

Journal of Hunan University (Natural Sciences)

Vol. 53 No. 5

May 2026

Available online at

<https://jionuns.com>



Open Access Article

 <https://doi.org/10.55463/issn.1674-2974.53.5.3>

Robust Extensions of Hotelling's T^2 and Related Multivariate Tests: A Systematic Review of Simulation Evidence

Siti Nor Ain Zainon ^{1*}, Shamsuritawati Sharif ²

¹Centre for Matriculation, Universiti Islam Antarabangsa Sultan Abdul Halim Mu'adzam Shah, Kedah, Malaysia,

²School of Quantitative Sciences, Universiti Utara Malaysia, Kedah, Malaysia,

* Corresponding author: sitinorain@unishams.edu.my

Article History:

Received: March 20, 2026

Revised: April 29, 2026

Accepted: May 15, 2026

Published: May 29, 2026

Abstract: Hotelling's T^2 test is a foundational multivariate statistical method widely used for hypothesis testing involving mean vectors. However, its classical formulation relies on strong assumptions, including multivariate normality, low dimensionality relative to sample size, and the absence of outliers. In recent decades, a growing body of literature has proposed robust extensions of Hotelling's T^2 test to address violations of these assumptions, particularly in high-dimensional and contaminated data settings. Despite rapid methodological development, simulation-based evidence on the performance of these robust extensions has not yet been systematically synthesized. Guided by the PRISMA framework, this study conducts a systematic literature review of simulation studies published between 1974 and 2025 that examine robust variants of Hotelling's T^2 test and related multivariate tests. Searches were conducted in Scopus and Web of Science, resulting in 35 eligible studies for qualitative synthesis. Using thematic analysis, three major themes were identified: robustness in high-dimensional and small-sample regimes; robustness to distributional deviations and outlier contamination; and calibration and computational robustness through resampling and adaptive procedures. The review highlights the consistent performance advantages of robust methods over the classical Hotelling's T^2 test under assumption violations, identifies methodological gaps in simulation design, and provides recommendations for future research.



Copyright: © 2026 by the authors. Licensee JHU

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>)

Keywords: Hotelling's T^2 test; robust statistics; high-dimensional data; multivariate tests; simulation study; systematic literature review.

Hotelling's T^2 及相关多元检验的稳健扩展：模拟证据的系统综述

摘要：

Hotelling's T^2 检验是一种基础性的多元统计方法，广泛用于涉及均值向量的假设检验。然而，其经典形式依赖于较强的假设条件，包括多元正态性、相对于样本量较低的维度以及不存在异常值。近几十年来，越来越多的文献提出了Hotelling's T^2 检验的稳健扩展方法，以应对这些假设被违反的情况，尤其是在高维数据和受污染数据情境下。尽管方法学发展迅速，关于这些稳健扩展方法性能的模拟研究证据尚未得到系统综合。本文以PRISMA框架为指导，对1974年至2025年间发表的模拟研究进行系统文献综述，重点考察 Hotelling's T^2 检验的稳健变体及相关多元检验方法。研究在Scopus和Web of Science数据库中进行检索，最终纳入35项符合条件的研究进行定性综合。通过主题分析，识别出三大主要主题：高维和小样本情境下的稳健性；对分布偏离和异常值污染的稳健性；以及通过重抽样和自适应程序实现的校准与计算稳健性。本综述强调，在假设被违反的情况下，稳健方法相较于经典Hotelling's T^2 检验具有一致的性能优势，同时识别了模拟设计中的方法学缺口，并为未来研究提出建议。

关键词：

Hotelling's T^2 检验；稳健统计；高维数据；多元检验；模拟研究；系统文献综述。

1. Introduction

The classical Hotelling's T^2 test has long served as the multivariate analogue of the univariate t -test and remains a cornerstone of multivariate statistical inference [1]. Its applications span diverse fields, including quality control, bioinformatics, finance, and the social sciences. Nevertheless, Hotelling's T^2 test is known to have important limitations when its underlying assumptions are violated [2]. These limitations are particularly pronounced in modern data environments characterized by high dimensionality, where p is comparable to or exceeds n , as well as by non-normal distributions [3] and the presence of outliers [4,5].

Recent advances in statistical theory have led to the development of numerous extensions and alternatives to Hotelling's T^2 test. These include regularized covariance-based tests [6–8], robust location and scatter estimators [9,10], nonparametric rank-based methods, and resampling-calibrated procedures. Many of these methods have been evaluated primarily through simulation studies, which allow researchers to assess empirical size, power, and robustness under controlled violations of classical assumptions.

Despite the abundance of simulation-based evaluations, the literature remains fragmented.

Individual studies often focus on specific aspects of robustness, such as high dimensionality or contamination, making it difficult for applied researchers to identify which extensions are most effective across a range of realistic conditions. Moreover, existing reviews tend to provide descriptive summaries of methods without systematically integrating simulation findings to reveal cross-cutting patterns and comparative insights. Therefore, a systematic synthesis of simulation evidence is needed. This study addresses this gap by conducting a systematic literature review of simulation-based studies on robust extensions of Hotelling's T^2 test and related multivariate tests. It also provides a thematic synthesis that integrates findings across studies to identify key performance patterns, methodological strengths, and practical implications.

2. Material and methods

This section first outlines the reporting standard adopted in this study, namely the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [11]. It then presents the formulation of the research questions, followed by a detailed description of the systematic search procedures, including the identification, screening, and eligibility stages. The

section also explains the quality appraisal process and the methods used for data extraction and analysis.

2.1 PRISMA Guidelines

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA provides a transparent, structured, and reproducible framework for identifying, screening, assessing the eligibility of, and synthesizing relevant literature, thereby improving the methodological rigor and reporting quality of systematic reviews. A PRISMA flowchart illustrating the study selection process is presented in Figure 1.

As an internationally recognized reporting standard, PRISMA guides researchers in formulating clear research questions, establishing explicit inclusion and exclusion criteria, and systematically screening large volumes of scientific literature over a specified period [12]. In this study, the PRISMA framework informed the design and implementation of a rigorous and transparent search strategy aimed at identifying empirical and simulation-based studies on robust extensions of Hotelling's T^2 test and related multivariate test statistics. It also helped ensure consistency, reproducibility, and completeness throughout the review process [13–15].

