UNIVERSITÀ DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN :
Economia e Management dell'Innovazione e della Sostenibilità
CICLO XXXIV
in CO-TUTELA con:
UNIVERSITÀ DEGLI STUDI DI FERRARA

Title:Nonparametric Hypothesis Testing for Multivariate and Complex Data on Sustainability

Coordinatore:
Prof. Stefano Azzali

Tutore:
Prof. Dr. Bonnini Stefano

Dottorando:
Assegie Getnet Melak

Anni Accademici : 2018/2019- 2020/2021

DECLARATION

I declare that the thesis entitled "Nonparametric Hypothesis Testing for Multivariate and Complex Data on Sustainability" has not been submitted in any form to the University and college previously for the award of any other degree.

DEDICATION

To my dearest family!

ACKNOWLEDGEMENTS

Writing the Ph.D. dissertation during the Covid-19 pandemic is not an easy task. Hence, first and foremost, I thank Almighty God for making me persistent in my Ph.D. journey.

My special gratitude goes to my tutor Bonnini Stefano (Prof.Dr.), especially for supporting me to develop the `R` script for the simulation. I truly, appreciate his kindness, friendship, and easy-going manner to help me. The structure of this thesis would not have been like this without his contributions, guidance, and assistance. Once again, I thank Bonnini Stefano (Prof.Dr.) for valuable and constructive comments, suggestions, and ideas.

I am grateful to the University of Parma, which sponsored this Ph.D. program. In addition, I thank Addis Ababa University, which hosted me during the Ph.D. mobility period. In line with this, I thank Mekonnen Tadesse (A.Prof.) and Merga Belina (Dr.) for their collaboration as reference persons. Moreover, I thank Ferrara University, Ethiopian Central Statistical Agency, Woldia University, and Woldia staff. Besides, I am very saddened by the unexpected pass away of my beloved collaborator Birhanu Teshome (Dr.).

I dedicate a special appreciation to my family members. In particular, my father and mother for their various advice, support, and inspiration in every aspect of my life.

I thank all my best friends who contribute in one way or another way to my thesis: Ermiyas Kefelegn, Worash Abocheregn, Lezashwork Fasik, Mengesha Asefa, Belete Dejen, Bayisa Tesfaye, Mariyamawit Fikresilassie, Riccardo Secomandi, and Michela Brunori.

CONTENTS

ABSTRACT

The dissertation comprises three chapters, in which the whole thesis focuses on the nonparametric solution for hypothesis testing of multivariate and complex datasets. The dataset's complexity includes the violation of parametric assumptions, small sample size, one-sided alternative hypothesis, and missing data. In the second chapter, we review about permutation test for analyzing complex datasets. We attempt to figure out the limitation of the previous studies and suggest some possible remedies. In chapter 3, we study the power performance and asymptotic properties of the combined permutation test (CPT) for complex data. The simulation results reveal that the CPT is the only nonparametric solution to tackle the loss of degrees of freedom when the number of response variables is greater than the sample size. For the two-sample test, the most powerful CPT is that based on the Tippett combination when the percentage of true partial alternative hypotheses is $\leq 30\%$, that based on the Fisher combination when the percentage is $> 30\%$ and $< 100\%$, and that based on the Liptak combination when the percentage is 100%. Finally, we analyzed the multidimensional sustainable development goals in Ethiopia using CPT. Moreover, we advance the power behavior of the CPT for multivariate analysis of variance, especially for the "big dataset". The simulation proves that the power of CPT increases as the number of samples and variables of the dataset increases. Besides, the proportion of true partial alternative hypotheses is more vital than the absolute number of variables in explaining the power improvement of CPT. Finally, we apply the CPT to study the organizational well-being of University workers. In chapter 4, we propose CPT for testing the significance of the multivariate linear regression model coefficients. The simulation results prove that the proposed CPT is exact, unbiased, and consistent to test the significance of coefficients. The power of CPT increases as the number of dependent

variables increases with fixed sample size. We applied the CPT to analyze multidimensional private firm performance in Ethiopia. Finally, chapter 5 consists of the summary of findings and future research work guidelines.

