Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2175

Research Article

Generalizing univariate empirical goodness-of-fit tests to multivariate normality

Buğra Kaan TİRYAKİ[1,*]

[1]*Department of Computer Science, Faculty of Science, Karadeniz Technical University, Trabzon, 61080, Türkiye*

ARTICLE INFO

Article history
Received: 18 February 2025
Revised: 01 June 2025
Accepted: 11 August 2025

Keywords:
Empirical Goodness-of-Fit Test; Multivariate Empirical Distribution; Multivariate Normality Test; Univariate Normality Test

ABSTRACT

Assessing multivariate normality is a critical prerequisite in numerous statistical analyses and machine learning applications, as many inference procedures assume multivariate normality. However, traditional tests for multivariate normality often suffer from low interpretability, limited scalability in higher dimensions, and computational inefficiencies, making them less practical for modern data-driven research. This study aims to address these limitations by developing a generalized and unified framework that systematically extends several well-established univariate empirical goodness-of-fit tests, including the Kolmogorov–Smirnov, Cramér–von Mises, Anderson–Darling, Kuiper, and Watson tests, into the multivariate context. The proposed methodology involves a variable transformation step using Rosenblatt's transformation to ensure independence among dimensions, followed by the calculation of empirical and theoretical distribution values to construct multivariate test statistics. Comprehensive simulation studies were conducted to evaluate the performance of the proposed tests under various scenarios, including different sample sizes, dimensionalities, and covariance structures. The results show that the proposed tests preserve Type I error rates near the nominal 5 percent level and exhibit strong statistical power, with the multivariate Anderson–Darling test performing particularly well in small-sample settings. Additionally, critical value tables were constructed, and their accuracy was confirmed with average absolute percentage errors consistently below five percent. The applicability of the proposed framework was further demonstrated using real-world biometric and environmental datasets. Overall, this study provides a theoretically sound, computationally efficient, and interpretable alternative to existing methods for assessing multivariate normality, offering practical value for researchers dealing with moderate to high-dimensional multivariate data.

Cite this article as: Tiryaki BK. Generalizing univariate empirical goodness-of-fit tests to multivariate normality. Sigma J Eng Nat Sci 2026;44(5):3543–3566.

INTRODUCTION

Goodness-of-fit tests play a vital role in statistical science. Before applying many statistical procedures, it is essential to assess whether the data follow a normal distribution, as this determines the suitability of the test. In particular, the assumption of normality is a key factor in selecting the appropriate statistical method for research [1].

***Corresponding author.**
*E-mail address: btiryaki@ktu.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

The above-mentioned assumption forms the basis of many different statistical techniques like multivariate analysis of variance, discriminant analysis, canonical correlation analysis, factor analysis, regression, and t-test [2]. All these tests provide valid results only if the underlying data is normally distributed. Hence, it is important to test the normality assumption while conducting any statistical test.

Various goodness-of-fit tests for univariate data have been developed to test the normality assumption, among which the popular ones include Chi-square [3], Kolmogorov-Smirnov [4, 5], Lilliefors, Anderson-Darling [6], Cramér–von Mises [7-9], Watson [10, 11], and Kuiper [12, 13] tests. In the multivariate context, several approaches have also been proposed, such as Mardia's test [14], Royston's test [15], and the Henze–Zirkler test [16]. However, the majority of empirical goodness-of-fit methods in the literature remain limited to the univariate setting or to very low-dimensional cases, typically up to three variables. Although some studies have attempted to extend univariate tests to the multivariate domain, these efforts are often restricted to bivariate data and fail to scale effectively to higher dimensions due to computational complexity. Furthermore, to date, no study has systematically and comprehensively adapted a broad set of univariate empirical goodness-of-fit tests for evaluating multivariate normality.

This gap forms the primary motivation for the present study, which aims to address the lack of easily applicable multivariate versions of empirical goodness-of-fit tests, particularly for high-dimensional data. In practical applications, problems are often multivariate, and applying univariate normality tests to each variable separately can lead to misleading conclusions. Therefore, a multivariate normality test should be conducted before analyzing a multivariate dataset. In this study, several well-known empirical goodness-of-fit tests are adapted to evaluate multivariate normality effectively.

Mecklin and Mundfrom classified multivariate normality tests in the literature into four main categories [17]: (i) methods based on graphical and correlational approaches, (ii) methods derived by adapting univariate goodness-of-fit tests, (iii) methods using skewness and kurtosis coefficients, and (iv) methods based on consistent statistical approaches. Techniques that do not fall into these categories are typically classified as other or miscellaneous methods.

The Q–Q plot, the most commonly used graphical technique, offers a visual understanding of how close the data follows the normal distribution. This plot involves plotting the ordered data against the theoretical z-value corresponding to each one of them. The data is said to be normally distributed when the points on the plot fall roughly on the line y=x. Graphical techniques such as P–P plots, histograms, and boxplots can also be used to check the normality of the data. Several researchers have modified the above methods to test multivariate normality [18-20].

Studies that use modifications of univariate tests of goodness-of-fit in the multivariate case belong to methods of adaptation of empirical goodness-of-fit tests. A lot of research aimed at solving the problem of multivariate normality have used adaptations of the empirical distribution function [21-26].

Measures of skewness and kurtosis help to understand the peculiarities of a distribution. Studies that use these measures are included into the methods based on skewness and kurtosis. Nevertheless, these measures cannot be used in all cases to find out whether the distribution is normal. There are a number of ways of testing multivariate normality based on skewness and kurtosis [27-33].

The empirical characteristic function approach methods are studied in this category. Methods that are based on this idea use characteristic function of normal distribution, especially when multivariate normality is tested. There are a lot of researches based on [16, 34-38].

The methods which do not fall into any of the four basic classes are covered in the class known as other methods. For example, one of the commonly applied methods in multivariate normality tests consists of treating the components of the multivariate sample as univariate variables and applying the corresponding normality test. The conclusion that the whole sample is not multivariate normally distributed should be made if any component has a non-normal distribution. Nevertheless, it should be stressed that for multivariate distribution to be normally distributed all its univariate marginals should also have normal distribution. Therefore, using this method alone might result in wrong conclusions [2, 39]. Moreover, there are special methods developed for particular cases.

The current research aims at developing multivariate versions of univariate goodness-of-fit test statistics in order to test for multivariate normality. The focus is on testing procedures that quantify the difference between empirical and theoretical cumulative distribution functions. Classical tests have been extended to be able to analyze multivariate data, thus allowing us to test for multivariate normality. The power of the suggested approaches is compared, and real data sets are used for validation.

While several earlier studies have attempted to adapt univariate goodness-of-fit tests to the multivariate setting, they are often limited to low-dimensional data (typically $d \leq 2$ or 3), lack generalizability, or present test statistics that are computationally demanding in higher dimensions. Moreover, no study has systematically extended a broad set of univariate empirical tests such as Kolmogorov–Smirnov, Anderson–Darling, Cramér–von Mises, Kuiper, and Watson to the multivariate domain while maintaining practical applicability. In recent years, several important studies have been conducted to enhance the understanding and applicability of multivariate normality testing in both theoretical and practical domains. For instance, Chen and Xia [40] have suggested an approach for high-dimensional normality testing using nearest-neighbors which is able to guarantee a correct Type I error rate control irrespective of whether the number of dimensions is larger than the sample size. Likewise, Song and Zhao [41] have recommended

a Jarque-Bera test along with principal component analysis which exhibited higher power for high-dimensional data. Móri et al. [42] have made improvements to the energy distance test. Feng et al. [43] have also contributed to this field by developing an innovative goodness-of-fit test for the assessment of multivariate normality based on entropy. Finally, concerning practical applications of machine learning, it has been demonstrated that multivariate normality testing is important for the construction of synthetic biomedical data sets as described by Heine et al. [44]. Tests such as Anderson-Darling and Mardia's test were used to check the statistical validity of the created datasets.

Additionally, recent applications of statistical models and experiments have emphasized the importance of distributional assumptions for the proper functioning of statistical models and experiments, although formal tests for multivariate normality were rarely conducted. In particular, Al-Marsomi and Al-Zwainy [45] developed a structural equation model (SEM) based on PLS-SEM for analyzing critical success factors of regional development programs. Although the authors underlined the advantages of PLS-SEM due to the lack of distributional assumptions, no multivariate normality tests were conducted in this work. Another study by Risan et al. [46] offered an effective management framework for engineering and scientific experiments in which statistical analysis was considered as one of the main steps. Although the authors indicated the need to verify assumptions about normality and homogeneity of variances before the interpretation of the statistical model, there was no formal procedure suggested in the paper to conduct a multivariate normality test. These are some of the examples showing the current gap between the theoretical importance of distributional assumptions and the practice of their testing. This study aims at overcoming this gap by developing a series of tests for checking multivariate normality.

The purpose of this research is to solve this problem via developing an integrated and computationally convenient methodology for extending the tests for general multivariate normality testing. The developed methodology is tested through power analysis based on simulations and application to real data sets.

In the existing literature, several comparative studies including the multivariate Kolmogorov-Smirnov test have already been carried out. For instance, Kesemen et al. [47] proposed a new test for multivariate normality and compared it with the Kolmogorov-Smirnov test under different distributional settings. Given that such comparative analyses are already well-established, the present study did not aim to repeat them. Rather than following traditional approaches, the primary objective of this study is to propose a generalized family of empirical goodness-of-fit tests that rely on the discrepancy between empirical and theoretical distribution functions. These tests are then specifically compared to the multivariate Kolmogorov–Smirnov test in terms of performance and applicability.

UNIVARIATE EMPIRICAL GOODNESS-OF-FIT TESTS

An empirical goodness-of-fit test assesses how well a theoretical distribution represents the data by comparing it with the empirical distribution function. Prominent examples of univariate empirical goodness-of-fit tests include the Kolmogorov–Smirnov test [4, 5], the Cramér–von Mises test [7-9], the Anderson–Darling test [6], the Kuiper test [12, 13], and the Watson test [10, 11].

Table 1. Univariate empirical goodness-of-fit tests

The Statistic of Kolmogorov-Smirnov Test
$D_n = \sup_x \lvert F_n(x) - F_0(x) \rvert$
The Statistic of Cramér-von Mises Test
$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F_0(x_{(i)}) - F_n(x_{(i)}) + \frac{1}{2n} \right]^2$
The Statistic of Anderson-Darling Test
$A^2 = -n - \sum_{i=1}^{n} \left[\left(2F_n(x_{(i)}) - \frac{1}{n} \right) \log\left(F_0(x_{(i)}) \right) + \left(2\left(1 - F_n(x_{(i)}) \right) + \frac{1}{n} \right) \log\left(1 - F_0(x_{(i)}) \right) \right]$
The Statistic of Kuiper Test
$D^+ = \sup_x \{ F_n(x) - F_0(x) \}$, $D^- = \sup_x \{ F_0(x) - F_n(x) \}$,
$V = D^+ + D^-$
The Statistic of Watson Test
$U^2 = W^2 - n\left(\frac{1}{n} \sum_{i=1}^{n} x_i - 0.5 \right)$

Let $x = \{(x_i), i = 1, \dots n\}$, with $x_i \in$ R, denote a random sample of independent and identically distributed (i.i.d) from an unknown distribution function F. The hypothesis is formulated as H_0: $F(x) = F_0(x)$ versus H_1: $F(x) \neq F_0(x)$, where F_0 is a specified theorical distribution. Based on this framework, several goodness-of-fit tests exist to evaluate the discrepancy between empirical and theoretical distributions. A selection of these tests is summarized in Table 1

Here, $F_0(x)$ is the univariate normal distribution function and $F_n(x)$ is the empirical distribution function of the sample. Also, the ordered sample is represented by $x_{(i)}$. The test process is concluded by comparing each test statistic with the relevant critical table value.

Generalization of Empirical Goodness-of-Fit Tests

In this section, the univariate empirical goodness-of-fit tests are extended to multivariate normality tests. These tests include the Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling, Kuiper, and Watson tests, respectively. To compute the test statistics, both empirical and theoretical distribution values need to be determined. The challenge in the multivariate case is that, unlike in the univariate case, the variables cannot be assumed to be independent from the distribution [47]. Therefore, the variables must first be rendered independent to address this issue.

Consider a random sample $\mathbf{X} = \{(x_{1i}, x_{2i}, \dots, x_{di}): i = 1, 2, \dots, n\} \subset \mathbb{R}^d$ consisting of i.i.d. realizations from an unknown d-variate distribution function F. The objective is to assess the validity of the null hypothesis stating that $F(\mathbf{x}) = F_0(\mathbf{x})$, in contrast to the competing claim that $F(\mathbf{x})$ differs from $F_0(\mathbf{x})$, where $F_0(\mathbf{x})$ denotes the distribution function of a multivariate normal distribution $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with a specified mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. According to Theorem 1, each observation in the sample can be represented parametrically in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as shown in Equation (1).

$$\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu} \tag{1}$$

Here, the variables $\boldsymbol{Z}_k$ follow a distribution characterized by a mean of zero and a variance of one [33, 48].