2.2 Research Questions

To guide the review, the following research questions were formulated:

1. What types of robust extensions of Hotelling's T^2 test have been evaluated through simulation studies?
2. Under which data conditions do robust Hotelling-type tests outperform the classical Hotelling's T^2 test?
3. What methodological trends and gaps emerge from recent simulation-based evidence?

2.3 Systematic Searching Strategy

The systematic searching strategy consisted of three phases: identification, screening, and eligibility.

2.3.1 Identification

The identification stage was conducted to broaden and refine the search vocabulary, thereby maximizing the likelihood of retrieving all relevant studies. This step is widely recommended in systematic review methodology because keyword enrichment reduces the risk of missing relevant records due to terminological variation across disciplines and databases [16,17].

Following established guidance for systematic reviews, the core concepts, namely Hotelling's T^2 test, robustness, and evaluation, were expanded by systematically identifying

related terms and synonyms. This process involved using an online thesaurus, examining keywords reported in influential prior studies, consulting database-generated keyword suggestions, and incorporating expert feedback to improve both coverage and precision [18].

The final set of enriched keywords was then operationalized into structured search strings using standard database search techniques, including Boolean operators, phrase searching, truncation, wildcards, and field codes. This approach is consistent with PRISMA-informed expectations for transparent and reproducible reporting of search procedures [19].

The systematic search was conducted in Scopus and Web of Science in November 2025 using the full search strings presented in Table 1. In addition, manual searching was performed by examining potentially relevant records identified within both databases, including studies retrieved through reference list screening and related-record functions. Through this process, 45 potentially relevant articles were identified. After the removal of one duplicate record, 44 articles were retained for the subsequent stages of the review.

Table 1. The search string.

Database	Date of Search	Search String
Scopus	16 Nov 2025	("Hotelling T ² " OR "Hotelling T ² " OR "Hotelling's T-squared" OR "Hotelling T-squared" OR "Hotelling's T ² " OR "Hotelling T ² ") AND (robust OR "robust statist*" OR "robust method*" OR "robust test*" OR "robust estimat*" OR "robust infer*") AND "evaluation" ("Hotelling T ² " OR "Hotelling T ² " OR "Hotelling's T-squared" OR "Hotelling T-squared" OR "Hotelling's
Web of Science	16 Nov 2025	"Hotelling T-squared" OR "Hotelling T-squared" OR "Hotelling's

T2" OR
 "Hotelling
 T^2") AND
 (robust OR
 "robust
 statist*" OR
 "robust
 method*" OR
 "robust
 test*" OR
 "robust
 estimat*" OR
 "robust
 infer*")
 AND
 "evaluation"

2.3.2 Screening

Following the identification stage, a screening process was conducted on the 44 retrieved records. Consistent with established systematic review protocols, the screening criteria included literature type, language, and year of publication, as summarized in Table 2.

Peer-reviewed journal articles were prioritized as the primary source of evidence because they typically report original empirical or simulation-based findings, which are essential for ensuring the analytical rigor of a systematic literature review [16,20]. Only English-language publications were included to ensure accurate interpretation and consistency in data extraction, a practice commonly adopted in systematic reviews to minimize potential misinterpretation arising from translation constraints [18].

In determining the temporal scope, the review was guided by the maturity of the research field, which requires a sufficiently large and developed body of literature to enable meaningful synthesis and pattern identification [21]. Accordingly, the review encompassed studies published over a 51-year period, from 1974 to 2025, thereby capturing both the foundational development of Hotelling’s T² test and its subsequent robust extensions. During the screening process, one article was excluded because it was published in German and did not meet the language inclusion criterion.

Table 2. The selection criterion is searching

Criterion	Inclusion	Exclusion
Language	English	Non-English
Timeline	1974 – 2025	< 1974
Literature type	Article, Conference paper, Review	Book
Publication Stage	Final	In Press
Subject Area	All	None

2.3.3 Eligibility

The third phase of the systematic review involved eligibility assessment, which served as a second-level screening process to ensure that all records retained after the initial screening fully met the predefined inclusion criteria. This phase followed a structured set of inclusion and exclusion criteria to enhance transparency and reproducibility.

Each record was re-evaluated for relevance based on its title and abstract. When ambiguity remained, the full text was consulted to confirm eligibility in a transparent and reproducible manner consistent with PRISMA 2020 guidance [19]. The review process adhered to PRISMA 2020 reporting standards; however, no formal protocol was pre-registered.

Full-text assessment also enabled the reviewers to verify critical methodological features, particularly whether each study reported simulation-based evidence relevant to robust extensions of Hotelling’s T² test and related multivariate tests. This step strengthened the internal validity of the synthesis.

Following this detailed appraisal, eight articles were excluded for one or more of the following reasons: (i) the full text was irretrievable; (ii) the study lacked a DOI or sufficient bibliographic information; (iii) the study was not aligned with the research objective based on the predefined criteria; or (iv) the publication fell outside the eligible time frame, that is, prior to 1974. The remaining 35 studies satisfied the eligibility criteria and were retained for quality appraisal and inclusion in the final qualitative synthesis (see Figure 1), in line with established methodological guidance for systematic literature reviews.

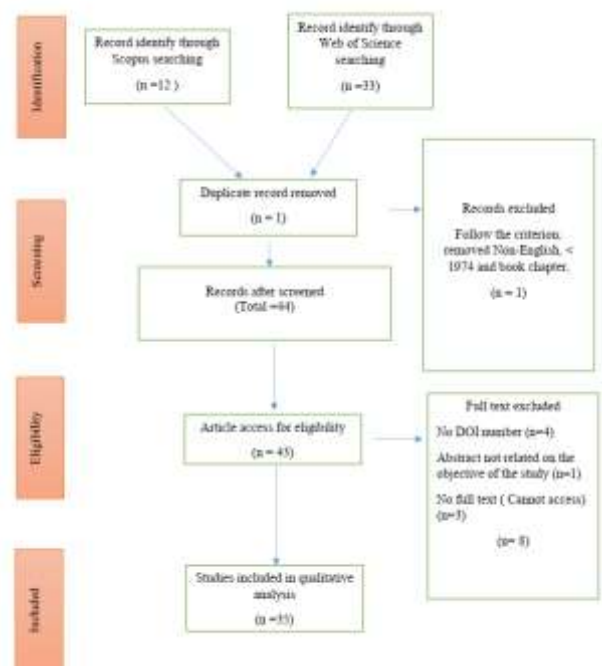


Figure 1. Systematic review process (developed by author)

2.4 Data Extraction and Analysis

Following the eligibility stage, the remaining articles were subjected to systematic assessment and analysis. In addition, a quality assessment of the included studies was conducted to enhance the methodological rigor of the review. Given the simulation-based nature of the selected studies, a simplified appraisal framework was adopted. Each study was evaluated based on the following criteria: (i) clarity of the simulation design; (ii) appropriateness of performance metrics, such as size, power, and robustness; (iii) transparency in reporting experimental settings, including sample size, dimensionality, and distributional assumptions; and (iv) consistency and interpretability of the results. The quality assessment was used to support the interpretation of the findings rather than as a basis for exclusion.