CHAPTER 1

INTRODUCTION

The dissertation comprises three chapters and focuses on a nonparametric solution for hypothesis testing of multivariate and complex datasets. The complexity of the dataset includes the violation of parametric assumptions, small sample size (especially when the number of variables is larger than the sample size), one-sided alternative hypothesis, and missing data. Hence, throughout the dissertation, we propose a family of nonparametric tests called combined permutation tests (CPT) for hypothesis testing. In addition, throughout the dissertation, we consider multivariate problems. In what follows, all the simulations and application of the proposed method to real datasets considered high dimensional responses by developing specific `R` scripts for each chapter.

A nonparametric test is vital when the traditional parametric approach is not suitable for the features of sample data and the study design. For example, the formal definition of big data may comprise volume, variety, and velocity. Roughly speaking, volume in this dissertation context covers the case in which the number of response variables is much larger than the sample size. For this reason, the volume of the big data may create a complex dataset for which the parametric approach has got problems due to the loss of degrees of freedom. For instance, in hypothesis testing of multivariate location parameters for two or more groups using the standard parametric test such as the Hotelling T^2 test, it is impossible when the sample size is smaller than the number of response variables.

The CPT is a vital statistical tool for testing coefficients' significance linear and non-linear models, especially for small sample sizes.

CPT is a suitable methodology when the problem can be broken down into many partial or sub-problems. In other words, when we have multivariate outcomes or multiple groups, we break down the complex null hypothesis into sub null hypotheses based on each aspect. Moreover, when we have a matrix of regression coefficients, we consider sub hypotheses based on the single coefficients. As a result, we need partial permutation test statistics for testing each sub null hypothesis, and finally, we combine them to make global information using the combined permutation test. For this reason, the p-values of each partial permutation test are combined using suitable functions. However, choosing the best combining function that determines the most powerful combined permutation test is challenging. Moreover, the power of the combined permutation test is highly dependent on the true partial alternative hypotheses. Hence, in this dissertation, the relationship between power and percentage of true partial alternative hypothesis is defined through simulation studies.

In the parametric methods, the functional form and the probability distribution of the data must be known to conduct estimation and hypothesis testing. However, they are optional in the nonparametric approach. For instance, before applying the parametric methods for estimation and hypothesis testing, linear, non-linear, polynomial, logit, log-log, exponential, and normality form of the sample data must be known. On the other hand, neither functional form of the data nor probability distribution of sample data is required to apply the combined permutation test.

Moreover, the complex nature of the economic dataset often requires permutation tests. For instance, datasets about sustainability (economic, social, and environmental dimensions) require distribution-free methods since the standard parametric approach is not suitable due to high dimensional outcomes, missing data, and correlated response variables (for example, sustainability indicators). Hence, we extend the application of the permutation test to problems concerning sustainable development goals and private firm performance in Ethiopia, useful to design suitable economic policy and public policy.

In the second chapter, we study an important statistical tool for analyzing complex datasets when parametric tests are not flexible, feasible, and powerful due to the violation of assumptions. In particular, we review the application of the permutation test for hypothesis testing in the comparative studies. Moreover, we review the application of the permutation test for the multivariate regression model. In addition, we attempt to figure out the limitations of the previous studies and suggest some possible remedies.

Chapter 3 studies power performance and asymptotic properties of the combined permutation for complex data. The standard parametric tests may not be suitable in specific conditions, such as in the presence of small sample sizes. We compare the power behavior of combined permutation tests for various scenarios through a simulation study. Moreover, we compare the power behavior of different combining functions as a function of the number of variables and the percentage of true partial alternative hypotheses to set the general rule. Finally, in the framework of sustainable development goals, we apply the permutation test to compare the living conditions of refugees and the hosting community households in Ethiopia.

Moreover, we study the power behavior of the permutation tests for multivariate analysis of variance. Hence, in this chapter, we compare the power behavior of different combination-based tests as a function of the proportion of true partial alternative hypotheses and the number of groups. However, for the multivariate multisample location problem, a comparative study of the power behavior of the most crucial CPT as the number of variables diverges is missing. Hence, we study the "big data" problem within the permutation MANOVA framework.

