong, The efficiency of value-at-risk models during extreme market stress in cryptocurrencies, *Sustainability*, **15** (2023), 4395. https://doi.org/10.3390/su15054395
19. T. J. Linsmeier, N. D. Pearson, Risk measurement: An introduction to value at risk, 1996. https://doi.org/10.22004/ag.econ.14796
20. J. Mallela, M. Gupta, Value at risk performance and backtesting of cryptocurrencies. *Manag. Quest*, **6** (2024), 51–73.
21. G. McLachlan, D. Peel, *Finite mixture models*. John Wiley & Sons, 2000. https://doi.org/10.1002/0471721182
22. I. Morkūnaitė, D. Celov, R. Leipus, Evaluation of Value-at-Risk (VaR) using the Gaussian mixture models, *Res. Stat.*, **2** (2024), 2346075. https://doi.org/10.1080/27684520.2024.2346075
23. F. M. Müller, M. B. Righi, Comparison of value at risk (VaR) multivariate forecast models, *Comput. Econ.*, **63** (2024), 75–110. https://doi.org/10.1007/s10614-022-10330-x
24. S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, 2008.
25. A. Panda, S. G. Deb, Are VaR models effective in capturing downside risk in alternative investment funds? Insights from a cross-country study, *Int. J. Financ. Eng.*, **11** (2024), 2350044. https://doi.org/10.1142/S2424786323500445
26. V. Piparo, Backtesting expected shortfall: An application in cryptocurrencies, 2025.
27. S. M. S. Seyfi, A. Sharifi, H. Arian, Portfolio Value-at-Risk and expected-shortfall using an efficient simulation approach based on Gaussian Mixture Model, *Math. Comput. Simulat.*, **190** (2021), 1056–1079. https://doi.org/10.1016/j.matcom.2021.05.029
28. B. Silahli, K. D. Dingec, A. Cifter, N. Aydin, Portfolio value-at-risk with two-sided Weibull distribution: Evidence from cryptocurrency markets, *Financ. Res. Lett.*, **38** (2021), 101425. https://doi.org/10.1016/j.frl.2019.101425

29. K. Tan, M. Chu, Estimation of portfolio return and value at risk using a class of Gaussian mixture distributions, *Int. J. Bus. Financ. Res.*, **6** (2012), 97–107.
30. S. Vahabzadeh, S. S. Haghshenas, S. J. Ghoushchi, G. Guido, V. Simic, D. Marinkovic, A new framework for risk assessment of road transportation of hazardous substances, *Facta Univ.-Ser. Mech.*, **23** (2025), 479–509. https://doi.org/10.22190/FUME240801007V
31. Y. Zhang, S. Nadarajah, A review of backtesting for value at risk, *Commun. Stat.-Theor. M.*, **47** (2018), 3616–3639. https://doi.org/10.1080/03610926.2017.1361984