Prior to analysis, all relevant information was systematically extracted from the selected studies, guided explicitly by the research questions of the review. Because this systematic review focused on primary empirical and simulation-based evidence, the data extraction process initially concentrated on three core sections of each article: the abstract, results, and discussion. These sections most directly report methodological design, simulation settings, and key findings. Subsequently, other sections of the articles were examined to identify additional information relevant to the objectives of the review. All extracted data were organized into a structured table to ensure transparency and consistency during the synthesis process [21].

A qualitative synthesis was then conducted using thematic analysis, which is widely employed in systematic reviews to identify, analyze, organize, and report recurring patterns within extracted data [22]. Thematic analysis offers several advantages for evidence synthesis: it enables the comparison of diverse perspectives, facilitates the identification of similarities and differences across studies, and allows unanticipated insights to emerge while maintaining a systematic and transparent analytical structure [23].

Through iterative examination of the extracted data, patterns and relationships were identified and grouped into higher-order categories. As a result of this process, three overarching themes were generated: (i) high-dimensional robustness; (ii) distributional and outlier robustness; and (iii) resampling and calibration. These themes form the basis for organizing and interpreting the simulation-based evidence on robust extensions of Hotelling's T^2 test and related multivariate tests.

3. Result and Finding

The thematic synthesis resulted in three main themes, reflecting dominant research directions in the simulation literature.

3.1 Theme 1: Robustness to High-Dimensionality and Small-Sample Regimes ($p \geq n$)

Most studies have focused on adapting Hotelling's T^2 test for high-dimensional contexts where the sample covariance matrix is singular or unstable [24,25]. Simulation evidence consistently shows that regularized shrinkage-based and shrinkage-based extensions outperform classical Hotelling's T^2 test in controlling Type I error and achieving higher power [26,27]. Block and omnibus test statistics further improve performance under complex correlation structures [28,29]. These findings underscore the shortcomings of classical Hotelling's T^2 test in modern high-dimensional settings.

3.2 Theme 2: Robustness to Distributional Deviations and Outlier Contamination

Another dominant theme relates to robustness to non-normality and outliers [30]. Simulation studies comparing classical Hotelling T^2 test with robust alternatives such as M-estimators, minimum covariance determinants (MCDs), and rank-based tests have revealed significant improvements in empirical size control under heavy-tailed distributions and contamination [31]. Although power loss is sometimes observed under clean normal data, the robust method consistently outperforms classical Hotelling T^2 test in contaminated scenarios.

3.3 Theme 3: Calibration and Computational Robustness via Resampling and Adaptive Procedures

The third theme highlighted the role of permutation tests, bootstrap calibration, and adaptive tuning strategies. Rather than modifying the test statistics alone, these approaches improved finite-sample inference by aligning empirical and nominal significance levels. Simulation evidence showed that resampling-calibrated tests provided stable performance across mixed stress conditions, although increased computational cost remained a key limitation [32].

Based on the source materials provided, the following table summarizes the included studies and their relevance to the three specified robustness themes.

Table 3. Distribution of Included Simulation Studies Across Robustness Themes

Number	Author(s)	Year	Test/Method Type	Theme 1: High-Dimensional Robustness	Theme 2: Distributional & Outlier Robustness	Theme 3: Resampling & Calibration
1	Ahsan, Mashuri, & Khusna	2018	Robust Hotelling's T^2 control chart (SDCM)		✓	✓
2	Wang & McDermott	1998	Conditional Hotelling's T^2 -type test		✓	
3	Everitt	1979	Monte Carlo robustness study of T^2 tests		✓	
4	Sivabhaskar et al.	2021	T^2 statistic for Dose-Volume Histograms		✓	
5	Gupta & Kabe	1999	T^2 test on normal mixture populations		✓	
6	Chou, Mason, & Young	1999	T^2 based on alternative covariance estimators		✓	
7	Williams, Sullivan, & Birch	2009	T^2 via successive differences covariance estimator		✓	
8	Srivastava & Awan	1982	Robustness of T^2 under mixture populations		✓	
9	Srivastava	1983	T^2 distribution under contaminated normals		✓	
10	Tiku & Balakrishnan	1988	Robust Hotelling-type T^2 (MML estimators)		✓	
11	Balci	1982	Two-sample T^2 for simulation model validation		✓	
12	Keriya	1981	Robustness property of Hotelling's T^2 test		✓	
13	Cedeño Viteri et al.	2012	Nearest In-Control Neighbor (NICN) T^2 approach	✓	✓	
14	Carvalho Nascimento et al.	2025	Bootstrap tests for two independent mean vectors	✓	✓	✓
15	Dong et al.	2016	Shrinkage-based diagonal Hotelling's T^2 tests	✓	✓	
16	Mehmood	2016	T^2 based variable selection in PLS regression	✓	✓	✓
17	Khatun et al.	2018	Adaptive double sampling Hotelling's T^2 charts			✓
18	Al-Labadi et al.	2022	Bayesian Hotelling's T^2 test (relative belief ratio)		✓	✓
19	Haddad, Syed-Yahaya, & Alfaro	2012	Robust Hotelling's T^2 charts (Winsorized MOM)		✓	✓