In chapter 4, we study the combined permutation tests for testing the significance of coefficients of multivariate regression models. However, the standard multivariate parametric tests may not suit hypothesis testing on regression coefficients in various conditions, such as small sample sizes and non-normal errors. Hence, we simulated datasets to compare the power behavior of combined permutation tests with the typical parametric multivariate test (Pillai's Trace test) for various scenarios. In addition, we investigate the power behavior under different correlation matrices of the multivariate response and mild multicollinearity. Moreover, we study the power behavior of Fisher and Tippett combinations as a function of the number of non-zero coefficients or the percentage of true alternative hypotheses. Finally, we apply the combined permutation tests to analyze multidimensional private firm performance in Ethiopia.

The whole dissertation is structured as follows. Following Chapter 1 general introduction, Chapter 2 is a theoretical chapter, which covers the general literature review about the permutation tests. In Chapter 3, we study the power and asymptotic properties of combined permutation tests for the two-sample problem with high dimensional responses, and it deals with combined permutation tests for multivariate multisample problems. Moreover, Chapter 4 examines the robustness of combined permutation tests on the significance of the multivariate regression model coefficients under various scenarios. Finally, Chapter 5 is dedicated to the conclusions of the thesis and possible future research areas.

CHAPTER 2

REVIEW ABOUT THE PERMUTATION APPROACH IN HYPOTHESIS TESTING

2.1 Introduction to Permutation Tests

In many fields, the application of permutation tests for real data analysis has increased. For instance, many papers use permutation tests in ecology, neurosciences, biostatistics and econometrics[1, 2, 3, 4, 5, 6]. However, [7] initially introduced the permutation test in a problem for paired agricultural data collected by Charles Darwin. Since then, the permutation approach has become a valid solution for testing problems when the parametric methods are not suitable, feasible, and powerful due to the violations of assumptions[8, 9, 10, 11, 12].

Due to the fact that they are flexible, powerful and robust, permutation tests are becoming very popular in several empirical disciplines for any type of complex data structure [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37].

Furthermore, [38] provided the formal and concise definition of permutation test :

> "Permutation tests for a hypothesis exist whenever the joint distribution of the observations under the hypothesis has a certain kind of symmetry, namely when a set of permutations of the observations leave the distribution the same (the distributions are invariant under a group of permutations)".

In many cases, the permutation test is mainly based on the equality of distributions under the null hypothesis.

Although the permutation test is a type of re-sampling method like bootstrapping, unlike bootstrapping, the sampling process in the permutation test takes place by conditioning to the observed dataset[39, 40, 5]. In other words, given the sample dataset, we could re-sample observations many times without replacement to estimate the empirical null distribution of the permutation test statistic. In multivariate permutation tests, the dependency between response variables is taken into account without modeling the dependency structure of the joint distribution. Extensive simulation results in previous studies revealed that the permutation tests are much powerful than the bootstrap method [41, 42, 43, 40, 44]. Moreover, the bootstrap test is much more computationally intensive than the permutation test.

The permutation test has essential advantages and a few disadvantages. In other words, thanks to the improvement of computers power and speed, computational limitations of permutation tests could be compensated by their huge advantages [45].

There are many R packages and source codes developed in the R programming language dedicated to permutation tests. Traditional software like SPSS and STATA have also an option for the permutation test. Some examples of useful R packages for permutation tests include `lmperm` package [46], `coin` package [47], `flip` package [48], and others. Similarly, some useful R packages and source functions for the application of permutation tests are described in [39] and [4].

2.1.1 Mild Assumptions of Permutation Tests

Even though the permutation tests are nonparametric methods, they are based on some mild assumptions commonly met in many real-world situations[49, 50, 51]. For instance, [52] focuses on the mild condition of exchangeability. Moreover, other researchers studied conditions and properties of permutation tests such as the appropriate formulations of the null hypothesis, unbiasedness and consistency [53, 54, 39]. On the other hand, [55] argued that permutation tests could also be applied for non-exchangeable error terms.

Exchangeability is a simple and weak condition. For instance, in experimental designs, the random allocation of subjects to treatments is sufficient to justify the exchangeability. Moreover, the independence of observations may be sufficient in an observational study.