Theorem 1. Let $\mathbf{X}$~N($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$), and define $\boldsymbol{Z} = \boldsymbol{BX} + \boldsymbol{b}$, where B is a matrix and $\boldsymbol{b}$ is a constant vector. Then $\boldsymbol{Z}$ follows a multivariate normal distribution, specifically $\boldsymbol{Z} \in N(\boldsymbol{B\mu} + \boldsymbol{b}, \boldsymbol{B\Sigma B}')$.

Proof. The first part of the proof focuses on showing that any linear combination of the elements of $\boldsymbol{Z}$ can be expressed as a linear combination of the elements of $\boldsymbol{X}$. In other words, the objective is to establish that $\boldsymbol{a}'\boldsymbol{Z}$ is normal for every **a**. However,

$$\boldsymbol{a}'\mathbf{Z} = \boldsymbol{a}'\boldsymbol{BX} + \boldsymbol{a}'\boldsymbol{b} = (\boldsymbol{B}'\boldsymbol{a})'\boldsymbol{X} + \boldsymbol{a}'\boldsymbol{b} = \boldsymbol{c}'\boldsymbol{X} + d, \tag{2}$$

where $\boldsymbol{c} = \boldsymbol{B}'\boldsymbol{a}$ and $d = \boldsymbol{a}'\boldsymbol{b}$**.** Since $\boldsymbol{c}'\boldsymbol{X}$ is normal by the characterization that $\boldsymbol{X}$ is multivariate normal if and only if $\boldsymbol{a}'\boldsymbol{X}$ is univariate normal for all constant vectors $\boldsymbol{a}$, and since $\boldsymbol{d}$ is fixed, it follows that $\boldsymbol{a}'\boldsymbol{Z}$ is also normally distributed [48]**.**

Since the density of $\boldsymbol{Z}$ is known, the density of $\boldsymbol{X}$ is calculated as follows::

$$\begin{aligned} f_{\mathbf{Z}}(z) &= \prod_{k=1}^{n} f_{Z_k}(z_k) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_k^2}{2}\right\} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{n} z_k^2\right\} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\{\boldsymbol{z}'\boldsymbol{z}\}, \quad \boldsymbol{z} \in \mathbb{R}^{\boldsymbol{n}}. \end{aligned} \tag{3}$$

Also, since det $\boldsymbol{\Sigma} > 0$, the inverse $\boldsymbol{\Sigma}^{-1/2}$ exists. The variables $\boldsymbol{X}_k$ can be expressed as linear combinations of the variables $\boldsymbol{Z}_k$. The values of $Z_1, Z_2, \dots, Z_d$ are calculated as in Equation (4).

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \tag{4}$$

It is well established that linear transformations can be applied to multivariate normal vectors in such a way that the resulting vector retains normality and may exhibit component-wise independence. Specifically, any orthogonal transformation applied to a normal vector with independent components and equal variances yields another multivariate normal vector whose components remain independent [48].

Theorem 2. Let **X** be a multivariate normal random vector. If the components of **X** are uncorrelated, then they are also statistically independent.

Proof. It suffices to show that uncorrelated components of a multivariate normal vector are independent, since the converse holds by definition. Assuming that $Cov(X_i, X_j) = 0$ for all $i \neq j$, it follows that the covariance matrix $\boldsymbol{\Sigma}$ is diagonal, with diagonal entries $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. If any variance term $\sigma_k^2 = 0$, the corresponding component is degenerate and thus trivially independent from the others. For the purposes of the argument, one may therefore assume that all variances are strictly positive. Under this condition, the covariance matrix $\boldsymbol{\Sigma}$ is invertible, and its inverse $\boldsymbol{\Sigma}^{-1}$ is also diagonal with entries $1/\sigma_1^2, 1/\sigma_1^2, \dots, 1/\sigma_n^2$.

It can thus be concluded that the corresponding density function is equal to:

$$\begin{aligned} f_X(x) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\prod_{k=1}^{n}\sigma_k} \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}\frac{(x_k-\mu_k)^2}{\sigma_k^2}\right\} \\ &= \prod_{k=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{(x_k-\mu_k)^2}{2\sigma_k^2}\right\}, \end{aligned} \tag{5}$$

which proves the desired independence [48].

Let $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since the matrix $\boldsymbol{\Sigma}$ is nonnegative-definite, there exists an orthogonal matrix **C** such that $\boldsymbol{C}'\boldsymbol{\Sigma C} = \boldsymbol{D}$, where the diagonal matrix **D** has its diagonal elements taken to be the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of $\boldsymbol{\Sigma}$. Set $\boldsymbol{Z} = \boldsymbol{C}'\boldsymbol{X}$. It follows from the established result (see Theorem 1) that $\boldsymbol{Z} \in N(\boldsymbol{C}'\boldsymbol{\mu}, \boldsymbol{D})$. The components of **Z** are thus uncorrelated and independent. This establishes the following Theorem 3:

Theorem 3. Let $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, define $\boldsymbol{Z} = \boldsymbol{C}'\boldsymbol{X}$, where **C** is an orthogonal matrix satisfying $\boldsymbol{C}'\boldsymbol{\Sigma C} = \boldsymbol{D}$, with $\boldsymbol{D}$ being a diagonal matrix. Then $\boldsymbol{Z}$~$N(\boldsymbol{C}'\boldsymbol{\mu}, \boldsymbol{D})$. Furthermore, the components of **Z** are mutually independent, and the

variance of each component Z_k is given by $Var(Z_k) = \lambda_k$, where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of $\mathbf{\Sigma}$.

Under the condition that the variables are conditionally independent of each other, the theoretical distribution function can be expressed using the marginal distribution functions with the aid of the Rosenblatt transform [49], as shown in Equation (8).

Theorem 4. (Rosenblatt) Let $\boldsymbol{X}$ be a joint random density vector whose components $X_1, X_2, \dots, X_d$, the joint distribution function is obtained as follows:

$$f_0(x_1, \dots, x_d) = f_1(x_1) f_2(x_2|x_1) \dots f_d(x_d|x_1, x_2, \dots, x_d) \quad (6)$$

and define the transformation $Y = T(\boldsymbol{X})$ by

$$\begin{aligned} Y_1 &= F(X_1) \\ Y_j &= F\big(X_j|X_1, \dots, X_{j-1}\big), \quad j = 2, \dots, d. \end{aligned} \quad (7)$$

This transformation yields $Y_1, Y_2, \dots, Y_d$ as i.i.d random variables following a uniform distribution in the interval [0,1] [22].

Using the independent and uncorrelated $Z_1, Z_2, \dots, Z_d$ values from Equation (4), the following equation is obtained with the help of Equation (7).

$$F_0(z_1, \dots, z_d) = F(z_1) F(z_2) \dots F(z_d) \quad (8)$$

A new random variable U_k is constructed by applying the marginal distribution function to the corresponding random variable [50].

$$U_j = F\big(z_j\big), \; j = 1, \dots, d. \quad (9)$$

Assuming that $X_1, X_2, \dots, X_d$ follow the distribution F_0, the transformed variables $Z_1, Z_2, \dots, Z_d$, as defined in Equation (4), are mutually independent standard normal random variables, with their joint distribution denoted by F^*.

For this instance, the theoretical distribution function is calculated by Equation (8) as follows:

$$F_0^*(\mathbf{z}) = \prod_{k=1}^{d} U_k \,. \quad (10)$$

Each U_k is an independent random variable drawn from the uniform distribution on [0,1]. Also, the multiplication of U is $F^*(\mathbf{z})$ and its distribution function is $F_U^{(d)}(u)$ which is given as follows [51].

$$F_U^{(d)}(u) = \sum_{k=1}^{d} \frac{(-1)^{k-1}}{(k-1)!} u \log^{k-1}(u). \quad (11)$$

Then, new obtained theoretical distribution function values $C_0(\mathbf{z})$ are calculated as in Equation (12):

$$C_0(\mathbf{z}) = F_U^{(d)}\big(F_0^*(\mathbf{z})\big). \quad (12)$$

Let $Z = \{z_{1i}, z_{2i}, \dots, z_{di}\}$, $i = 1, \dots, n$ represent the transformed version of the original observations $X = \{x_{1i}, x_{2i}, \dots, x_{di}\}$, $i = 1, \dots, n$ obtained via the transformation specified in Equation (4). The empirical distribution values of the independent $z_{1i}, z_{2i}, \dots, z_{di}$ values $F_n^*(\mathbf{z})$ are calculated as in Equation (13):

$$\begin{aligned} F_n^*(\boldsymbol{z}) &= \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(F_0^*(z_{1j}, z_{2j}, \dots, z_{dj}), F_0^*(\mathbf{z})) \,, \\ \mathbf{1}(\mathrm{A}, \mathrm{B}) &= \begin{cases} 1, & A \le B \\ 0, & otherwise \end{cases} \end{aligned} \quad (13)$$

where $\mathbf{1}(.)$ is the indicator function [24].

Both the theoretical and the empirical distribution values of $z_{1i}, z_{2i}, \dots, z_{di}$ values are obtained with the help of Equation (12) and Equation (13), respectively. The desired

Table 2. Multivariate empirical goodness-of-fit tests for multivariate normality

The Statistic of Multivariate Kolmogorov-Smirnov Test
$MVKS = \sup_{z} \lvert F_n^*(\mathbf{z}) - C_0(\mathbf{z}) \rvert$
The Statistic of Multivariate Cramér-von Mises Test
$MVCvM = \frac{1}{12n} + \sum_{i=1}^{n} \left[C_0\big(z_{1_i}, z_{2i}, \dots, z_{di}\big) - F_n^*\big(z_{1_i}, z_{2i}, \dots, z_{di}\big) + \frac{1}{2n} \right]^2$
The Statistic of Multivariate Anderson-Darling Test
$MVAD = -n - \sum_{i=1}^{n} \left[\left(2F_n^*\big(z_{1_i}, z_{2i}, \dots, z_{di}\big) - \frac{1}{n} \right) \log\left(C_0\big(z_{1_i}, z_{2i}, \dots, z_{di}\big) \right) + \left(2\left(1 - F_n^*\big(z_{1_i}, z_{2i}, \dots, z_{di}\big)\right) + \frac{1}{n} \right) \log\left(1 - C_0\big(z_{1_i}, z_{2i}, \dots, z_{di}\big)\right) \right]$
The Statistic of Multivariate Kuiper Test
$D^+ = \sup_{z} \{F_n^*(\mathbf{z}) - C_0(\mathbf{z})\}$, $D^- = \sup_{z} \{C_0(\mathbf{z}) - F_n^*(\mathbf{z})\}$
$MVKUI = D^+ + D^-$
The Statistic of Multivariate Watson Test
$MVWAT = MVCvM - n(\bar{\mathbf{z}} - 0.5)$

goodness-of-fit test will be performed with the help of theoretical distribution values $C_0(\mathbf{z})$ and empirical distribution values $F_n^*(\mathbf{z})$. An adaptation of univariate empirical goodness-of-fit tests to multivariate normality is given in Table 2. Throughout this study, the test statistic for the multivariate Kolmogorov-Smirnov test is referred to as $MVKS$, the multivariate Cramér-von Mises test as $MVCvM$, the Multivariate Anderson-Darling test as $MVAD$, the multivariate Kuiper test as $MVKUI$ and the multivariate Watson test as $MVWAT$.

Algorithm for Generalizing Empirical Goodness-of-Fit Tests to Multivariate Normality

An algorithm has been developed to extend empirical tests designed for univariate distributions to the context of multivariate normality. The desired univariate empirical goodness-of-fit test can be adapted to the multivariate normality and the test process can be performed easily with Algorithm 1.

Algorithm 1. Algorithm of the generalization of empirical goodness-of-fit tests for multivariate normality

Initial Parameter Determination

Step 1. Specify the sample ($\mathbf{x}$), significance value (α).

Step 2. Derive the statistical parameters ($\boldsymbol{\mu}$) and (Σ), representing the sample's location and scale, respectively.

If they are unknown, estimate from the sample.

Ensuring Variable Independence

Step 3. Center the data by subtracting the mean vector ($\boldsymbol{\mu}$) from the sample values $\mathbf{x}$.

Step 4. Standardize the data by pre-multiplying the centered observations with $\Sigma^{-1/2}$, the inverse matrix square root of the covariance matrix (Σ).

Calculation of the multivariate test statistics

Step 5. Evaluate $F_0^*(\mathbf{z})$.

Step 6. Compute the theoretical distribution values $C_0(\mathbf{z})$ as defined in Equation (12).

Step 7. Compute the empirical distribution values $F_n^*(\mathbf{z})$via Equation (13).

Step 8. Calculate the multivariate test statistic by substituting the $C_0(\mathbf{z})$ and/or $F_n^*(\mathbf{z})$ in the desired goodness-of-fit test.

Decision Rule

Step 9. Compare the calculated test statistic with the relevant univariate critical table value,

If the test statistic is less than the critical value, accept the null hypothesis (H_0), Otherwise reject it.