Number	Author(s)	Year	Test/Method Type	Theme 1: High-Dimensional Robustness	Theme 2: Distributional & Outlier Robustness	Theme 3: Resampling & Calibration
20	Zwick	1986	Rank and Normal Scores alternatives to T^2		✓	✓
21	Williams et al.	2006	T^2 via successive differences covariance estimator		✓	✓
22	Makinde & Omotoso	2020	Robust diagonalized and regularized T^2 tests	✓	✓	
23	More et al.	2024	T^2 for anomaly detection in control systems		✓	✓
24	Amanullah, Yu, & Sun	2018	Multivariate differential expression by T^2 (MDEHT)		✓	✓
25	Liu & Shao	2013	Global test based on maximum of T^2 statistics	✓	✓	
26	Nescolarde et al.	2016	Paired one-sample Hotelling's T^2 test		✓	
27	Li et al.	2020	Adaptable RHT (ARHT) ridge-regularized test	✓	✓	✓
28	Algina & Oshima	1990	Robustness of independent samples T^2		✓	
29	Kharbach et al.	2016	Robust variant multivariate control charts		✓	
30	Pardo	2011	T^2 Test with normal mixture populations		✓	✓
31	Lix & Keselman	2002	Robust alternatives to Hotelling's T^2		✓	
32	Perlman	2019	Parsimonious variable selection for T^2 test		✓	
33	Salmanpour et al.	2021	High Dimensional Hotelling T^2 Test	✓	✓	✓
34	Romer & Richards	2013	Monotone incomplete multivariate normal T^2		✓	✓
35	Park et al.	2021	Multivariate rank-based MDR (MR-MDR)	✓	✓	✓

Table 3 summarizes the distribution of 35 simulation-based studies on Hotelling's T^2 test and related multivariate tests across the three robustness themes identified in this systematic review: high-dimensional robustness, distributional and outlier robustness, and resampling and calibration robustness. Critical inspection of the table reveals a strong dominance of studies addressing distributional and outlier robustness (Theme 2). Most of the included studies, which encompass both early foundational work and more recent applications, focus on assessing the sensitivity of Hotelling's T^2 test to deviations from multivariate normality, including mixed distributions, heavy tails, and contamination. This dominance reflects the historical development of the literature, where robustness was initially conceptualized primarily in terms of resistance to non-normality and influential observations, typically assessed through Monte Carlo simulation designs.

In contrast, high-dimensional robustness (Theme 1) is represented by a smaller but more recent group of studies, with most contributions appearing after 2010 and increasing after 2020. These studies explicitly address the breakdown of the classic Hotelling's T^2 test when the number of variables approaches or exceeds the sample size, a setting in which the sample covariance matrix becomes singular or unstable. As shown in Table 3, the methods included in this theme often rely on regularization, shrinkage, diagonalization, or variable selection strategies. Simulation evidence from these studies consistently shows that regularized or adaptable Hotelling-type tests far outperform classical statistics in terms of Type I error control and power under the condition $p \geq n$. The growing presence of this theme highlights the increasing relevance of high-dimensional data across application domains and confirms high dimensionality as a key driver of contemporary methodological innovation in multivariate testing [28,33].

Robustness of resampling and calibration (Theme 3) appears as the least represented theme in Table 3, yet it plays a critical methodological role. Only a limited subset of studies explicitly incorporates bootstrap procedures, permutation tests, or adaptive calibration mechanisms. Importantly, all studies classified under Theme 3 also overlap with Themes 1 and/or 2, indicating that resampling-based strategies are typically used as complementary tools rather than stand-alone solutions. The simulation findings reported in this study suggest that calibration techniques can significantly improve finite-sample performance, especially in complex settings characterized by high co-dimensionality and non-normality. However, their relatively sparse representation underscores a major gap in the

literature, especially given the increasing availability of computational resources and the demonstrated effectiveness of resampling methods in aligning empirical and nominal significance levels [34,35].

An important insight that emerges from Table 3 is that robustness in multivariate tests is inherently multidimensional rather than singular. Several studies cover a variety of themes, reflecting the shift towards integrated frameworks that incorporate covariance regularization, robust estimation, and calibration strategies. At the same time, the table clearly shows that no single methodological approach dominates all three dimensions of robustness. Some methods exhibit robust performance in high-dimensional settings but remain sensitive to contamination, while others provide excellent resistance to outliers at the expense of increased computational complexity. This pattern reinforces the conclusion that the selection of robust Hotelling-type tests must be context-dependent, guided by the underlying data structure, risk of contamination, and computational constraints. Overall, Table 3 provides a strong empirical justification for the thematic framework adopted in this review and highlights both the progress made and methodological gaps that require further investigation.

4. Discussion

This study aims to systematically synthesize the existing simulation-based literature on robust extensions of Hotelling's T^2 test and related multivariate tests, with particular emphasis on methodological responses to the challenges posed by modern data environments. Like many classical statistical procedures, Hotelling's T^2 test was originally developed under restrictive assumptions, including low dimensionality, multivariate normality, and the absence of anomalous observations. However, these assumptions are frequently violated in contemporary applications, thereby necessitating the development of robust testing procedures capable of maintaining valid inference under such conditions [28]. Based on a systematic review of studies retrieved from major academic databases, this study synthesizes evidence from 35 simulation-based articles that explicitly evaluate the performance of robust Hotelling-type tests under a variety of adverse data conditions. The thematic synthesis reveals three overarching robustness themes: high-dimensional robustness, distributional and outlier robustness, and resampling and calibration robustness. Collectively, these themes represent the dominant methodological directions in this field. Their prominence reflects a growing recognition of the limitations of the classical Hotelling's T^2 test and the need for alternative procedures capable of controlling Type I error and maintaining statistical power when standard assumptions are violated [32,33]. The simulation

evidence consistently shows that the classical Hotelling's T^2 test suffers from substantial size distortion and power degradation in high-dimensional or contaminated settings, whereas robust extensions offer marked improvements in empirical performance.

With respect to high-dimensional robustness, the reviewed studies show a strong research focus on scenarios in which the number of variables is comparable to or exceeds the sample size. In such settings, the sample covariance matrix becomes singular or unstable, making the classical Hotelling's T^2 test infeasible. Adjusted covariance estimators, shrinkage-based tests, and block-type or omnibus procedures have been shown to effectively address this challenge by stabilizing covariance estimates and increasing power across a variety of dependence structures and signal configurations [29]. However, although these approaches improve stability and feasibility in high-dimensional settings, they often rely on tuning parameters and structural assumptions, which may limit their robustness under model misspecification. In contrast, alternative approaches that do not depend heavily on covariance estimation may offer greater flexibility but at the cost of reduced efficiency. The prevalence of this theme underscores the central role of high-dimensional data in motivating recent methodological innovations in multivariate hypothesis testing.