Furthermore, depending on the goal of analysis, different types of exchangeablility conditions are introduced[52]: preserving transforms, asymptotic exchangeability, partial exchangeability, and weak exchangeability. For example, if the interest of the study is testing the significance of coefficients in the presence of nuisance parameters in the linear model, exchangeability-preserving transform is considered[52]. Exchangeability and conditioning on the sample dataset imply an important invariance property[54, 52]. In other words, under exchangeability, the joint distribution of the observations is invariant under the resampling of subjects. As a result, sufficient permutation test statistics can be computed.

2.1.2 Computation of P-Values in Permutation Tests

In the permutation test, the hypothesis testing procedure takes place by calculating the p-values of the permutation test statistic[56]. She studied the multivariate permutation analysis of variance for multifactor and complex designs by calculating the asymptotic permutation p-values. Her method is based on computing the distance matrix from the permuted datasets and ranking the distances to construct the permutation test. However, the robustness depends on the choice of the method for computing distances, and it also requires balanced replicates for each cell. Moreover, the author noted that computing fully nonparametric p-values for higher-order interaction is challenging. Hence, a sort of restricted permutations or the so-called synchronized permutation could provide an approximate result [57, 58].

In hypothesis testing, the permutation test is the method of computing p-value empirically conditional on sample data to make the appropriate decision about the null hypothesis. For a preassigned level of significance α, we could compare the p-value obtained from the permutation test with α to accept or reject the null hypothesis in favor of the alternative hypothesis.

In the literature, the conditional Monte Carlo method is useful for the computation of p-values according to the permutation principle[59, 60, 61, 57, 53, 39, 62, 63]. Often, a number of conditional Monte Carlo permutations of 1000 should be sufficient to have a good approximation of the null permutation distribution and obtain p-values very close to the exact ones [64, 65]. Some authors use 10000 permutations[39, 66]. To get exact p-values all, the possible permutations n! must be considered. However, when the sample size is large, determining all the possible permutations is computationally challenging, and a conditional Monte Carlo approach is considered for computing the permutation distribution of the test statistic and for estimating the p-values[67, 68].

Furthermore, [67] used a simulation study to determine the appropriate number of permutations to get well-approximated p-values. He found that the number of permutations(R) required for estimating the power of the permutation test and estimating p-values for practical data analysis is different depending on the level of significance. For instance, 1000 and 5000 permutations are sufficient for estimating power performance and p-values respectively for practical data analysis at 5% nominal level. Moreover, due to the invariance property, the permutation distribution of the test statistic obtained from all possible permutations (n!), and with conditional Monte Carlo method asymptotically provide the same conclusion.

This chapter considers a review of permutation tests for comparative studies and linear regression models. The chapter is structured as follows. Section 2 is about a permutation test for comparative studies. The permutation solutions for regression are presented in section 3. Section 4, presents the permutation solution for the comparisons of dependent sample. Section 5 covers permutation solution for missing data problems. Finally, section 6 comprises the conclusions.

2.2 Permutation Test for Comparative Studies

When researchers conduct the hypothesis testing to compare groups based on a shift in location or scale parameter for univariate and multivariate continuous outcomes, the standard parametric approach might not be a suitable choice for various reasons such as skewed data, excess kurtosis, non-normal data, unequal variances, small sample sizes, missing data, and others. In addition, comparing the difference in proportion between groups for categorical variables using the standard parametric test may not be appropriate. Hence, the nonparametric solution using permutation tests is an inevitable choice for such problems.

2.2.1 Permutation Approach for Two-Sample Problems

When the assumptions of normality and equal variance are violated in two-sample location problems, the permutation test is a suitable solution[69, 45, 70, 71, 72, 73, 74, 75, 76, 77]. The reason is that the standard Student's t-test for univariate response and its multivariate extension Hotelling T-square test cannot provide reliable results when these assumptions are not satisfied.

Furthermore, when the type of alternative hypothesis is directional, there is no parametric solution to test the shift in location between two samples for multivariate response variables[78, 4]. In other words, when there is a positive and negative treatment effect or shift in location parameter under the alternative hypothesis, the standard parametric two-sample test based on t-test for univariate response and Hotelling T-square test for the multivariate test have no extension for the multi-sided alternative hypothesis. Hence, the permutation solution provides a valuable contribution to such problems.