EXPERIMENTAL RESULTS

The tests suggested as the multivariate empirical goodness-of-fit tests are investigated in three different stages. In the first stage, the critical values of these tests are determined using Monte Carlo simulations. In the second stage, the tests are compared in terms of their Type I error rates, whereas in the third stage, the power of the tests is determined. The computations have been done using a computer which has a processing unit with similar performance to that of the i7-4740 and with 16 GB of RAM. Moreover, the experiments have been conducted using MATLAB® R24b software. In order to reduce the time spent in doing the simulations, MATLAB's $parfor$ function, which uses the power of multicore processors, has been used.

The overall computational complexity of the proposed test framework is approximately $O(R \cdot n^2)$, where n is the sample size and R = 100,000 is the number of Monte Carlo trials. The quadratic dependency on n stems from pairwise operations in the test statistics. Although the procedure is computationally demanding, the critical values are precomputed and reused, improving practical efficiency.

Despite the quadratic computational complexity inherent in empirical distribution-based tests ($O(n^2)$), several strategies have been employed to mitigate computational burden. First, all critical values were computed in advance and stored, making the method practical for repeated use without rerunning simulations. Second, parallel computation using MATLAB's "$parfor$" was utilized to efficiently handle simulations across multiple cores. Furthermore, experimental evaluations up to d = 10 and n = 500 demonstrated that the method remains computationally feasible under realistic conditions. In future work, GPU-based acceleration and quasi-Monte Carlo integration are being explored to further improve performance for large-scale applications.

To ensure a thorough evaluation of the proposed multivariate empirical goodness-of-fit tests, simulation experiments were conducted using a wide range of sample sizes: n = 10, 30, 50, 100, 300, and 500. These values were purposefully selected to reflect practical sample conditions frequently encountered in multivariate data analysis. A very low sample size (n = 10) enables investigation of the test's properties when data is scarce, whereas a moderate sample size (n = 30 − 100) reflects an average situation from practice. A large sample size (n = 300, 500) will reveal the properties of tests under asymptotics. Thus, the methodology is evaluated with regard to both information poor and rich cases.

Computation of Critical Tables for Multivariate Empirical Goodness-of-Fit Tests

In this section, the critical values for the $MVCvM$, $MVAD$, $MVKUI$ and $MVWAT$ tests are obtained. Since the critical values for the MVKS test were previously reported [47], they are not repeated here. While calculating the critical tables, the number of variables d was set to 1, 5, and 10, respectively, and the quantiles considered were 0.85, 0.90, 0.95, 0.975, and 0.99. Additionally, the critical values of the proposed multivariate normality tests were computed

for various sample sizes $n \in \{10, 30, 50, 100, 300, 500\}$ and compared with the corresponding univariate critical values.

Moreover, the critical values of the multivariate tests depend on the distribution assumed under the null hypothesis H_0 and on whether the distribution parameters are known. For this reason, critical values were obtained for both scenarios: when the mean vector and covariance matrix are known, and when they are unknown.

If mean and variance are known for the normality test, a modification including the sample size is used for the Cramér-von Mises test (CvM) test, the Kuiper (KUI) test and the Watson test (WAT) test [11, 13]. On the other hand, if mean and variance are unknown for the normality test, a modification including the sample size is used for

CvM, Anderson-Darling (AD), KUI and WAT test [11 ,13]. The decision is made by comparing the new test statistic calculated with this modification and the critical table value. The same modifications are also used to obtain the critical table values of the proposed $MVCvM$, $MVAD$, $MVKUI$ and $MVWAT$ tests in this study. These modifications are given in the Table 3. On the other hand, while there are no modifications for the AD test when the parameters of the normal distribution are known, there are two modifications proposed by Stephens and D'Agostino when the parameters of the normal distribution are unknown. One of these modifications is used in this study.

In addition, the Mean Absolute Percentage Error (MAPE) is employed as a performance metric to quantify the similarity between the obtained critical values and those from the univariate critical tables. In case the MAPE values are obtained quite low, the error between the critical table values is low and the same table values can be used to generalize the univariate normality test for the multivariate case. Also, the MAPE is calculated by each significance level. So, all the levels considered are averaged out. Its formula is given as below [52-54].

$$\text{MAPE} = \frac{1}{TNC}\sum_{i=1}^{TNC}\frac{|CT_i^S - CT_i^C|}{CT_i^C} \tag{14}$$

Table 3. Modifications for empirical goodness of fit tests

	MEAN and VARIANCE	
TEST	**Known**	**Unknown**
CvM	$\left(W^2 - \frac{0.4}{n} + \frac{0.6}{n^2}\right)\left(1 + \frac{1}{n}\right)$	$W^2\left(1 + \frac{0.5}{n}\right)$
AD	-	$A^2\left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$
KUI	$V(\sqrt{n} + 0.155 + 0.24/\sqrt{n})$	$V(\sqrt{n} + 0.05 + 0.82/\sqrt{n})$
WAT	$\left(U^2 - \frac{0.1}{n} + \frac{0.1}{n^2}\right)\left(1 + \frac{0.8}{n}\right)$	$U^2\left(1 + \frac{0.5}{n}\right)$

The total number of elements in the critical table is denoted by TNC. For each index i, CT_i^S, is the critical value computed via simulation, and CT_i^C is the critical value derived from the corresponding univariate test.

Moreover, during the construction of the critical value tables, samples are generated from multivariate normal distributions under the assumption of dependence among variables. In the multivariate setting, both the mean vector and the covariance matrix are assigned through random generation. It is noteworthy that identical critical values may be obtained for different configurations of mean vectors and covariance matrices. In this process, the mean vector $\boldsymbol{\mu}$ is sampled from a uniform distribution over the interval [−1000, 1000]. The diagonal entries of the covariance matrix Σ, representing the variances, are independently drawn from a uniform distribution on [0, 100], whereas the off-diagonal elements, corresponding to correlation coefficients, are sampled from a uniform distribution on [−1, 1].

This wide range for the mean vector ($\mu \in [-1000, 1000]$) was intentionally selected to simulate diverse location parameters and to robustly test the stability of the proposed methods under various distributional shifts. Although this range may exceed typical values encountered in standardized datasets, it reflects potential use cases in applied domains such as signal processing or environmental measurements where variable magnitudes can vary substantially. Moreover, since the generated data strictly follow the multivariate normal distribution, the absolute scale of μ does not distort the underlying distributional assumption. This justification has been explicitly provided to ensure the transparency and reproducibility of the simulation design.

Computation of Critical Tables for Multivariate Cramér-von Mises Test

The proposed multivariate goodness-of-fit tests are evaluated starting with the multivariate Cramér-von Mises ($MVCvM$) test. Table 4 summarizes the critical values derived from 100,000 Monte Carlo trials, considering both known and unknown distributional parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).

When Table 4 is examined, it is seen that the critical table values of the $MVCvM$ in each dimension are quite close to the univariate CvM test critical table values. In the case of known mean vector and covariance matrix, the MAPE values for d = 1, 5, 10 are calculated as 0.005, 0.010, and 0.006, respectively. Additionally, for the unknown case, the MAPE values for d = 1, 5, 10 are obtained as 0.004, 0.063, and 0.081, respectively. The fact that the MAPE values are quite low indicates that the error between the critical table values is minimal and the same table values can be used to generalize the univariate CvM test to the multivariate case.

To investigate the behavior of the $MVCvM$ test's critical values across different dimensions and sample sizes, the tabulated values for dimensions d = 1, 5, 10 and sample sizes n = 10, 30, 50, 100, 300, 500 are visualized. The corresponding plot is provided in Figure 1.

Table 4. Simulation results for critical table values of *MVCvM* test

			μ and Σ							
			Known				Unknown			
		Sample Size	0.85	0.90	0.95	0.99	0.85	0.90	0.95	0.99
CvM	$d = 1$	-	0.284	0.347	0.461	0.743	0.091	0.104	0.126	0.179
MVCvM	$d = 1$	10	0.276	0.345	0.464	0.747	0.091	0.103	0.126	0.176
		30	0.280	0.345	0.460	0.743	0.091	0.104	0.126	0.178
		50	0.281	0.345	0.460	0.746	0.091	0.104	0.126	0.178
		100	0.281	0.345	0.461	0.742	0.090	0.104	0.126	0.180
		300	0.285	0.347	0.462	0.737	0.090	0.103	0.125	0.179
		500	0.285	0.349	0.460	0.740	0.090	0.104	0.125	0.179
	$d = 5$	10	0.271	0.337	0.452	0.740	0.095	0.108	0.131	0.179
		30	0.282	0.348	0.466	0.747	0.098	0.111	0.137	0.193
		50	0.285	0.349	0.467	0.759	0.097	0.111	0.135	0.192
		100	0.283	0.345	0.462	0.744	0.096	0.109	0.134	0.188
		300	0.286	0.352	0.466	0.750	0.098	0.113	0.138	0.195
		500	0.284	0.346	0.461	0.754	0.096	0.109	0.132	0.185
	$d = 10$	10	0.276	0.344	0.465	0.752	0.095	0.109	0.132	0.185
		30	0.280	0.344	0.461	0.748	0.098	0.112	0.135	0.191
		50	0.281	0.344	0.461	0.743	0.099	0.113	0.137	0.194
		100	0.284	0.347	0.461	0.759	0.099	0.114	0.140	0.200
		300	0.285	0.348	0.461	0.733	0.098	0.112	0.135	0.192
		500	0.282	0.347	0.461	0.740	0.098	0.114	0.139	0.194

Figure 1 illustrates the critical values of the $MVCvM$ test across various sample sizes (n = 10, 30, 50, 100, 300, 500) and dimensions (d = 1, 5, 10) for both known (left column: (a), (c), (e)) and unknown (right column: (b), (d), (f)) parameter settings. Each plot includes critical values at different quantiles (0.85, 0.90, 0.95, 0.99), represented respectively with circle, square, pentagon, and hexagon markers, while the '×' marker denotes the simulated critical values (SCV) obtained from 100,000 trials.

Across all dimensions, the critical values appear to remain relatively stable as sample size increases, particularly in the known parameter case. Moreover, the simulated critical values (SCV) closely follow the trend of tabulated quantiles, especially at lower significance levels, indicating consistency and robustness of the $MVCvM$ test across varying conditions. However, in the unknown parameter scenario (right column), minor fluctuations are observed in the SCV, particularly in higher dimensions and smaller sample sizes, which is expected due to increased estimation variability.

In general, these graphs show that the tabled univariate CvM statistics give a good estimate even in multivariate situations. This shows the validity of using the critical value tables for the univariate case in multivariate $MVCvM$ applications.

Computation of Critical Tables for Multivariate Anderson-Darling Test

Following the MVCvM test, a similar simulation-based procedure was applied to compute the critical values of the MVAD test. The results for this test are summarized in Table 5, under both known and unknown parameter assumptions.

When Table 5 is examined, it is observed that the critical table values of the $MVAD$ in each dimension are quite close to the univariate AD test critical table values. As the sample size changes, slight deviations in the critical values occur due to randomness. In addition, as the sample size increases, the critical table values for the $MVAD$ test become closer to each other. In the case of known mean vector and covariance matrix, the MAPE values for d = 1, 5, 10 are calculated as 0.003, 0.014, and 0.009, respectively. On the other hand, for the unknown case, the MAPE values for d = 1, 5, 10 are calculated as 0.002, 0.085, and 0.102, respectively.

To explore how the critical values of the $MVAD$ test change with respect to dimension and sample size, the computed values for dimensions d = 1, 5, 10 and sample sizes n = 10, 30, 50, 100, 300, 500 are presented graphically. The corresponding results are depicted in Figure 2.