The second theme, distributional and outlier robustness, highlights the vulnerability of Hotelling's T^2 test to deviations from multivariate normality and the presence of influential observations. Simulation studies in this category typically use heavy-tailed distributions, contamination models, and leverage-point outliers to evaluate test performance. These findings consistently show that robust location and scatter estimators, such as M-estimators, the minimum covariance determinant (MCD), and its adjusted variants, as well as rank- and sign-based methods, substantially improve empirical size control under contamination while maintaining reasonable power [28,32]. Nevertheless, these robust methods may exhibit lower efficiency when classical assumptions hold, and some estimators, such as MCD, can be computationally demanding, particularly in higher dimensions. Compared with parametric approaches, nonparametric and rank-based methods provide distribution-free advantages but may sacrifice statistical power under ideal conditions. These results are particularly relevant for applied researchers working with real-world data, where ideal distributional assumptions are rarely met.

The third theme, resampling and calibration robustness, reflects the growing recognition that modifying the test statistic alone may not be sufficient to ensure reliable finite-sample inference. Studies

under this theme emphasize permutation tests, bootstrap calibration, and adaptive tuning strategies to align empirical significance levels with nominal thresholds. Simulation evidence suggests that such calibration techniques are particularly effective in complex settings involving combinations of high dimensionality and non-normality. However, compared with direct modification approaches, resampling-based methods tend to be computationally intensive and may introduce additional variability depending on the resampling scheme and the number of iterations used. Although they offer flexibility and broad applicability, their scalability remains a key limitation in large-scale or real-time applications.

An important insight emerging from this review is that robustness in multivariate testing is inherently multifaceted rather than singular. Some methods provide strong protection against high dimensionality but remain sensitive to contamination, whereas others are highly robust to outliers but require substantial computation. This pattern highlights a fundamental trade-off between robustness, efficiency, and computational cost across different methodological approaches. Therefore, the selection of an appropriate robust Hotelling-type test should be guided by the specific characteristics of the data and the objectives of the statistical inference. From this perspective, the literature shows that no single method uniformly dominates across all scenarios, reinforcing the importance of simulation-based benchmarking in methodological development [28].

Finally, this review distinguishes between direct and indirect robustness strategies. Direct strategies explicitly modify the test statistic or covariance structure through techniques such as regularization or robust estimation. Indirect strategies, by contrast, improve the reliability of inference through resampling, calibration, or adjustment procedures without changing the underlying statistic. Direct methods often provide more interpretable and computationally efficient solutions, whereas indirect methods offer greater flexibility in handling complex data conditions. These approaches should therefore be regarded as complementary rather than competing strategies in modern robust multivariate inference.

Overall, the findings of this systematic review demonstrate a clear shift in the literature toward a robustness-oriented framework for multivariate testing. Although significant progress has been made, future simulation studies should integrate multiple sources of data complexity within unified experimental designs, adopt standardized benchmarking practices, and report computational considerations more transparently. Such improvements would further advance the methodological maturity and practical applicability of robust extensions of Hotelling's T^2 test.

5. Future Research Directions

One of the primary objectives of conducting a systematic literature review is to identify prevailing research patterns and, importantly, to highlight areas where further investigation is needed. Although this review identified a growing number of simulation-based studies on robust extensions of Hotelling's T^2 test and related multivariate tests, the overall volume of high-quality empirical evidence remains limited relative to the complexity of modern data environments. The systematic search and screening process resulted in a final set of 35 eligible studies, suggesting that considerable scope exists for expanding research in this field, particularly through comprehensive and comparative simulation designs.

A key direction for future research involves the evaluation of robustness under more realistic and compounded data conditions. Most existing simulation studies focus on isolated challenges, such as high dimensionality, distributional deviations, or outlier contamination, while relatively few examine their combined effects. Given that real-world datasets often exhibit multiple sources of complexity simultaneously, future studies should assess whether currently proposed robust Hotelling-type tests remain effective under such conditions. Additionally, greater attention should be paid to issues of computational scalability and efficiency, especially for resampling-based and adaptively calibrated procedures, which may become impractical as dimensionality and sample size increase. Finally, future research should broaden the scope of performance evaluation and methodological development. Beyond empirical size and power, simulation studies should consider robustness to tuning parameter choices, sensitivity to covariance structures, and stability under model misspecification. There is also potential for developing hybrid and adaptive testing frameworks that integrate multiple robustness strategies, such as regularisation, robust estimation, and resampling-based calibration. Addressing these directions will help avoid methodological "specialisation traps" and contribute to the development of more versatile and practically applicable robust extensions of Hotelling's T^2 test.

6. Conclusion

Hotelling's T^2 test remains one of the most influential multivariate statistical procedures and continues to serve as a benchmark for testing mean vectors across a wide range of scientific disciplines. However, like many classical statistical methods, its traditional formulation faces significant challenges in contemporary data environments characterized by high dimensionality, deviations from normality, and the presence of outliers. As a result, the development of robust extensions and related multivariate tests has

become increasingly important for ensuring valid and reliable inference in modern applications.

The findings of this study are drawn from a systematic synthesis of simulation-based evidence examining robust extensions of Hotelling's T^2 test and related multivariate tests. The review identifies three dominant robustness-oriented themes: high-dimensional robustness, distributional and outlier robustness, and resampling and calibration robustness. Collectively, these themes reflect the major directions of methodological development in this field. These robust testing strategies offer practical guidance to researchers and applied statisticians by highlighting alternative approaches that consistently outperform the classical Hotelling's T^2 test under assumption violations, particularly in high-dimensional and contaminated data settings.

However, the relative performance of these methods varies depending on the underlying data conditions. Methods designed for high-dimensional settings tend to improve stability and power but may rely on tuning parameters and structural assumptions. Robust estimators for outlier resistance enhance reliability under contamination but may reduce efficiency when classical assumptions hold. Similarly, resampling-based approaches offer flexibility and improved finite-sample inference, yet they often involve substantial computational costs and scalability limitations. These observations highlight inherent trade-offs between robustness, efficiency, and computational feasibility across different methodological approaches.