The permutation test for two-sample problems when the treatment effect is some known constant is well investigated in the literature, see for example[79, 4, 80, 81, 82]. However, this permutation solution needs to be applied when the treatment effect is not constant or the so-called heterogeneous treatment effect. In such a case, there is an interesting recent development of permutation test for testing the heterogeneous treatment effect when there is an unknown nuisance parameter[59]. This approach used a martingale transformed test statistic for two samples to control the limiting rejections probabilities based on the pivotal statistic. The simulation results revealed that the method of[59] controls the type I error rate.

Similarly, many authors recommended the use of pivotal statistics in permutation tests to get good approximation[1, 83, 66]. However, since the permutation test is in the family of distribution-free tests, pivotal quantity is neither necessary nor sufficient conditions to apply permutation tests. However, the permutation test also provides well-approximated results if the researcher uses the pivotal test statistic.

The traditional parametric approach can not apply when we have mixed multivariate variables. On the other hand, as[72] investigated in his research work, the permutation principle provides a solution. For instance, in the case of two groups, he used the sample total for a numeric variable and the Fisher exact type permutation test statistic for the categorical variable to test a significant shift in the location parameter. Then, to synthesize the information of the two test statistics, he considered the combination of p-values of the test statistics using the nonparametric combination method[39]. In particular, he used the Tippett combining function since he suspected that the first aspect corresponding to sample total has a large p-value under the alternative hypothesis for heavy-tailed distributions. Finally, he classified the numerical variable as dichotomous based on the median of the variables. Technically, he considered the robustness of the permutation test in the presence of a mixed response.

He simulated data from continuous and discrete distributions. He also studied the robustness of the permutation test for the two-sample location problem under different heavy-tailed distributions like the Cauchy, the half-Cauchy, and some mixtures of normal distributions with unequal variances using Monte Carlo simulations compared with student t-test. Consequently, his proposal of the bi-aspect permutation test provided a well-approximated type I error rate robust and powerful results.

2.2.2 Permutation Solution for Two-sample Location-scale Problems

The traditional parametric approach, does not provide solutions for testing location and scale parameter jointly [4]. In other words, if the researcher is interested in examining the treatment effects in the first moment and the second moment, that is, so-called change in both mean and variance, the traditional parametric test could not jointly account for such a problem.

In contrast, the permutation approach solves such a complex problem. For instance, [39] used the multi-aspect permutation principle to test a shift in location parameter and changes in variability between two samples for univariate as well as multivariate outcomes. The permutation solution for the location-scale problem is somehow analogous to that for the mixture variables problem. [39] tested the significance of location and scale parameters using the difference of means and differences of the sum of squares as a statistic for the test on location and scale parameter respectively. In addition, a multi-aspect permutation test for the first and second moment in practical landmark shape data analysis is found in[84].

Moreover, [45] proposed an interesting study for a completely randomized experimental design to test a shift in location parameter and change in scale parameter for two samples. The method of[45] to jointly test the significance of location and scale parameter is somehow different from[39]. This is because[45] considered an omnibus statistic. Whereas in[39] different test statistics for the different aspects were considered. The three omnibus permutation tests were Euclidean commensuration, Hotelling commensuration, and the permutation version of the Bartlett-Nanda-Pillai test. A simulation study under various distributions was carried out to compare them with the parametric Bartlett-Nanda-Pillai trace test. The simulation results revealed that the Euclidean commensuration permutation test was the most performant irrespective of the considered distributions for a bivariate response. The power of the tests was not performant for the other distance. In other words, the power performance of those methods depends on the choice of distance metrics. In addition, the Hotelling commensuration permutation test was not stable due to the loss of degrees of freedom.

2.2.3 Permutation Approach for Small Sample Multivariate Problems

Although the literature about solutions for multivariate tests with the number of variables greater than the number of observations is not rich, some authors addressed the problem[74, 39, 71]. In two-sample or multi-sample tests, when sample sizes are smaller than the number of response variables, the classic parametric tests such as the Hotelling T^2- test cannot be used due to the loss of degrees of freedom. In this case permutation solutions are possible [39]. For these tests the so-called finite-sample consistency is satisfied. This property implies that, under some conditions, the power of the test in H_1 increases with the number of variables irrespective of the finite sample sizes.