Figure 2 displays the critical values of the $MVAD$ test for dimensions d = 1, 5, 10 across increasing sample sizes, under both known ((a), (c), (e)) and unknown ((b), (d),

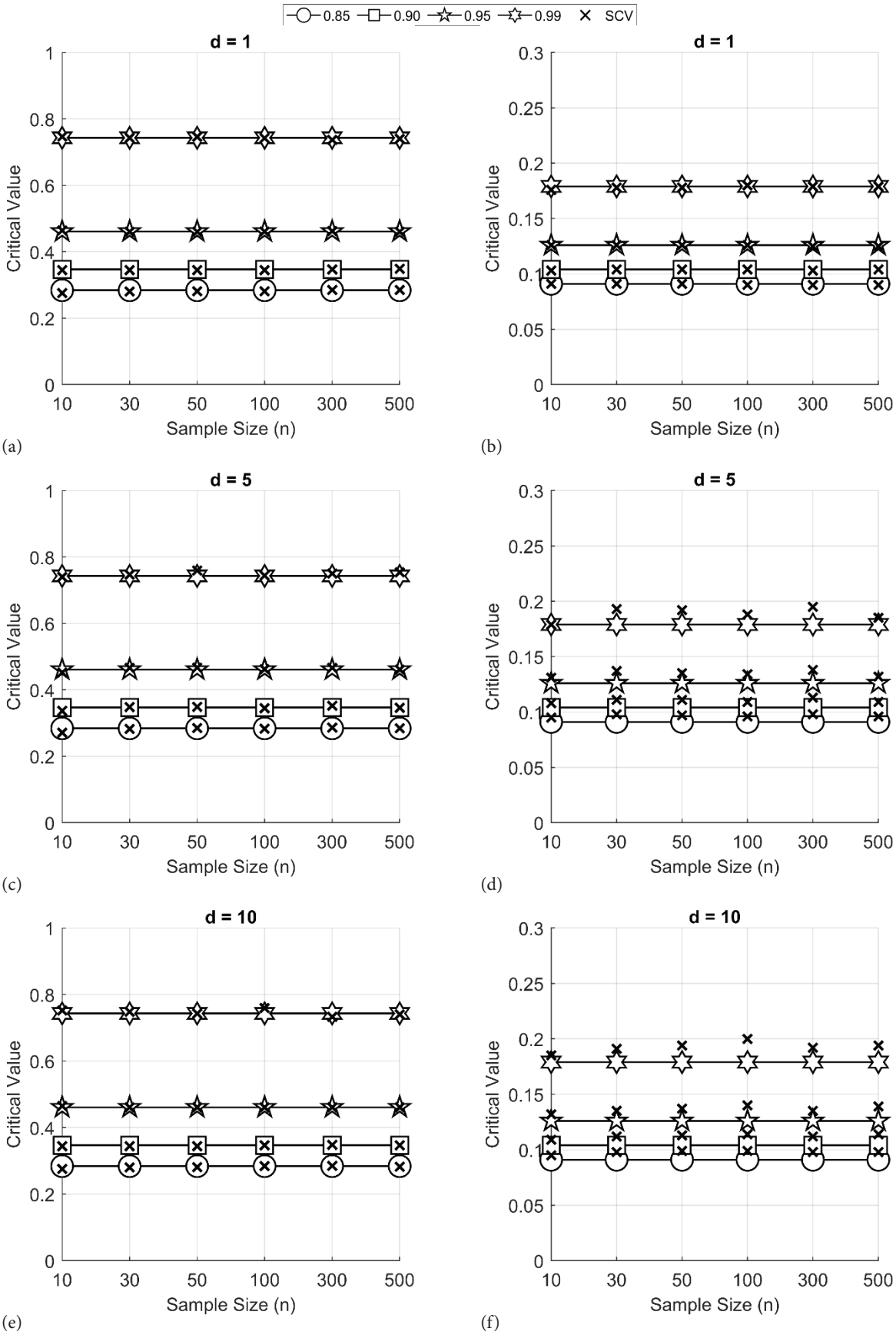

Figure 1. Critical values of the *MVCvM* test for different sample sizes (n = 10, 30, 50, 100, 300, 500) and dimensions (d = 1, 5, 10). Left column: known parameter case ((a), (c), (e)); right column: unknown parameter case ((b), (d), (f)).

Table 5. Simulation results for critical table values of *MVAD* test

			μ and Σ							
			Known				Unknown			
		Sample Size	0.85	0.90	0.95	0.99	0.85	0.90	0.95	0.99
AD	$d = 1$	$n \geq 8$	1.621	1.933	2.492	3.878	0.561	0.631	0.752	1.035
MVAD	$d = 1$	10	1.625	1.942	2.510	3.920	0.566	0.636	0.755	1.028
		30	1.620	1.932	2.498	3.903	0.561	0.632	0.753	1.034
		50	1.621	1.933	2.493	3.890	0.561	0.632	0.751	1.035
		100	1.624	1.937	2.498	3.881	0.559	0.629	0.751	1.035
		300	1.622	1.932	2.493	3.865	0.559	0.630	0.751	1.032
		500	1.618	1.928	2.490	3.881	0.560	0.630	0.751	1.035
	$d = 5$	10	1.609	1.904	2.468	3.716	0.613	0.689	0.816	1.108
		30	1.604	1.926	2.469	3.969	0.612	0.688	0.821	1.120
		50	1.631	1.949	2.541	3.819	0.611	0.687	0.818	1.133
		100	1.574	1.887	2.464	3.931	0.605	0.677	0.806	1.103
		300	1.610	1.918	2.470	3.830	0.617	0.696	0.832	1.129
		500	1.604	1.950	2.508	3.897	0.611	0.683	0.805	1.100
	$d = 10$	10	1.624	1.927	2.485	3.936	0.619	0.700	0.845	1.199
		30	1.630	1.930	2.481	3.923	0.614	0.691	0.820	1.121
		50	1.623	1.937	2.496	3.864	0.618	0.696	0.824	1.127
		100	1.645	1.950	2.502	3.913	0.618	0.694	0.823	1.113
		300	1.619	1.944	2.548	4.023	0.620	0.698	0.832	1.132
		500	1.608	1.931	2.496	3.942	0.621	0.699	0.831	1.140

(f)) parameter settings. The critical values are provided at quantiles 0.85, 0.90, 0.95, and 0.99, represented respectively by circle, square, pentagon, and hexagon markers. The '×' markers indicate the simulated critical values (SCV) derived from 100,000 trials.

For the known parameter case (left panel), critical values continue to be very stable with respect to sample sizes and follow an almost constant trend. The calculated SCVs correspond well to the quantiles given in tables, which means that the $MVAD$ test is stable and gives reliable critical values regardless of how high the dimension is. In the unknown parameter case (right panel), although SCVs also show some deviations from theoretical quantiles, they are small enough and become more noticeable at high dimension and low sample size levels. Such deviations are explained by the increased variability due to estimation of the mean vector and covariance matrix.

To conclude, the presented graphs prove the validity of the use of critical values from the table of the univariate Anderson-Darling test for the multivariate case.

Computation of Critical Tables for Multivariate Kuiper Test

Building on the preceding analysis, the $MVKUI$ test was also examined using the same simulation framework to derive its critical values. The computed values are provided in Table 6 for both known and unknown settings of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

In Table 6, the critical table values of the $MVKUI$ in each dimension are quite close to the univariate KUI test critical table values. When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, the MAPE values for $d = 1, 5, 10$ are calculated as 0.004, 0.003, and 0.003, respectively. However, for the unknown case, the MAPE values for $d = 1, 5, 10$ are calculated as 0.003, 0.029, and 0.027, respectively.

To evaluate the behavior of the MVKUI test's critical values under different dimensional and sample size conditions, computed values for selected settings are visualized in Figure 3.

Figure 3 presents the critical values of the $MVKUI$ test across various sample sizes ($n = 10, 30, 50, 100, 300, 500$) and dimensions ($d = 1, 5, 10$), under both known ((a), (c), (e)) and unknown ((b), (d), (f)) parameter settings. The plots include tabulated critical values corresponding to quantiles 0.85, 0.90, 0.95, and 0.99, indicated respectively by circle, square, pentagon, and hexagon markers. The '×' symbol represents the simulated critical values (SCV) obtained from 100,000 Monte Carlo trials.

However, in the case of knowing the parameters (the left column), the critical values do not change dramatically with the increase in sample size, while the SCV values agree with the tabled quantiles for all dimensions. The above observations show the reliability and consistency of the MVKUI test regardless of the number of dimensions.

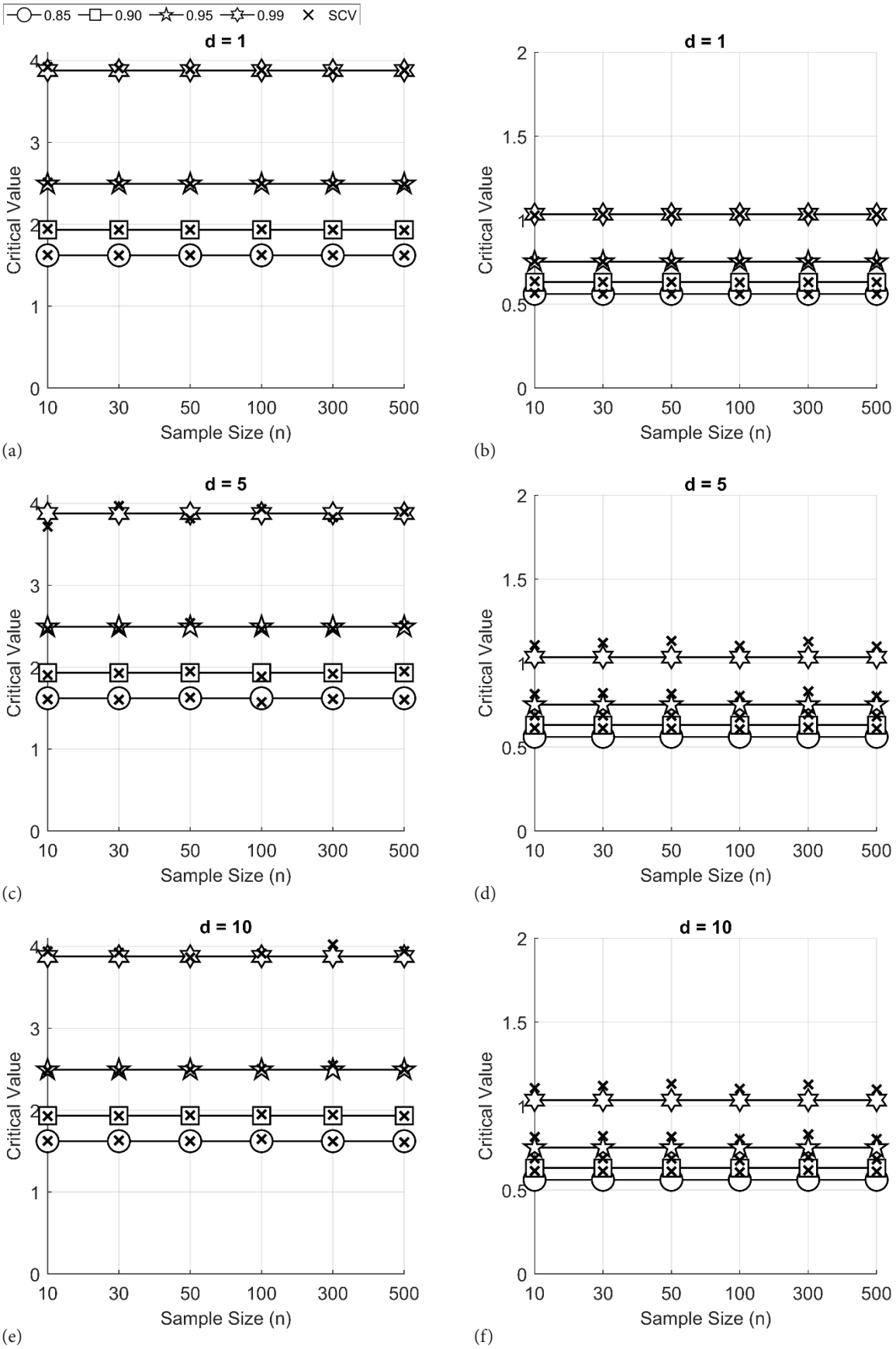

Figure 2. Critical values of the *MVAD* test for different sample sizes (n = 10, 30, 50, 100, 300, 500) and dimensions (d = 1, 5, 10). Left column: known parameter case ((a), (c), (e)); right column: unknown parameter case ((b), (d), (f)).

Table 6. Simulation results for critical table values of *MVKUI* test

			μ and Σ							
			Known				Unknown			
		Sample Size	0.85	0.90	0.95	0.99	0.85	0.90	0.95	0.99
KUI	$d = 1$	-	1.537	1.620	1.747	2.001	1.320	1.386	1.489	1.693
MVKUI	$d = 1$	10	1.535	1.618	1.743	1.985	1.329	1.395	1.496	1.698
		30	1.528	1.611	1.739	1.990	1.316	1.382	1.487	1.697
		50	1.527	1.610	1.737	1.988	1.315	1.383	1.487	1.691
		100	1.531	1.613	1.742	2.001	1.319	1.384	1.486	1.696
		300	1.531	1.614	1.742	1.991	1.321	1.389	1.494	1.701
		500	1.532	1.613	1.746	1.995	1.324	1.392	1.495	1.698
	d = 5	10	1.532	1.618	1.743	1.992	1.366	1.436	1.542	1.748
		30	1.528	1.611	1.740	2.004	1.353	1.423	1.527	1.734
		50	1.527	1.611	1.743	1.999	1.350	1.417	1.528	1.744
		100	1.532	1.613	1.735	1.978	1.353	1.421	1.524	1.738
		300	1.536	1.619	1.750	2.002	1.359	1.428	1.537	1.757
		500	1.530	1.612	1.741	1.997	1.358	1.424	1.532	1.748
	$d = 10$	10	1.532	1.614	1.743	1.987	1.305	1.374	1.479	1.686
		30	1.532	1.614	1.740	1.994	1.353	1.422	1.528	1.743
		50	1.530	1.615	1.745	2.004	1.354	1.425	1.535	1.746
		100	1.533	1.618	1.747	2.002	1.360	1.428	1.536	1.756
		300	1.529	1.612	1.739	1.976	1.357	1.429	1.536	1.759
		500	1.534	1.618	1.746	1.994	1.367	1.435	1.541	1.747

In case of unknown parameters (the right column), there is some variation in the SCV values, especially for small sample sizes and for higher dimensions. Such variations could be associated with the increased variation due to the estimation of the mean vector and the covariance matrix. Still, the simulated values are in good agreement with the tabled quantiles, which implies that the reliability of the MVKUI test is also acceptable in cases of estimation uncertainty.