Despite progress in simulation-based research, several challenges remain. These include the limited number of studies that jointly consider multiple sources of data complexity, such as high dimensionality combined with heavy-tailed distributions and outliers, as well as the lack of standardized simulation benchmarks and transparent reporting of computational costs and tuning procedures. In addition, many simulation designs rely on simplified or idealized data-generating mechanisms, which may limit the generalizability of findings to real-world applications where multiple complexities coexist.

Therefore, future research should develop simulation designs that incorporate more realistic data-generating mechanisms, examine the scalability of robust procedures, and evaluate their performance across multiple application domains. Addressing these gaps is essential for advancing the practical applicability and methodological maturity of robust Hotelling-type multivariate testing methodologies.

Declarations

Author Contributions

Conceptualization and study design were carried out by the authors. The study methodology was developed

based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines. The systematic search strategy consisted of three phases: identification, screening, and eligibility. Data extraction and analysis were conducted collaboratively by the authors. All authors contributed to the writing, review, and approval of the final manuscript.

Data Availability Statement

The data presented in this study are available from the corresponding author upon reasonable request.

Funding

This study was funded by the researchers from Universiti Utara Malaysia and Universiti Islam Antarabangsa Sultan Abdul Halim Mu'adzam Shah.

Acknowledgements

The authors gratefully acknowledge Universiti Utara Malaysia for its institutional support and assistance in facilitating this research. The authors also appreciate the technical and administrative support provided throughout the completion of this study. Additional guidance and resources provided by the university are sincerely valued and contributed significantly to the successful preparation of this manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this manuscript. In addition, all ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, duplicate publication and/or submission, and redundancy, have been fully observed by the authors.

References

- [1] D. C. Montgomery, *Introduction to Statistical quality control*. John Wiley & Sons, Inc., 2013.
- [2] M. Arslan, U. Shahzad, A. Yeganeh, H. Zhu, J. Malela-Majika, and S. Ahmad, "A Robust L-Comoments Covariance Matrix-Based Hotelling's T^2 Control Chart for Monitoring High-Dimensional Non-Normal Multivariate Data in the Presence of Outliers," *Qual. Reliab. Eng. Int.*, vol. 41, no. 7, pp. 3308–3317, Nov. 2025, doi: 10.1002/qre.70025.
- [3] Vijayalakshmi S, N. Sebastian, and S. Ta, "Robust Bivariate Control Chart for Process Monitoring with Rational Subgroups," Jul. 2025, doi: 10.1002/qre.70022.
- [4] J. L. Alfaro and J. F. Ortega, "A comparison of robust alternatives to Hotelling's T^2 control chart," *J. Appl. Stat.*, vol. 36, no. 12, pp. 1385–1396, 2009, doi: 10.1080/02664760902810813.
- [5] D. A. Nugroho and Suwanda, "Diagram Kontrol T^2 Hotelling Minimum Volume Ellipsoid (Penerapan pada Pengontrolan Fluktuasi Saham)," *J. Ris. Stat.*, pp. 20–27, Jul. 2022, doi: 10.29313/jrs.vi.718.
- [6] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Stat.*, vol. 36, no. 1, pp. 199–227, 2008, doi: 10.1214/009053607000000758.
- [7] E. M. Issouani, P. Bertail, and E. Gautherat, "Exponential bounds for regularized Hotelling's T^2 statistic in high dimension," *J. Multivar. Anal.*, vol. 203, p. 105342, Sep. 2024, doi: 10.1016/j.jmva.2024.105342.
- [8] J. Kalina, "A Bayesian Framework for Regularized Estimation in Multivariate Models Integrating Approximate Computing Concepts," Sep. 12, 2025, *arXiv: arXiv:2509.10045*. doi: 10.48550/arXiv.2509.10045.
- [9] C. Agostinelli, A. Leung, V. J. Yohai, and R. H. Zamar, "Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination," *TEST*, vol. 24, no. 3, pp. 441–461, Sep. 2015, doi: 10.1007/s11749-015-0450-6.
- [10] R. Mohd Salleh, "A robust Estimation method of location and scale with application in monitoring process variability," Universiti Teknologi Malaysia, 2013.
- [11] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
- [12] H. A. Mohamed Shaffril, S. F. Samsuddin, and A. Abu Samah, "The ABC of systematic literature review: the basic methodological guidance for beginners," *Qual. Quant.*, vol. 55, no. 4, pp. 1319–1346, Aug. 2021, doi: 10.1007/s11135-020-01059-6.
- [13] R. Aggarwal *et al.*, "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis," *Npj Digit. Med.*, vol. 4, no. 1, p. 65, Apr. 2021, doi: 10.1038/s41746-021-00438-z.
- [14] B. S. Kelly *et al.*, "Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE)," *Eur. Radiol.*, vol. 32, no. 11, pp. 7998–8007, Apr. 2022, doi: 10.1007/s00330-022-08784-6.
- [15] H. Sitompul, M. Saifi, B. Hutahayan, and Sunarti, "Use of Renewable Energy to Enhance Firm Performance: A Systematic Review," *Sustainability*, vol. 16, no. 21, p. 9157, Oct. 2024, doi: 10.3390/su16219157.
- [16] S. Kraus, M. Breier, and S. Dasí-Rodríguez, "The art of crafting a systematic literature review in entrepreneurship research," *Int. Entrep. Manag. J.*, vol. 16, no. 3, pp. 1023–1042, Sep. 2020, doi: 10.1007/s11365-020-00635-4.
- [17] W. Mengist, T. Soromessa, and G. Legese, "Method for conducting systematic literature review and meta-analysis for environmental science research," *MethodsX*, vol. 7, p. 100777, 2020, doi: 10.1016/j.mex.2019.100777.
- [18] H. A. Mohamed Shaffril, S. F. Samsuddin, and A. Abu Samah, "The ABC of systematic literature review: the basic methodological guidance for beginners," *Qual. Quant.*, vol. 55, no. 4, pp. 1319–1346, Aug. 2021, doi: 10.1007/s11135-020-01059-6.
- [19] "PRISMA 2020 statement paper," PRISMA statement. Accessed: Jan. 10, 2026. [Online]. Available: <https://www.prisma-statement.org/prisma-2020-statement>