Furthermore, [74] proposed an invariant inter-point distance method using the permutation principle to compare the location parameters for multivariate two-sample problems in the presence of high dimensional response variables with fixed small sample sizes. They simulated data to compare the power behavior of the permutation test with nonparametric tests (such as multivariate generalization of the run test[85], nearest neighbor [86], and Rosenbaum test statistics [87] under normal and Laplace distributions. They found that the permutation test is powerful in testing the significance of location and scale parameters in the presence of small sample sizes with divergent response variables under normality and Laplace distributions. Finally, they proved that the permutation test is more powerful than other nonparametric tests.

Moreover, the power of multivariate generalization of the run test, nearest neighbor, and Rosenbaum test statistic dropped to zero as the dimension of the response variables increased indefinitely. In contrast, the power of the permutation test based on the inter-point distance tends to one when the number of components of the response variable increases for a given fixed sample size. The simulation also revealed the cut-off(minimum) sample sizes (5 or 4) for each group under the divergent number of response variables in which the power behavior of the permutation test could approximate one. The simulation proved that the permutation test is consistent and powerful when the number of responses diverges with small fixed sample sizes even under the Behrens-Fisher problem. However, their method assumed the uniformly bounded forth moments and week correlation among the components of the response variables.

Furthermore, [88] proposed a recent development to improve the power behavior of a multivariate version of the two-sample permutation test, the so-called combination-based permutation test. The main property of the combination-based permutation tests is that they condense the statistical information of all response variables into one statistic to decide on the null hypothesis, and they implicitly take into account the dependency between response variables.

Another contribution to the multivariate tests with a large number of response variables is [50]. Through Monte Carlo simulations, they found that the power behavior of Hotelling T-square and rank-based test decrease as the number of variables increase with fixed sample size, while the power of combination-based permutation tests increased monotonically as the number of responses variables increased with fixed sample size. The power of the combination-based permutation tests is good even for small shifts in mean differences.

Nevertheless, when the sample size is less than the number of response variables, the Hotelling T-square test, and the nonparametric rank-based test cannot be applied. The proposed multi-aspect strategy improved the power of the permutation test for skewed distribution. However, the iterated combining techniques do not improve the power of the test. Thus, other techniques are necessary to support the decision about the appropriate combining functions to maximize the power of the combined permutation test.

2.2.4 Permutation Test for the Two-sample Behrens-Fisher Problem

A comparative study on location is called the Behrens-Fisher problem when we have unequal unknown variances of the populations[89]. [70] proposed an asymptotic test based on the permutation version of Welch test, which is an extension of Pitman's permutation test for the Behrens-Fisher problem on the equality of two population means. Although the permutation approach provides an exact test under the IID assumption in the null hypothesis, the Janssen method based on the studentized statistic does not require the IID assumption to obtain an asymptotic test. [70] applied this method for a composite null hypothesis against one-sided alternatives. The simulation results revealed that the permutation test is more powerful than the Welch test[90] for skewed distributions. The limitation of this method is that it is not stable for an unbalanced design.

According to [39] studentized and non-studentized permutation test statistics are permutationally equivalent due to the conditioning on the sample data and invariance property, and they provide approximately similar results. Thus, eliminating the denominator of the studentized permutation test statistic may improve the computational aspect of this test statistic.

Furthermore, many authors solved the Behrens-Fisher problem nonparametrically using the permutation approach. For more example see[91, 89, 40, 92, 93]. In addition, more recent solutions to the multivariate Behrens-Fisher problem for the dissimilarity-based measure is found in[94], and an approximate randomization test for the high-dimensional two-sample Behrens-Fisher problem under arbitrary covariances are found in[41].

2.2.5 Permutation Principle for Multi-Sample Problems

In addition to a vast literature on two-sample permutation tests, there are many scientific contributions (theoretical and applied) on the univariate and multivariate multisample extensions, such as the permutation analysis of variance (ANOVA) and the permutation multivariate analysis of variance (MANOVA) [42, 95, 96, 43, 80, 82, 97, 83, 98, 99].