Thus, it follows from the plots that the tabulated critical values for the univariate Kuiper test can be applied to the multivariate case successfully. The $MVKUI$ test demonstrates practical applicability and statistical reliability across a range of dimensional and sampling scenarios.

Computation of Critical Tables for Multivariate Watson Test

As a final step in the critical value computation framework, the $MVWAT$ test was subjected to the same simulation procedure. Table 7 presents the resulting critical values based on 100,000 trials, for both known and unknown configurations of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

When Table 7 is examined, the critical table values of the $MVWAT$ in each dimension are quite close to the univariate WAT test critical table values both known and unknown parameter cases. When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, the MAPE values for $d = 1, 5, 10$ are calculated as 0.004, 0.005, and 0.007, respectively. On the other hand, when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, the MAPE values for $d = 1, 5, 10$ are calculated as 0.006, 0.059, and 0.072, respectively.

Figure 4 illustrates the critical values of the $MVWAT$ test across various dimensional and sample size settings, enabling an assessment of how these values behave under different data configurations.

Figure 4 illustrates the critical values of the $MVWAT$ test across different sample sizes (n = 10, 30, 50, 100, 300, 500) and dimensions (d = 1, 5, 10), under both known ((a), (c), (e)) and unknown ((b), (d), (f)) parameter conditions. The plotted values correspond to the quantiles 0.85, 0.90, 0.95, and 0.99, denoted respectively by circle, square, pentagon, and hexagon markers, while the '×' symbol represents the simulated critical values (SCV) based on 100,000 Monte Carlo trials.

Under the known parameter scenario (left column), the critical values show high stability across all sample sizes and dimensions. The SCV values are closely aligned with the tabulated quantiles, confirming the consistency of the MVWAT test in controlled settings. In the case of unknown parameters (right column), the simulated values still follow the tabulated critical values closely, with only minor fluctuations observed particularly at lower sample sizes. These deviations are likely attributed to variability introduced through the estimation of the mean vector and covariance matrix.

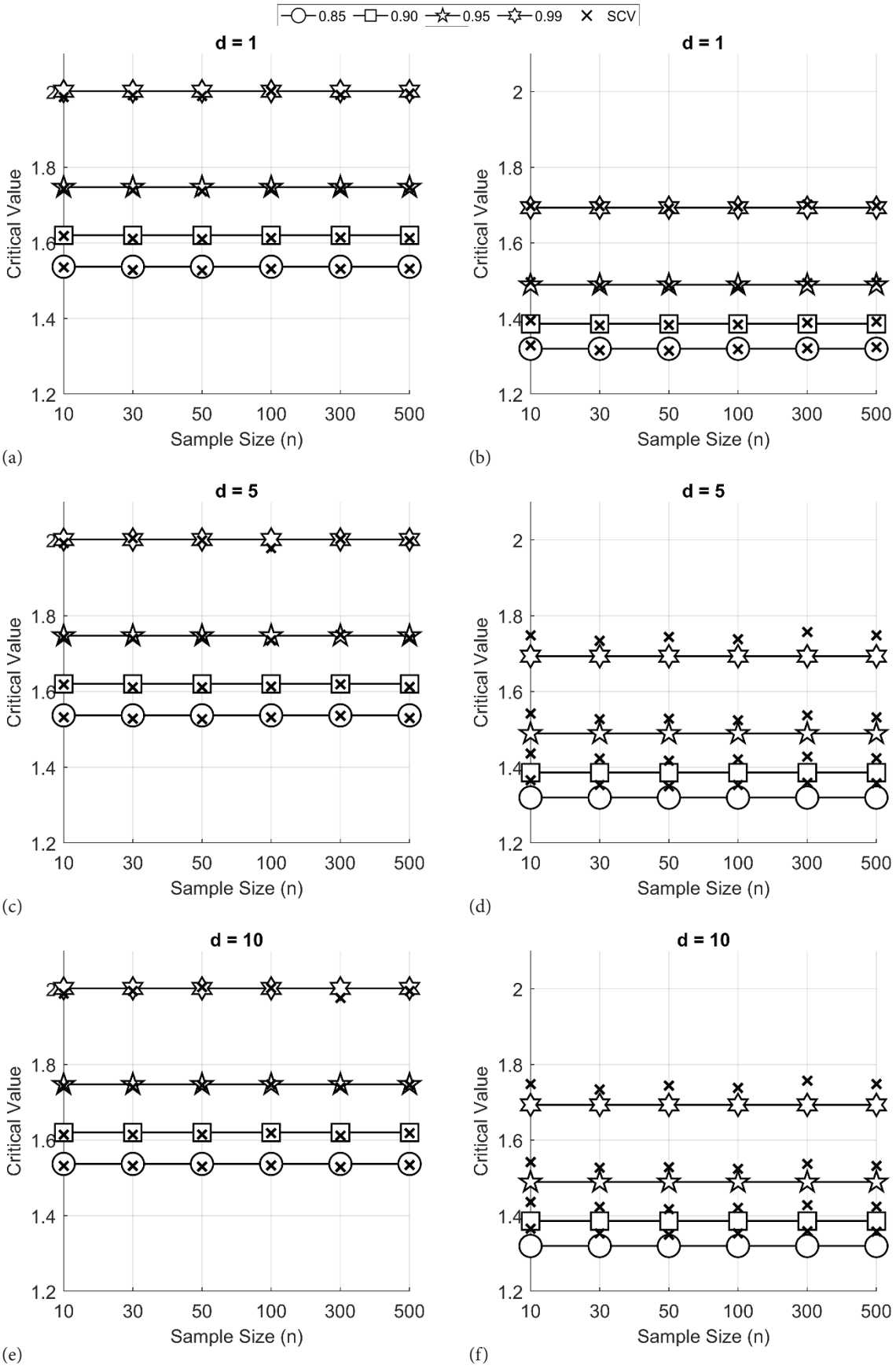

Figure 3. Critical values of the *MVKUI* test for different sample sizes (n = 10, 30, 50, 100, 300, 500) and dimensions (d = 1, 5, 10). Left column: known parameter case ((a), (c), (e)); right column: unknown parameter case ((b), (d), (f)).

Table 7. Simulation results for critical table values of *MVWAT* testt

			μ and Σ							
			Known				Unknown			
		Sample Size	0.85	0.90	0.95	0.99	0.85	0.90	0.95	0.99
WAT	$d = 1$	-	0.131	0.152	0.187	0.267	0.085	0.096	0.117	0.164
MVWAT	$d = 1$	10	0.130	0.151	0.186	0.264	0.086	0.097	0.117	0.165
		30	0.131	0.152	0.187	0.266	0.085	0.096	0.116	0.163
		50	0.131	0.152	0.188	0.268	0.085	0.096	0.117	0.165
		100	0.131	0.151	0.187	0.269	0.084	0.096	0.116	0.163
		300	0.131	0.152	0.187	0.264	0.084	0.096	0.117	0.167
		500	0.131	0.152	0.187	0.268	0.084	0.096	0.116	0.164
	d = 5	10	0.131	0.152	0.188	0.269	0.084	0.097	0.119	0.167
		30	0.132	0.152	0.188	0.267	0.089	0.102	0.124	0.173
		50	0.130	0.151	0.186	0.270	0.089	0.102	0.124	0.176
		100	0.132	0.152	0.184	0.263	0.090	0.103	0.124	0.175
		300	0.132	0.153	0.187	0.269	0.090	0.104	0.126	0.178
		500	0.130	0.150	0.187	0.267	0.090	0.103	0.125	0.177
	$d = 10$	10	0.131	0.152	0.187	0.261	0.084	0.098	0.120	0.172
		30	0.132	0.153	0.187	0.269	0.089	0.102	0.124	0.177
		50	0.131	0.152	0.189	0.265	0.090	0.103	0.126	0.177
		100	0.132	0.152	0.188	0.271	0.092	0.104	0.127	0.180
		300	0.130	0.150	0.185	0.262	0.090	0.103	0.126	0.180
		500	0.132	0.152	0.188	0.267	0.092	0.104	0.126	0.179

Overall, the results suggest that the tabulated univariate Watson critical values provide a reliable approximation for multivariate contexts, supporting the robustness and practical usability of the $MVWAT$ test across a wide range of scenarios.

Type I Error Comparisons of Multivariate Empirical Goodness-of-Fit Tests

The $MVKS$, $MVCvM$, $MVAD$, $MVKUI$ and $MVWAT$ tests are evaluated and compared based on their Type I error rates across varying sample sizes (n = 15, 30, 50, 100, 300, 500). Type I errors are obtained for both known and unknown mean vector-covariance matrix cases. To compute the Type I error rates for each test, 10,000 trials are performed, and the corresponding univariate critical values are used. In the comparison of Type I errors, the null hypothesis is determined as the five-variate and ten-variate normal distributions, respectively. If the sample follows a multivariate normal distribution, each test is expected to reject the null hypothesis at the 5% significance level.

The mean vector of the multivariate normal distribution is defined as zero, and the covariance matrices ($\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$) are specified for both uncorrelated and correlated cases. In the uncorrelated case, the covariance matrix $\boldsymbol{\Sigma}_1$ is constructed with diagonal elements equal to 1 and all off-diagonal elements set to 0. For the correlated case, $\boldsymbol{\Sigma}_2$ maintains unit variances on the diagonal, while the off-diagonal elements are fixed at 0.5.

The Type I error values of $MVKS$, $MVCvM$, $MVAD$, $MVKUI$ and $MVWAT$ tests for both uncorrelated ($\boldsymbol{\Sigma}_1$) and correlated ($\boldsymbol{\Sigma}_2$) covariance matrices are given in Table 8 for the known mean vector and covariance matrix.

When Table 8 is examined, the Type I error values of all multivariate goodness-of-fit tests vary around at the level of 5% for both the uncorrelated ($\boldsymbol{\Sigma}_1$) and correlated ($\boldsymbol{\Sigma}_2$) cases of the covariance matrix. Regardless of changes in sample size, the Type I error rates of all tests remain approximately at the 5% significance level. According to Table 8, all the multivariate normality tests perform accurately and consistently across both the uncorrelated and correlated cases, as well as in five and ten dimensional settings.

In contrast, the Type I error rates observed for the $MVKS$, $MVCvM$, $MVAD$, $MVKUI$ and $MVWAT$ tests for both uncorrelated ($\boldsymbol{\Sigma}_1$) and correlated ($\boldsymbol{\Sigma}_2$) covariance matrices are presented in Table 9 when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. In these scenarios, the specified covariance matrices are utilized to generate random samples from five and ten dimensional multivariate normal distributions. Since the population parameters are typically unknown in practical applications, both the mean vector and the covariance matrix are estimated from the observed data. In such cases, the standard Kolmogorov–Smirnov (KS) test becomes invalid due to its reliance on known parameters. This is overcome through the use of the KS test based on Lilliefors' correction, where