- [20] P. A. Alexander, "Methodological Guidance Paper: The Art and Science of Quality Systematic Reviews," *Rev. Educ. Res.*, vol. 90, no. 1, pp. 6–23, Feb. 2020, doi: 10.3102/0034654319854352.
- [21] Y. Xiao and M. Watson, "Guidance on Conducting a Systematic Literature Review," *J. Plan. Educ. Res.*, vol. 39, no. 1, pp. 93–112, 2017, doi: 10.1177/0739456X17723971.
- [22] V. Braun and V. Clarke, Eds., "Thematic analysis," in *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, Washington: American Psychological Association, 2012, pp. 57–71. doi: 10.1037/13620-004.
- [23] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, "Thematic Analysis: Striving to Meet the Trustworthiness Criteria," *Int. J. Qual. Methods*, vol. 16, no. 1, p. 1609406917733847, Dec. 2017, doi: 10.1177/1609406917733847.
- [24] G. V. Moura, A. A. P. Santos, and E. Ruiz, "Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection," *J. Bank. Finance*, vol. 118, 2020, doi: 10.1016/j.jbankfin.2020.105882.
- [25] C. Wu, G. Xu, X. Shen, and W. Pan, "A regularization-based adaptive test for high-dimensional generalized linear models," *J. Mach. Learn. Res.*, vol. 21, p. undefined-undefined, 2020.
- [26] O. Ledoit and M. Wolf, "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *Ann. Stat.*, vol. 48, no. 5, Oct. 2020, doi: 10.1214/19-AOS1921.
- [27] B. D. Robinson and V. Latimer, "Spectrally Robust Covariance Shrinkage for Hotelling's T^2 in High Dimensions," Jun. 11, 2025, arXiv: arXiv:2502.02006. doi: 10.48550/arXiv.2502.02006.
- [28] Y. Huang, C. Li, R. Li, and S. Yang, "An overview of tests on high-dimensional means," *J. Multivar. Anal.*, vol. 188, p. 104813, Mar. 2022, doi: 10.1016/j.jmva.2021.104813.
- [29] R. Modarres, "Hotelling T^2 test in high dimensions with application to Wilks outlier method," *Stat. Pap.*, vol. 65, no. 8, pp. 5203–5218, Oct. 2024, doi: 10.1007/s00362-024-01587-5.
- [30] P. Stoica, P. Babu, and P. Varshney, "Robust Estimation of the Covariance Matrix From Data With Outliers," *IEEE Open J. Signal Process.*, vol. 5, pp. 1061–1072, 2024, doi: 10.1109/OJSP.2024.3473610.
- [31] P. M. Shanthini and B. Sumithra, "Improving Reliability in Data Analysis: Techniques in Robust Statistics," vol. 89, no. 12, Dec. 2024.
- [32] H. Bulut, S. Iftikhar, N. Faiz, and O. Albalawi, "A Robust High-Dimensional Test for Two-Sample Comparisons," *Axioms*, vol. 13, no. 9, p. 585, Aug. 2024, doi: 10.3390/axioms13090585.
- [33] H. Li, A. Aue, D. Paul, J. Peng, and P. Wang, "An adaptable generalization of Hotelling's T^2 test in high dimension," *Ann. Stat.*, vol. 48, no. 3, Jun. 2020, doi: 10.1214/19-AOS1869.
- [34] M. C. Nascimento, L. H. C. Braz, and D. F. Ferreira, "Proposition of Bootstrap Tests for Comparisons Between Two Independent Mean Vectors in High Dimensionality," May 2025.
- [35] J. Zhou, Y. Chen, J. Yan, and S. Fang, "The Hotelling's T^2 Tests Revisited," *J. Syst. Sci. Complex.*, vol. 38, no. 6, pp. 2643–2663, Dec. 2025, doi: 10.1007/s11424-025-5041-3.

参考文献:

- [1] D. C. Montgomery, 《统计质量控制导论》. John Wiley & Sons, Inc., 2013.
- [2] M. Arslan, U. Shahzad, A. Yeganeh, H. Zhu, J. Malela-Majika, and S. Ahmad, "基于稳健L矩协方差矩阵的Hotelling's T^2 控制图: 用于在存在异常值情况下监测高维非正态多元数据," *Qual. Reliab. Eng. Int.*, vol. 41, no. 7, pp. 3308–3317, Nov. 2025, doi: 10.1002/qre.70025.
- [3] Vijayalakshmi S, N. Sebastian, and S. Ta, "用于具有合理子组的过程监测的稳健二元控制图," Jul. 2025, doi: 10.1002/qre.70022.
- [4] J. L. Alfaro and J. F. Ortega, "Hotelling's T^2 控制图稳健替代方法的比较," *J. Appl. Stat.*, vol. 36, no. 12, pp. 1385–1396, 2009, doi: 10.1080/02664760902810813.
- [5] D. A. Nugroho and Suwanda, "最小体积椭圆Hotelling's T^2 控制图: 在股票波动控制中的应用," *J. Ris. Stat.*, pp. 20–27, Jul. 2022, doi: 10.29313/jrs.vi.718.
- [6] P. J. Bickel and E. Levina, "大规模协方差矩阵的正则化估计," *Ann. Stat.*, vol. 36, no. 1, pp. 199–227, 2008, doi: 10.1214/009053607000000758.
- [7] E. M. Issouani, P. Bertail, and E. Gautherat, "高维情境下正则化Hotelling's T^2 统计量的指数界," *J. Multivar. Anal.*, vol. 203, p. 105342, Sep. 2024, doi: 10.1016/j.jmva.2024.105342.
- [8] J. Kalina, "在多元模型中结合近似计算概念的正则化估计贝叶斯框架," Sep. 12, 2025, arXiv: arXiv:2509.10045. doi: 10.48550/arXiv.2509.10045.
- [9] C. Agostinelli, A. Leung, V. J. Yohai, and R. H. Zamar, "在单级和案例级污染存在下的多元位置与散布稳健估计," *TEST*, vol. 24, no. 3, pp. 441–461, Sep. 2015, doi: 10.1007/s11749-015-0450-6.
- [10] R. Mohd Salleh, "位置与尺度的稳健估计方法及其在过程变异性监测中的应用," *Universiti Teknologi Malaysia*, 2013.
- [11] M. J. Page et al., "PRISMA 2020声明: 系统综述报告的更新指南," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
- [12] H. A. Mohamed Shaffril, S. F. Samsuddin, and A. Abu Samah, "系统文献综述基础指南: 面向初学者的基本方法论指导," *Qual. Quant.*, vol. 55, no. 4, pp. 1319–1346, Aug. 2021, doi: 10.1007/s11135-020-01059-6.
- [13] R. Aggarwal et al., "医学影像中深度学习的诊断准确性: 系统综述与荟萃分析," *Npj Digit. Med.*, vol. 4, no. 1, p. 65, Apr. 2021, doi: 10.1038/s41746-021-00438-z.
- [14] B. S. Kelly et al., "放射学人工智能: 系统综述与方法评估 (RAISE)," *Eur. Radiol.*, vol. 32, no. 11, pp. 7998–8007, Apr. 2022, doi: 10.1007/s00330-022-08784-6.
- [15] H. Sitompul, M. Saifi, B. Hutahayan, and Sunarti, "利用可再生能源提升企业绩效: 系统综述," *Sustainability*, vol. 16, no. 21, p. 9157, Oct. 2024, doi: 10.3390/su16219157.
- [16] S. Kraus, M. Breier, and S. Dasi-Rodríguez, "创业研究中撰写系统文献综述的艺术," *Int. Entrep. Manag. J.*, vol. 16, no. 3, pp. 1023–1042, Sep. 2020, doi: 10.1007/s11365-020-00635-4.
- [17] W. Mengist, T. Soromessa, and G. Legese, "环境科学研究中开展系统文献综述与荟萃分析的方法," *MethodsX*, vol. 7, p. 100777, 2020, doi:

10.1016/j.mex.2019.100777.

[18] H. A. Mohamed Shaffril, S. F. Samsuddin, and A. Abu Samah,

“系统文献综述基础指南：面向初学者的基本方法论指导,” *Qual. Quant.*, vol. 55, no. 4, pp. 1319–1346, Aug. 2021, doi: 10.1007/s11135-020-01059-6.

[19] “PRISMA 2020声明论文,” PRISMA statement. Accessed: Jan. 10, 2026. [Online]. Available: <https://www.prisma-statement.org/prisma-2020-statement>

[20] P. A. Alexander, “方法论指导论文：高质量系统综述的艺术与科学,” *Rev. Educ. Res.*, vol. 90, no. 1, pp. 6–23, Feb. 2020, doi: 10.3102/0034654319854352.

[21] Y. Xiao and M. Watson, “开展系统文献综述的指导,” *J. Plan. Educ. Res.*, vol. 39, no. 1, pp. 93–112, 2017, doi: 10.1177/0739456X17723971.

[22] V. Braun and V. Clarke, Eds., “主题分析,” in *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, Washington: American Psychological Association, 2012, pp. 57–71. doi: 10.1037/13620-004.

[23] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, “主题分析：努力满足可信度标准,” *Int. J. Qual. Methods*, vol. 16, no. 1, p. 1609406917733847, Dec. 2017, doi: 10.1177/1609406917733847.

[24] G. V. Moura, A. A. P. Santos, and E. Ruiz, “高维条件协方差矩阵的比较：对投资组合选择的启示,” *J. Bank. Finance*, vol. 118, 2020, doi: 10.1016/j.jbankfin.2020.105882.

[25] C. Wu, G. Xu, X. Shen, and W. Pan, “面向高维广义线性模型的基于正则化的自适应检验,” *J. Mach. Learn. Res.*, vol. 21, p. undefined-undefined, 2020.

[26] O. Ledoit and M. Wolf, “大维协方差矩阵的解析非线性收缩,” *Ann. Stat.*, vol. 48, no. 5, Oct. 2020, doi: 10.1214/19-AOS1921.

[27] B. D. Robinson and V. Latimer, “高维情境下用于Hotelling’s T^2 的谱稳健协方差收缩,” Jun. 11, 2025, arXiv: arXiv:2502.02006. doi: 10.48550/arXiv.2502.02006.

[28] Y. Huang, C. Li, R. Li, and S. Yang, “高维均值检验综述,” *J. Multivar. Anal.*, vol. 188, p. 104813, Mar. 2022, doi: 10.1016/j.jmva.2021.104813.

[29] R. Modarres, “高维Hotelling’s T^2 检验及其在Wilks异常值方法中的应用,” *Stat. Pap.*, vol. 65, no. 8, pp. 5203–5218, Oct. 2024, doi: 10.1007/s00362-024-01587-5.

[30] P. Stoica, P. Babu, and P. Varshney, “存在异常值数据中的协方差矩阵稳健估计,” *IEEE Open J. Signal Process.*, vol. 5, pp. 1061–1072, 2024, doi: 10.1109/OJSP.2024.3473610.

[31] P. M. Shanthini and B. Sumithra, “提高数据分析可靠性：稳健统计技术,” vol. 89, no. 12, Dec. 2024.

[32] H. Bulut, S. Iftikhar, N. Faiz, and O. Albalawi, “用于两样本比较的稳健高维检验,” *Axioms*, vol. 13, no. 9, p. 585, Aug. 2024, doi: 10.3390/axioms13090585.

[33] H. Li, A. Aue, D. Paul, J. Peng, and P. Wang, “高维情境下Hotelling’s T^2 检验的自适应推广,” *Ann. Stat.*, vol. 48, no. 3, Jun. 2020, doi: 10.1214/19-AOS1869.

[34] M. C. Nascimento, L. H. C. Braz, and D. F. Ferreira, “用于高维情境下两个独立均值向量比较的Bootstrap检验提出,” May 2025.

[35] J. Zhou, Y. Chen, J. Yan, and S. Fang, “重新审视Hotelling’s T^2 检验,” *J. Syst. Sci. Complex.*, vol. 38, no. 6, pp. 2643–2663, Dec. 2025, doi: 10.1007/s11424-025-5041-3.

Manuscript Information

Word count: 8,009 words (excluding references).

Peer-Review Record

Fast-track status: Not fast-tracked.

First-round reviews received: 3 reports.

Revision cycles completed: 3 rounds.

Final version submitted: May 15, 2026

Disclaimer / Publisher’s Note

The statements, opinions, and data contained in this article are solely those of the authors and do not necessarily represent the views of the *Journal of Hunan University (Natural Sciences)* or its editorial team. The journal and its editors disclaim any responsibility for injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content of this article.