The standard parametric F test for ANOVA requires equal variances, normality, continuous response, and IID conditions. Similar assumptions are required for multivariate multi-sample or the so-called MANOVA. In addition, even if the assumptions of (M)ANOVA are satisfied, there are some conditions in which the classical (M)ANOVA cannot be applied[100, 101, 102]. Some examples are small sample sizes, correlated responses, and others. Hence, in such a problem, the permutation solution for (M)ANOVA is an unavoidable method of hypothesis testing for multiple samples.

Similar to what happens in the two-sample multivariate location problem, the loss of degrees of freedom due to a large number of response variables with fixed sample sizes is quite common in MANOVA. In this case, the MANOVA tests, as well as the nonparametric rank-based tests, could not be applied in the presence of small sample sizes with the number of responses diverges [88]. In such a situation, the nonparametric solution based on the permutation approach is an inevitable choice.

Permutation (M)ANOVA provides accurate and reliable (powerful) results [80]. In other words, hypothesis testing of significant mean difference among groups based on the permutation test under the null hypothesis yields a type I error rate that respects the predetermined level of significance α and rejects the null hypothesis when the alternative is true with high probability. According to[39] and [5] permutation tests are exact, unbiased, and consistent.

[80] compares the power behavior of permutation test based on different permutation test statistics for continuous and count data in case of unbalanced sample sizes, heteroscedasticity, correlation, and a large number of responses. The simulation results reveal that the power of the permutation test decreases in the presence of positive correlations among the response variables. It is highly unreliable for the negative correlation scenario. However, since the permutation test is a conditional inference that relies on the sample data, the response variables are supposed to be positively related. For this reason, in our opinion, simulating from a negative correlation structure does not make sense.

2.2.6 Permuting the Observations in Multi-sample Problems

In general, to get reliable and accurate results based on the permutation test, when H_o is true, the permutations of the statistical units for one way (M)ANOVA can be applied by exchanging the row vectors of the observations between treatment groups or factor levels irrespective of the outcome variables[95, 43, 80, 83]. However, the permutation principle for the two-way (M)ANOVA or general factorial design requires some restrictions. For this reason, researchers used synchronized permutation to permute row vectors of the observations between levels of one factor by keeping the other factors as a block[57, 4, 103, 58, 104, 83].

In addition, the synchronization process could be constrained or unconstrained depending on the sample size in each level [105, 103, 106]. In the case of constrained synchronization, if we have an unbalanced design, the smallest sample size determines the number of units to be exchanged within treatment levels, and the position of the units must be fixed. While in unconstrained synchronized permutations the position of the units doesn't matter to permute the observations (see[96] for a detailed explanation).

[96] developed a synchronized permutation test for unbalanced two-factor ANOVA with two levels. Since the error terms are not exchangeable between the two factors under the null hypothesis, they used the synchronized permutation principle. However, the error terms are exchangeable within the levels of one factor by considering the other factor as a block to get an approximate exact p-value. They used the weighted sum of observations within each cell as a permutation test statistic to test the significance of factors' main effect and interaction effects.

The conditional Monte Carlo simulations revealed that the power performance of the permutation test is somehow influenced by the permutation mechanism (constrained or unconstrained) and the type of weights. For instance, the constrained synchronized permutation mechanisms reduced the power of the permutation test. In addition, the test was conservative for restricted weighted and constrained synchronized permutation mechanisms. However, it seems that the power of the test using the constrained synchronized permutations is unstable for small sample sizes in the case of an unbalanced design.

2.2.7 The Permutation Test Statistic for (M)ANOVA

Due to the random nature of the permutation principle, obtaining the optimum permutation test statistic is among the challenging problems [88]. In permutation MANOVA, different test statistics have been considered, for example, distance-based and (M)ANOVA statistics. Thus, comparing the type, I error rate, and the power of the test based on those test statistics are crucial for selecting appropriate test statistics and improving the power performance of the permutation test. In line with this, there is an interesting paper about the power comparison of distance-based statistics, and MANOVA statistics in a multi-sample comparative study [107]. They classified the statistics as distance-based when they used the average distance of pairs of units or the dissimilarity measures, and variable-based statistics when they used summary statistics such as Wilks lambda, likelihood ratio type, and ANOVA F type statistic.