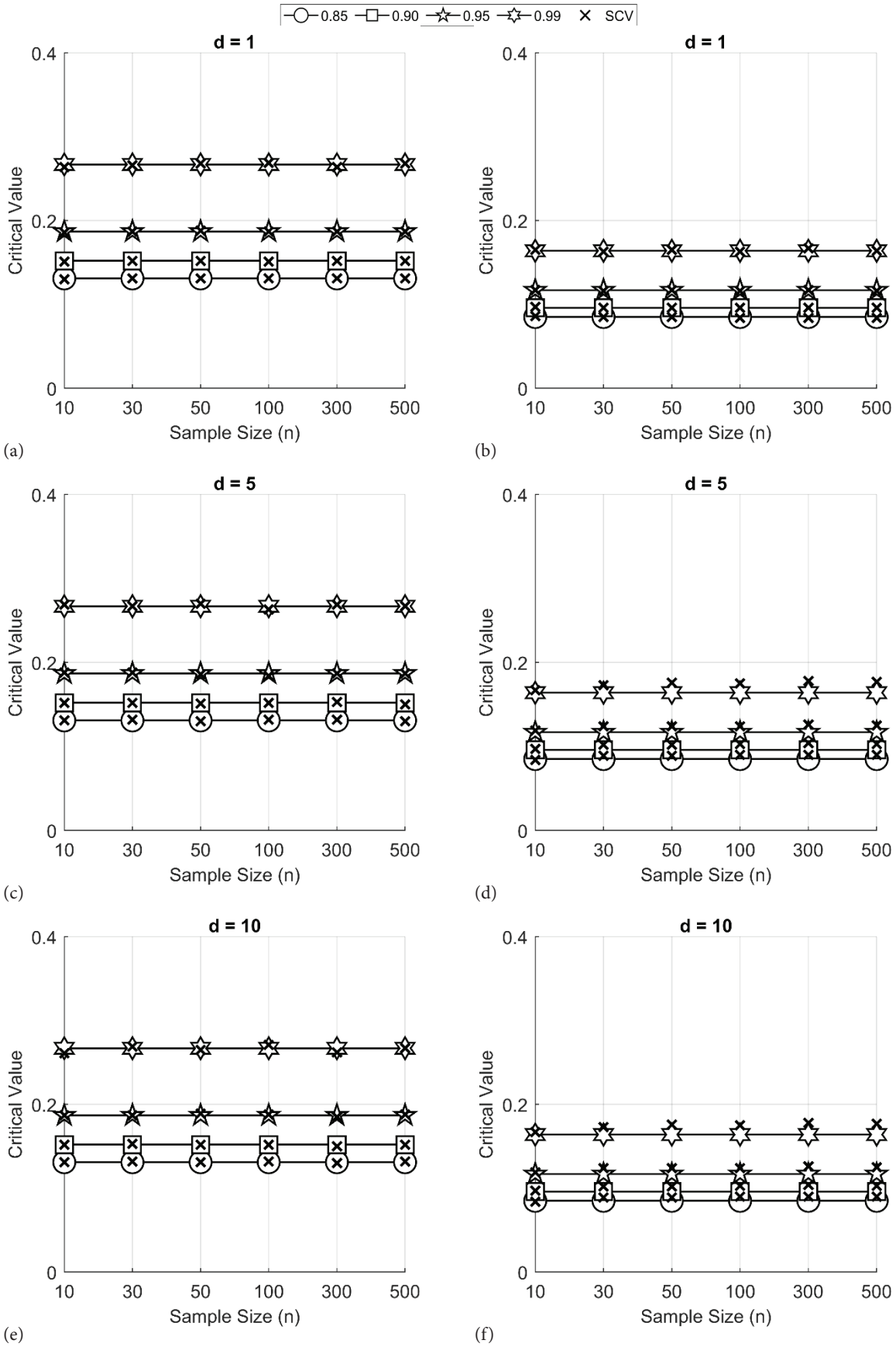

Figure 4. Critical values of the *MVWAT* test for different sample sizes (n = 10, 30, 50, 100, 300, 500) and dimensions (d = 1, 5, 10). Left column: known parameter case ((a), (c), (e)); right column: unknown parameter case ((b), (d), (f)).

Table 8. The Type I errors of the multivariate normality tests for the known mean and variance

			Sample Size					
Dimension	**Covariance Matrix**	**Test**	**15**	**30**	**50**	**100**	**300**	**500**
$d = 5$	Σ_1	*MVKS*	0.0477	0.0515	0.0502	0.0472	0.0498	0.0472
		MVCvM	0.0477	0.0506	0.051	0.0477	0.0532	0.053
		MVAD	0.047	0.0497	0.0518	0.047	0.0537	0.0512
		MVKUI	0.0487	0.0501	0.0525	0.0481	0.0507	0.0451
		MVWAT	0.048	0.0514	0.0502	0.0494	0.05	0.0452
	Σ_2	*MVKS*	0.0437	0.0499	0.0506	0.0485	0.0503	0.0486
		MVCvM	0.0483	0.0509	0.0501	0.0482	0.0528	0.0499
		MVAD	0.0475	0.0519	0.0499	0.0489	0.0547	0.0496
		MVKUI	0.0447	0.0544	0.0519	0.0476	0.0512	0.05
		MVWAT	0.0463	0.0532	0.0505	0.0495	0.0496	0.0486
$d = 10$	Σ_1	*MVKS*	0.0483	0.0528	0.0492	0.0521	0.047	0.0447
		MVCvM	0.0488	0.0516	0.0495	0.0515	0.0477	0.0501
		MVAD	0.0519	0.0535	0.0496	0.0504	0.0487	0.05
		MVKUI	0.0488	0.054	0.0554	0.0526	0.0496	0.0496
		MVWAT	0.0496	0.0506	0.0519	0.0514	0.049	0.0483
	Σ_2	*MVKS*	0.0481	0.0503	0.05	0.0497	0.0453	0.0467
		MVCvM	0.0468	0.0505	0.0493	0.0488	0.0442	0.0503
		MVAD	0.049	0.0515	0.0506	0.0493	0.0453	0.0482
		MVKUI	0.0507	0.0526	0.0527	0.0514	0.0478	0.0455
		MVWAT	0.0501	0.0515	0.0468	0.0505	0.0443	0.0474

Table 9. The Type I errors of the multivariate normality tests for the unknown mean and variance

			Sample Size					
Dimension	**Covariance Matrix**	**Test**	**15**	**30**	**50**	**100**	**300**	**500**
$d = 5$	Σ_1	*MVKS*	0.0496	0.0533	0.0521	0.0503	0.0536	0.0531
		MVCvM	0.0653	0.068	0.0639	0.0612	0.07	0.0633
		MVAD	0.0823	0.0746	0.0714	0.0664	0.0755	0.0693
		MVKUI	0.0652	0.0698	0.0651	0.0612	0.0651	0.0623
		MVWAT	0.0679	0.0711	0.0636	0.0616	0.0661	0.0606
	Σ_2	*MVKS*	0.0488	0.05	0.0544	0.0532	0.0541	0.0518
		MVCvM	0.0628	0.0654	0.068	0.0628	0.067	0.0647
		MVAD	0.0803	0.0714	0.0746	0.0686	0.0727	0.0694
		MVKUI	0.0644	0.0656	0.0652	0.0682	0.0659	0.065
		MVWAT	0.0679	0.0682	0.0666	0.0652	0.0629	0.0653
$d = 10$	Σ_1	*MVKS*	0.044	0.0528	0.0574	0.0557	0.0548	0.0526
		MVCvM	0.0582	0.0704	0.0677	0.0713	0.0739	0.0738
		MVAD	0.0796	0.0764	0.0758	0.0808	0.0807	0.0789
		MVKUI	0.0604	0.074	0.0688	0.07	0.0707	0.0659
		MVWAT	0.0627	0.0746	0.0685	0.0749	0.0704	0.0713
	Σ_2	*MVKS*	0.0469	0.0507	0.0534	0.0548	0.0511	0.0517
		MVCvM	0.0599	0.0691	0.0655	0.0684	0.0713	0.0669
		MVAD	0.0791	0.0747	0.0721	0.0761	0.0779	0.0733
		MVKUI	0.062	0.0697	0.0687	0.0723	0.0677	0.0663
		MVWAT	0.0648	0.0736	0.0665	0.0717	0.0695	0.0669

critical values are computed using simulations to factor in the estimation uncertainty. This adjustment was first applied in the univariate framework and then systematically generalized to the multivariate setting, thereby ensuring the validity and consistency of the proposed methodology under estimated parameters [55, 56].

When Table 9 is examined, the Type I error values of all multivariate goodness-of-fit tests vary around the 5% level for both the uncorrelated ($\boldsymbol{\Sigma}_1$) and correlated ($\boldsymbol{\Sigma}_2$) cases of the covariance matrix. The Type I error rates of all tests consistently remain close to the nominal 5% level, regardless of variations in sample size. According to Table 9, all the multivariate normality tests perform accurately and consistently across both uncorrelated-correlated cases, as well as in five and ten dimensional settings.

Furthermore, a closer inspection reveals that the observed deviations from the nominal significance level are slightly more pronounced in settings where the mean vector and covariance matrix are unknown and must be estimated. These deviations generally remain within a narrow margin of 1–2 percentage points and are considered to be the result of stochastic variation inherent in Monte Carlo simulations rather than a methodological limitation. Given the overall consistency and stability of Type I error rates across diverse configurations, the application of additional calibration procedures was deemed unnecessary.

Test Power Comparisons of Multivariate Empirical Goodness-of-Fit Tests

A power comparison of the multivariate empirical goodness-of-fit tests is conducted for both symmetric and non-symmetric distributions, in cases where the mean vector and covariance matrix are known and unknown. Student-t, Laplace, Cauchy, Triangular, and Uniform distributions are selected as multivariate symmetric distributions, while Exponential, Weibull, Lognormal, and Chi-square distributions are selected as multivariate non-symmetric distributions. The power values $1 - \beta$ of the $MVKS$, $MVCvM$, $MVAD$, $MVKUI$ and $MVWAT$ tests are compared by generating random samples from these distributions.

Table 10. Power comparison for the multivariate symmetric distributions with known mean vector and covariance matrix

		Sample Size					
Distribution	**Test**	**10**	**30**	**50**	**100**	**300**	**500**
Student-t	*MVKS*	0.0619	0.0662	0.073	0.0991	0.2088	0.3589
	MVCvM	0.063	0.0734	0.0835	0.1109	0.2231	0.3656
	MVAD	0.1037	0.1292	0.1523	0.2277	0.5469	0.7932
	MVKUI	0.0565	0.0637	0.0768	0.1022	0.2309	0.3731
	MVWAT	0.0566	0.0622	0.0738	0.0972	0.2277	0.3659
Laplace	*MVKS*	0.0696	0.0986	0.1299	0.22	0.5556	0.7702
	MVCvM	0.064	0.0972	0.1287	0.2178	0.5577	0.7865
	MVAD	0.0719	0.0991	0.1319	0.2216	0.6266	0.8862
	MVKUI	0.071	0.1142	0.1589	0.2847	0.7318	0.9258
	MVWAT	0.0728	0.1167	0.1652	0.305	0.7621	0.9378
Logistic	*MVKS*	0.5333	0.9167	0.992	0.9999	1	1
	MVCvM	0.5491	0.9209	0.9922	0.9999	1	1
	MVAD	0.8264	0.9932	0.9999	1	1	1
	MVKUI	0.4916	0.9141	0.992	1	1	1
	MVWAT	0.4997	0.9139	0.9913	1	1	1
Triangular	*MVKS*	0.1583	0.3211	0.4887	0.7913	0.9982	0.9999
	MVCvM	0.1681	0.3478	0.5209	0.8133	0.9987	1
	MVAD	0.2193	0.4335	0.628	0.8918	0.9997	1
	MVKUI	0.1281	0.2973	0.4675	0.7805	0.9985	1
	MVWAT	0.1325	0.3052	0.4785	0.7891	0.9991	1
Uniform	*MVKS*	0.0496	0.0721	0.0848	0.1164	0.2947	0.4541
	MVCvM	0.0489	0.0683	0.0824	0.1166	0.2929	0.4714
	MVAD	0.0425	0.0596	0.0725	0.1055	0.2992	0.5254
	MVKUI	0.0598	0.0881	0.12	0.1809	0.5069	0.7353
	MVWAT	0.0612	0.0881	0.1197	0.1939	0.5224	0.7522

Power Comparison for Multivariate Symmetric Distributions

The power values $(1 - \beta)$ of tests are obtained by generating random samples from the five-variate Student-t distribution with 10 degrees of freedom, the five-variate Laplace distribution with parameters ($\mu = 0$, $\sigma^2 = 1$), the five-variate Logistic distribution with parameters ($\mu = 0$, $\sigma^2 = 1$), the five-variate Triangular distribution with parameters ($a = -3$, $b = 3$), the five-variate Uniform distribution with parameters ($a = -\sqrt{3}$, $b = \sqrt{3}$). In the calculation of test power, the null hypothesis is specified as a five-dimensional standard normal distribution. The power values are computed at different sample sizes ($n = 10, 30, 50, 100, 300, 500$) and are reported in Table 10 for the case of a known mean vector and covariance matrix.

The test power values of all the multivariate tests increase for the Student-t, Laplace, Logistic, Triangular and Uniform distributions as the sample size increases. The $MVAD$ test has higher power than other tests at each sample size for the five-variate Student-t test. The $MVWAT$ test exhibits higher power than the others at each sample size for the five-variate Laplace distribution. While the $MVAD$ test has the highest power at small sample sizes, the power values approach to 1 for all tests as the sample size increases for the five-variate Logistic and Triangular distributions. The $MVKUI$ and $MVWAT$ tests tend to yield higher power values compared to the remaining methods.

The power values for the case of an unknown mean vector and covariance matrix are presented in Table 11.

The test power values of all multivariate tests increase for the multivariate symmetric distributions as the sample size increases. The $MVAD$ test shows the highest power at small sample sizes compared to the other tests. The $MVKUI$ and $MVWAT$ tests generally exhibit higher power values than the other empirical tests.

Table 11. Power comparison for the multivariate symmetric distributions with unknown mean vector and covariance matrix

		Sample Size					
Distribution	**Test**	**10**	**30**	**50**	**100**	**300**	**500**
	MVKS	0.0567	0.0834	0.1155	0.1953	0.4903	0.315
	MVCvM	0.067	0.1071	0.1496	0.257	0.6282	0.84
Student-t	*MVAD*	0.0849	0.1176	0.164	0.2858	0.6577	0.8591
	MVKUI	0.072	0.0903	0.1106	0.1552	0.3699	0.5634
	MVWAT	0.0731	0.096	0.1181	0.1872	0.4342	0.6461
	MVKS	0.0549	0.1616	0.2891	0.5553	0.9692	0.968
	MVCvM	0.0648	0.2054	0.3627	0.683	0.9945	0.9999
Laplace	*MVAD*	0.0734	0.2132	0.3755	0.6841	0.9933	1
	MVKUI	0.0684	0.1301	0.2048	0.4048	0.8949	0.9901
	MVWAT	0.0668	0.1535	0.243	0.4768	0.9336	0.9959
	MVKS	0.0463	0.0718	0.1004	0.1526	0.3767	0.1816
	MVCvM	0.0542	0.0912	0.1249	0.1992	0.4956	0.7176
Logistic	*MVAD*	0.0679	0.095	0.1357	0.2151	0.5076	0.7284
	MVKUI	0.0632	0.0759	0.0895	0.1172	0.2431	0.3933
	MVWAT	0.0594	0.0818	0.098	0.1368	0.3015	0.4661
	MVKS	0.0525	0.052	0.0684	0.0922	0.2109	0.0601
	MVCvM	0.0667	0.0701	0.0864	0.1206	0.2774	0.453
Triangular	*MVAD*	0.0811	0.0781	0.0938	0.1289	0.2878	0.4657
	MVKUI	0.0732	0.0685	0.0808	0.0938	0.1645	0.2638
	MVWAT	0.0742	0.0734	0.0845	0.107	0.1995	0.3221
	MVKS	0.0569	0.088	0.139	0.2892	0.7855	0.676
	MVCvM	0.0712	0.1167	0.1829	0.3932	0.9304	0.9984
Uniform	*MVAD*	0.0943	0.134	0.1985	0.4091	0.9425	0.9991
	MVKUI	0.0788	0.1011	0.1327	0.244	0.7203	0.944
	MVWAT	0.0785	0.1149	0.1543	0.2985	0.7806	0.9624

Power Comparison for Multivariate Non-Symmetric Distributions

The power values $(1-\beta)$ of the tests are obtained by generating random samples from the five-variate Exponential distribution with parameter $(\lambda = 1)$, the five-variate Weibull distribution with parameters $(a = 1, b = 5)$, the five-variate Log-Normal distribution with parameters $(\mu = 1, \sigma^2 = 1)$ and the five-variate Chi-Square distribution with 5 degrees of freedom. Test power is computed under the null hypothesis that assumes a five-dimensional standard normal distribution. In this case, the power values of all tests for the above-mentioned distributions are obtained as "1" for the known mean vector and covariance matrix. To observe the change in power values, the normal distribution parameters are not assumed to be standard in the null hypothesis, and the power values of the tests are recalculated by estimating them from the sample.

In cases where the parameters of the normal distribution are unknown, the Kolmogorov–Smirnov test is replaced by the Lilliefors test. In this case, the critical table of Lilliefors test is used instead of the critical table of KS test. The power comparisons for the remaining tests are obtained using the normal distribution defined by estimating the parameters from the sample in the null hypothesis. The power values computed for various sample sizes ($n = 10, 30, 50, 100, 300, 500$) and are presented in Table 12.

The test power values of all the multivariate tests increase for the multivariate non-symmetric distributions as the sample size increases. Although the test power values for the five-variate Exponential distribution are low at small sample sizes, they approach 1 as the sample sizes increase. In general, the $MVAD$ test shows higher power than the other empirical tests at small sample sizes for the multivariate non-symmetric distributions. In fact, the power values of some tests reach 1 at the sample sizes of 100, 300 and 500.

Real Dataset Apllications

Four different real datasets are selected to demonstrate the feasibility of the proposed multivariate empirical goodness-of-fit tests. These datasets include the sweat data [57], Meteorological data [58], Head Measurement data [59], and the Fisher Iris data [60]. Since both the mean vector and the covariance matrix are available for the sweat data, this enables a full comparison across all test statistics.

On the other hand, the test results are also included for the well-known multivariate normality tests such as

Table 12. Power comparison for the multivariate non-symmetric distributions with unknown mean vector and covariance matrix

		Sample Size					
Distribution	**Test**	**10**	**30**	**50**	**100**	**300**	**500**
Exponential	*MVKS*	0.0677	0.5479	0.9293	1	1	1
	MVCvM	0.1068	0.7931	0.9903	1	1	1
	MVAD	0.1146	0.9075	0.9994	1	1	1
	MVKUI	0.1209	0.7137	0.9699	1	1	1
	MVWAT	0.1152	0.8039	0.9894	1	1	1
Weibull	*MVKS*	0.0627	0.3118	0.6838	0.991	1	1
	MVCvM	0.0913	0.5605	0.8963	0.9997	1	1
	MVAD	0.098	0.7061	0.9741	1	1	1
	MVKUI	0.11	0.5138	0.8479	0.9982	1	1
	MVWAT	0.0987	0.5955	0.9084	0.9996	1	1
Log-normal	*MVKS*	0.098	0.8981	0.9992	1	1	1
	MVCvM	0.1423	0.9779	1	1	1	1
	MVAD	0.1619	0.995	1	1	1	1
	MVKUI	0.1402	0.9378	0.9994	1	1	1
	MVWAT	0.1345	0.9701	1	1	1	1
Chi-square	*MVKS*	0.0985	0.8784	0.999	1	1	1
	MVCvM	0.1565	0.979	1	1	1	1
	MVAD	0.1837	0.9956	1	1	1	1
	MVKUI	0.1646	0.9353	0.9997	1	1	1
	MVWAT	0.1594	0.9734	1	1	1	1

Henze-Zirkler (HEZ) [16], Royston (ROY) [15], Mardia's skewness coefficient (MSC) and Mardia's kurtosis coefficient (MKC) [14] to compare proposed multivariate tests. In these tests, there is no need to distinguish whether the mean and variance are known or not. So, these tests are used only in case of unknown mean and variance.

Multivariate normality test for the sweat data

The sweat data set was obtained by analyzing the sweating results from 20 healthy women [57]. Three components were measured in the analysis: sweat rate, sodium content, potassium content. The variables were respectively characterized by their mean, standard deviation, minimum, and maximum values as follows: (4.64 ± 1.69, 1.5, 8.5), (45.4 ± 14.14, 13.5, 71.6), and (9.97 ± 1.91, 7.1, 14). Also, mean vector ($\boldsymbol{\mu}$) and covariance matrix ($\boldsymbol{\Sigma}$) for this dataset are given as below, respectively [57].

$$\boldsymbol{\mu} = [4 \quad 50 \quad 10], \quad \boldsymbol{\Sigma} = \begin{bmatrix} 3 & 10 & -2 \\ 10 & 200 & -5 \\ -2 & -5 & 3 \end{bmatrix}.$$

The null and the alternative hypothesis for the sweat data are given as follows:

H_0: The sweat data follow the three-variate normal pdf with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

H_1: The sweat data do not follow the three-variate normal pdf with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The three-variate normality test results for the sweat data are given in Table 13.

Since the p values for all tests are greater than $\alpha = 0.05$, no test can reject the H_0 hypothesis. As a result, the sweat data have three-variate normal distribution at 95% confidence level.

Multivariate normality test for the meteorological data

The meteorological dataset was obtained by making measurements on five meteorological variables over a period of 11 years [58]. These variables are rainfall in November and December (in millimeters), average July temperature (in degrees Celsius), rainfall in July (in millimeters), radiation in July (in milliliters of alcohol) and average harvest yield (in quintals per hectare). The variables were respectively characterized by their mean, standard deviation, minimum, and maximum values as follows: (138 ± 44.42, 87.9, 220.9), (17.8 ± 1.28, 15.2, 19.7), (74.4 ±36.69, 1, 140.6), (1242.4 ± 230.03, 902, 1661) and (25.7 ± 2.12, 21.66, 28.4). Since the dataset lacks known values for the mean and covariance, maximum likelihood estimation (MLE) is employed to derive them. The null and the alternative hypothesis for the meteorological data are given as follows:

H_0: The meteorological data follow a five-variate normal distribution.

H_1: The meteorological data do not follow a five-variate normal distribution.

The five-variate normality test results for the meteorological data are given in Table 14.

Except for the $MVKS$ test, the other tests can not reject the H_0 hypothesis. At the 95% confidence level, while the meteorological data does not follow a five-variate normal distribution according to the $MVKS$ test, it follows a five-variate normal distribution according to other tests.

Multivariate normality test for the head measurements data

The head measurements dataset was collected as part of a preliminary study of a possible link between football helmet design and neck injuries [59]. In this data set, there are six variables as head width at widest dimension, head circumference, front to back measurement at eye level, eye to top of head measurement, ear to top of head measurement, and jaw width. The variables were respectively characterized by their mean, standard deviation, minimum, and maximum values as follows: (15.4 ± 0.67, 13.5, 17.5), (58.04 ± 1.88, 54.8, 63.5), (19.91 ± 0.75, 18.5, 21.8), (11.37 ± 1.68, 7.40, 15), (13.96 ± 0.96, 11.10, 16.5 and (12 ± 0.64, 10.5, 13.5). Also, there are 90 observations for each variable.

Table 13. Multivariate normality test results for sweat data

Test	Test Statistic	p	Result
MVKS	0.1283	0.8561	Cannot Reject
MVCvM	0.0567	0.8406	Cannot Reject
MVAD	0.4669	0.7786	Cannot Reject
MVKUI	0.1814	0.9418	Cannot Reject
MVWAT	0.0345	0.8918	Cannot Reject
HEZ	0.4764	0.7402	Cannot Reject
ROY	0.2493	0.9584	Cannot Reject
MSC	8.5075	0.5794	Cannot Reject
MKC	-0.7495	0.4536	Cannot Reject

Table 14. Multivariate normality test results for meteorological data

Test	Test Statistic	p	Result
MVKS	0.27	0.0241*	Rejected
MVCvM	0.1051	0.083	Cannot Reject
MVAD	0.5529	0.1214	Cannot Reject
MVKUI	0.399	0.0743	Cannot Reject
MVWAT	0.0993	0.0806	Cannot Reject
HEZ	0.7872	0.2277	Cannot Reject
ROY	0.8244	0.7753	Cannot Reject
MSC	20.5354	0.9753	Cannot Reject
MKC	-1.4129	0.1577	Cannot Reject

Since the dataset lacks known values for the mean and covariance, maximum likelihood estimation (MLE) is employed to derive them. The null and the alternative hypothesis for the head measurements data are given as follows:

H_0: The head measurements data follow a six-variate normal distribution.

H_1: The head measurements data do not follow a six-variate normal distribution.

The six-variate normality test results for the head measurements data are given in Table 15.

According to Table 15, while $MVKS$, $MVAD$, $MVWAT$, HEZ and MKC tests cannot reject the H_0 hypothesis,

$MVCvM$, $MVKUI$, ROY and MSC tests reject the H_0 hypothesis. At the 95% confidence level, while the head size data set follow a six-variate normal distribution according to the $MVKS$, $MVAD$, $MVWAT$, HEZ and MKC tests, it does not follow a six-variate normal distribution according to the other tests. In addition, the p value calculated for the $MVCvM$ test is quite close to the $\alpha = 0.05$.

Table 15. Multivariate normality test results for head measurements data

Test	Test Statistic	p	Result
MVKS	0.0873	0.0887	Cannot Reject
MVCvM	0.1251	0.0496*	Rejected
MVAD	0.7323	0.0535	Cannot Reject
MVKUI	0.1606	0.0327*	Rejected
MVWAT	0.1193	0.0445*	Rejected
HEZ	0.9405	0.2278	Cannot Reject
ROY	30.2178	0.0000*	Rejected
MSC	84.8693	0.0077*	Rejected
MKC	0.9268	0.3540	Cannot Reject

Multivariate normality test for the Fisher's Iris data

The Fisher Iris dataset is a classic and widely used benchmark in the field of pattern recognition and machine learning. It contains 150 observations of iris flowers, equally distributed among three species: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. Each observation includes four numerical features: sepal length, sepal width, petal length, and petal width, all measured in centimeters [60]. The variables were respectively characterized by their mean, standard deviation, minimum, and maximum values as follows: (5.84 ± 0.83, 4.30, 7.9), (3.06 ± 0.44, 2, 4.4), (3.76 ± 1.77, 1.0, 6.9 and (1.2 ± 0.76, 0.10, 2.5). The dataset is particularly valued for its simplicity, balanced class distribution, and suitability for classification and clustering tasks. Since the mean vector and the covariance matrix of this dataset are unknown, they are estimated using the MLE method. The null and alternative hypotheses for the head measurements data are formulated as follows:

H_0: The head measurements data follow a four-variate normal distribution.

H_1: The head measurements data do not follow a four-variate normal distribution.

The results of the four-variate normality tests applied to the Fisher Iris data are presented in Table 16.

Results obtained for such tests always showed the absence of any deviations from normality, since values of test statistics were below critical values at the significance level used. Agreement between various tests makes the statement of approximate normality of the data more reliable. Such results are consistent with the results found in other scientific sources, including the works of Mardia and Kent [61] and Looney [2].

Table 16. Multivariate normality test results for Fisher Iris data

	Setosa		Versicolor		Virninica	
Test	**Test Statistic**	***p* value**	**Test Statistic**	***p***	**Test Statistic**	***p***
MVKS	0.0755	0.6727	0.0637	0.8809	0.0708	0.7642
MVCvM	0.0463	0.5672	0.0330	0.7959	0.0518	0.4862
MVAD	0.3353	0.5014	0.2388	0.7727	0.4164	0.3241
MVKUI	0.1451	0.5210	0.1158	0.8755	0.1409	0.5763
MVWAT	0.0440	0.5642	0.0324	0.7767	0.0499	0.4692
HEZ	0.9411	0.0564	0.8325	0.2431	0.7505	0.5221
ROY	30.0454	0.0000*	7.8526	0.0848	8.1414	0.0678
MSC	23.1793	0.2801	23.7039	0.2556	24.7257	0.2121
MKC	0.7171	0.1953	-1.0342	0.3010	-0.3384	0.7350

CONCLUSION

In this study, several well-known univariate empirical goodness-of-fit tests were systematically generalized to assess multivariate normality. Critical value tables were constructed for scenarios where the mean vector and covariance matrix are either known or unknown. Simulation results demonstrated that the proposed tests maintain Type I error rates close to the nominal 5% level and exhibit robust statistical power across varying sample sizes and distributional conditions. Among these, multivariate Anderson-Darling test performed the best in terms of power performance especially for small sample cases. These findings establish the robustness and statistical validity of the proposed testing procedure.

Apart from the findings, this paper also makes recommendations regarding which tests should be chosen for particular cases. In this regard, it could be said that the multivariate Anderson-Darling test provides the best power properties and can thus be recommended if there is a need for very sensitive analysis. At the same time, one can use either the multivariate Kuiper test or the multivariate Watson test, since both demonstrate competitive results and have simpler algorithms than the first test.

One of the contributions of this study is that it presents a framework that links univariate goodness-of-fit testing to multivariate goodness-of-fit testing in an elegant way. Most of the existing tests for multivariate normality are either too narrow in scope or very complicated from a computational point of view. The framework presented here generalizes several traditional univariate tests into the multivariate world. This is a new approach not previously seen in the literature.

These tests have been designed such that they can be generally used irrespective of the size and dimension of data. The success of these tests on real world data sets is a clear indication of their applicability in areas like signal/image processing, biomedicine, environment studies, quality control, and reliability engineering, among others.

However, in extremely high dimensions, scalability issues may pose a problem to the algorithm. Further, there might be restrictions on the effectiveness of the variable independence transformation where data sets exhibit complex dependency structures. Future research can focus on the enhancement of computational efficiency and applicability to other types of distribution than normal.

Despite the fact that the empirical analysis in this paper is based on some lower dimensional scenarios only, the presented methodology, especially the independence-based variable transformation technique, is theoretically suitable for any number of dimensions (including $d > 10$). Nevertheless, because of the extensive length of this manuscript, simulations were not conducted for the case of higher dimensions. Future work could focus on the application of the proposed methodology to higher dimensions and its behavior in terms of efficiency and computations. In addition, although this methodology proved to be quite robust with respect to different sample sizes, it is important for users to know that the variability of parameter estimation can cause some sensitivity of the test statistic when $n < 20$.

It should be mentioned that the simulated covariance matrices generated through the synthetic way can miss the complexities inherent in real world data. The use of empirical or more realistic covariance patterns in the design of simulations is a future possibility that may be considered. It should also be mentioned that univariate critical values tables were used as references in the paper. Adaptive critical value approaches can be a future consideration taking into account the dependencies and high dimensionalities involved. Moreover, future studies will include tests using the proposed method under weakly non-normal distribution structures as well as mixed distributional structures which are prevalent in real world conditions but not considered in the present power analysis. Lastly, high dimensional scenarios and scenarios where the number of variables is near or larger than the number of samples ($d \approx n$) are of paramount importance in current statistical applications. This drawback has already been pointed out and constitutes a significant area for further research.

DATA AVAILABILITY STATEMENT

The author confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Genceli M. Kolmogorov-Smirnov, Lilliefors and Shaphiro-Wilk tests for normality. Sigma J Eng Nat Sci 2007;25:306–328.

[2] Looney SW. How to use tests for univariate normality to assess multivariate normality. Am Stat 1995;49:64–70. [CrossRef]

[3] Cochran WG. The χ2 test of goodness of fit. Ann Math Stat 1952;23:315–345. [CrossRef]

[4] Kolmogorov AN. Sulla determinazione empirica di una legge di distribuzione. Inst Ital Attuari Giorn 1933;4:83–91.

[5] Smirnov NV. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bull Math Univ Moscou 1939;2.

[6] Anderson TW, Darling DA. A test of goodness of fit. J Am Stat Assoc 1954;49:765–769. [CrossRef]

[7] Cramér H. On the composition of elementary errors: First paper: Mathematical deductions. Scand Actuar J 1928;1928:13–74. [CrossRef]

[8] von Mises R. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theorestischen Physik. Wien: Franz Deuticke; 1931.

[9] Anderson TW. On the distribution of the two-sample Cramér-von Mises criterion. Ann Math Stat 1962;33:1148–1159. [CrossRef]

[10] Watson GS. Goodness-of-fit tests on a circle. Biometrika 1961;48:109–114. [CrossRef]

[11] Stephens MA. EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc 1974;69:730–737. [CrossRef]

[12] Kuiper NH. Tests concerning random points on a circle. Nederl Akad Wetensch Proc Ser A 1960;63:38–47. [CrossRef]

[13] D'Agostino RB, Stephens MA. Goodness-of-fit techniques. Boca Raton, FL: CRC Press; 1986.

[14] Mardia KV. Measures of multivariate skewness and kurtosis with applications. Biometrika 1970;57:519–530. [CrossRef]

[15] Royston J. Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. J R Stat Soc C Appl Stat 1983;32:121–133. [CrossRef]

[16] Henze N, Zirkler B. A class of invariant consistent tests for multivariate normality. Commun Stat Theory Methods 1990;19:3595–3617. [CrossRef]

[17] Mecklin CJ, Mundfrom DJ. An appraisal and bibliography of tests for multivariate normality. Int Stat Rev 2004;72:123–138. [CrossRef]

[18] Beirlant J, Mason D, Vynckier C. Goodness-of-fit analysis for multivariate normality based on generalized quantiles. Comput Stat Data Anal 1999;30:119–142. [CrossRef]

[19] Liang JJ, Bentler PM. A t-distribution plot to detect non-multinormality. Comput Stat Data Anal 1999;30:31–44. [CrossRef]

[20] Ebner B, Henze N. Tests for multivariate normality-A critical review with emphasis on weighted L^2-statistics. Test 2020;29:845–892. [CrossRef]

[21] Cabaña A, Cabaña EM. Transformed empirical processes and modified Kolmogorov-Smirnov tests for multivariate distributions. Ann Stat 1997;25:2388–2409. [CrossRef]

[22] Justel A, Peña D, Zamar R. A multivariate Kolmogorov-Smirnov test of goodness of fit. Stat Probab Lett 1997;35:251–259. [CrossRef]

[23] Lopes RH, Reid I, Hobson PR. The two-dimensional Kolmogorov-Smirnov test. 2007. [CrossRef]

[24] Chiu SN, Liu KI. Generalized Cramér-von Mises goodness-of-fit tests for multivariate distributions. Comput Stat Data Anal 2009;53:3817–3834. [CrossRef]

[25] Koizumi K, Hyodo M, Pavlenko T. Modified Jarque-Bera type tests for multivariate normality in a high-dimensional framework. J Stat Theory Pract 2014;8:382–399. [CrossRef]

[26] Kurita E, Seo T. Multivariate normality test based on kurtosis with two-step monotone missing data. J Multivar Anal 2022;188:104824. [CrossRef]

[27] Mardia KV. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhya B 1974;36:115–128.

[28] Mardia KV. Assessment of multinormality and the robustness of Hotelling's T^2 test. Appl Stat 1975;24:163–171. [CrossRef]

[29] Small N. Marginal skewness and kurtosis in testing multivariate normality. Appl Stat 1980;29:85–87. [CrossRef]

[30] Dufour JM, Khalaf L, Beaulieu MC. Exact Skewness-Kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models. Oxf Bull Econ Stat 2003;65:891–906. [CrossRef]

[31] Balakrishnan N, Capitanio A, Scarpa B. A test for multivariate skew-normality based on its canonical form. J Multivar Anal 2014;128:19–32. [CrossRef]

[32] Wulandari D, Sutrisno S, Nirwana MB. Mardia's skewness and kurtosis for assessing normality assumption in multivariate regression. Enthusiastic Int J Appl Stat Data Sci 2021;1:1–6. [CrossRef]

[33] González-Estrada E, Villaseñor JA, Acosta-Pech R. Shapiro-Wilk test for multivariate skew-normality. Comput Stat 2022;37:1985–2001. [CrossRef]

[34] Fan Y. Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. J Multivar Anal 1997;62:36–63. [CrossRef]

[35] Henze N, Visagie J. Testing for normality in any dimension based on a partial differential equation involving the moment generating function. Ann Inst Stat Math 2020;72:1109–1136. [CrossRef]

[36] Dörr P, Ebner B, Henze N. A new test of multivariate normality by a double estimation in a characterizing PDE. Metrika 2021;84:401–427. [CrossRef]

[37] Al-Labadi L, Fazeli Asl F, Saberi Z. A necessary Bayesian nonparametric test for assessing multivariate normality. Math Methods Stat 2021;30:64–81. [CrossRef]

[38] Arnastauskaitė J, Ruzgas T, Bražėnas M. A new goodness of fit test for multivariate normality and comparative simulation study. Mathematics 2021;9:3003. [CrossRef]

[39] Bilodeau M, Brenner D. Theory of multivariate statistics. Berlin: Springer; 2008.
[40] Chen H, Xia Y. A normality test for high-dimensional data based on the nearest neighbor approach. J Am Stat Assoc 2023;118:719–731. [CrossRef]
[41] Song Y, Zhao X. Normality testing of high-dimensional data based on principle component and Jarque-Bera statistics. Stats 2021;4:216–227. [CrossRef]
[42] Móri TF, Székely GJ, Rizzo ML. On energy tests of normality. J Stat Plan Inference 2021;213:1–15. [CrossRef]
[43] Feng Z, Zhao M, Qi X. Multivariate normality test based on Shannon's entropy. Statistics 2025;213:1–20. [CrossRef]
[44] Heine J, Fowler EE, Berglund A, Schell MJ, Eschrich S. Techniques to produce and evaluate realistic multivariate synthetic data. Sci Rep 2023;13:12266. [CrossRef]
[45] Al-Zwainy F, Al-Marsomi M. Structural equation modeling of critical success factors in the programs of development regional. J Proj Manag 2023;8:119–132. [CrossRef]
[46] Risan HK, Serhan FM, Al-Azzawi AA. Management of a typical experiment in engineering and science. AIP Conf Proc 2024. [CrossRef]
[47] Kesemen O, Tiryaki BK, Tezel Ö, Özkul E. A new goodness of fit test for multivariate normality. Hacet J Math Stat 2021;50:872–894. [CrossRef]
[48] Gut A. An intermediate course in probability. New York, NY: Springer; 2009. [CrossRef]
[49] Rosenblatt M. Remarks on a multivariate transformation. Ann Math Stat 1952;23:470–472. [CrossRef]
[50] Liang JJ, Fang KT, Hickernell F, Li R. Testing multivariate uniformity and its applications. Math Comput 2001;70:337–355. [CrossRef]
[51] Dettmann CP, Georgiou O. Product of n independent uniform random variables. Stat Probab Lett 2009;79:2501–2503. [CrossRef]
[52] Lewis C. Demand forecasting and inventory control. London: Routledge; 2012. [CrossRef]
[53] Lazzeri F. Machine learning for time series forecasting with Python. Hoboken, NJ: John Wiley & Sons; 2020. [CrossRef]
[54] Truong NS, Ngo NT, Pham AD. Forecasting time-series energy data in buildings using an additive artificial intelligence model for improving energy efficiency. Comput Intell Neurosci 2021;2021:6028573. [CrossRef]
[55] Uhm T, Yi S. A comparison of normality testing methods by empirical power and distribution of P-values. Commun Stat Simul Comput 2023;52:4445–4458. [CrossRef]
[56] Terán-García E, Pérez-Fernández R. A robust alternative to the Lilliefors test of normality. J Stat Comput Simul 2024;94:1494–1512. [CrossRef]
[57] Johnson RA, Wichern DW. Applied multivariate statistical analysis. London, UK: Pearson; 2014. [CrossRef]
[58] Bibby J, Kent J, Mardia K. Multivariate analysis. London: Academic Press; 1979.
[59] Rencher AC. Methods of multivariate analysis. Hoboken, NJ: John Wiley & Sons; 2002. [CrossRef]
[60] Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugenics 1936;7:179–188. [CrossRef]
[61] Mardia K, Kent J. Rao score tests for goodness of fit and independence. Biometrika 1991;78:355–363. [CrossRef]