Nevertheless, the distance-based statistic requires the transformation or standardization of the original data. Besides, variable-based statistic requires normality assumption. Therefore, the permutation test statistics are an inevitable choice to make inferences based on mild conditions. In the permutation principle, the variable-based test statistics are often preferable to the distance-based ones. First, a different summary statistic for each aspect or variable is used, and then the information provided by partial tests is combined. As a result, the power of the combined permutation test is greater than the power of the other parametric and nonparametric competitors.

As defined by [107], the statistical power of the permutation test is: "... *the proportion of times it was significant at the α level, for N sets of data generated from the same distribution. For each of these data sets, statistical significance was assessed using R permutations* ...". As far as the distance-based statistic is concerned, the standardization and transformation of the original dataset might impact the power of the test. Thus, this issue and the robustness of variable-based statistics over distance-based statistics are studied[107]. The necessary procedures to test the multivariate data are transformation, standardization, choice of the distance metrics, and test statistic. Regarding the distance metrics, the most commonly used are Euclidean, Bray-Curtis, and Manhattan. However, the distance measures cannot account for the possible correlation of the multivariate outcomes.

They designed a simulation study to compare the power behavior under the null and the alternative hypothesis. The simulation results revealed that the p-values are more or less affected by the transformation of the original data. More precisely, this effect is pronounced for Bray-Curtis distance measures for unbalanced design. The standardization and transformation of data have a significant influence on the power of the test.

Furthermore, comparative studies based on the permutation tests for ANOVA were proposed using different test statistics such as F type statistic ATS[108], Wald type statistic WTS [43, 103], Fisher-Pitman [97], studentized test statistic[42, 43, 109, 89, 110], and distance-based statistic [80, 111]. Although Wald type statistic is useful for comparing the means of multivariate outcomes among groups, it requires the computation of the sample covariance matrix and difficulty of convergence for large sample size [103]. However, [42] introduced the modified form of the Wald type statistic that used the identity weight.

Regarding the impacts of distance metrics, the Euclidean distance-based test statistic seems more performant than other distance-based statistics [80, 111]. Some distance-based permutation statistics have no extension for a categorical response. However, [98] extended the application of distance-based permutation statistics for categorical and continuous outcomes for (M)ANOVA setup.

In the literature, the permutation solution for MANOVA used Eigenvalues by partitioning the total variability to each factors[80, 111, 83]. However, the Eigenvalues obtained from the distance matrix sometimes are negative. The negative Eigenvalues have got difficulty in interpreting the results, and some authors corrected it by adding a constant to the sum of squares [112]. Whereas[111] proposed permutation version of Euclidean-based statistic in hypothesis testing of MANOVA or MANCOVA by partitioning the total variability or the dissimilarity because the pseudo F- statistic could be derived from the Euclidean distance matrix to test the significance of the factors. They carried out a simulation study to test the accuracy of the proposed method called distance-based redundancy analysis (dB-RDA), which does not require any correction to have positive Eigenvalues with three competitors such as Bray-Curtis distances (Dir), the [113] method of distance-based redundancy analysis (LA) and axes (Pos) analysis for testing fixed, random, and mixed effects. The simulation results provided the preassigned nominal level for the fixed effect model. In contrast, the test was conservative for random and mixed linear models, especially for small sample sizes and lognormal distribution. The proposed distance-based redundancy analysis (dB-RDA) provided a well-approximated type I error rate. However, the simulation can be extended to check the good power performance under the alternative hypothesis.

In one-way ANOVA design, when the assumptions of the F test are violated, a permutation version of the Fisher-Pitman test could be used for testing the equality of location parameters among the groups [97]. The advantage of the Fisher-Pitman test is the way the permutation test statistic is computed. Since the total observations $N = \sum_{g=1}^{S} n_g$, number of groups S, and the total sample mean $\bar{Y}$ are permutationally constant due to the invariance property under exchangeability, the ANOVA F test could be simplified to the Fisher-Pitman permutation test statistic or the weighted sum of squares of the within-group sample means $\bar{Y}_g$,

$$\mathbf{T} = \sum_{g=1}^{S} n_g(\bar{Y}_g)^2 \tag{2.1}$$

Thus, it is simple to compute the Fisher-Pitman permutation test statistic, and somehow the computationally intensive nature of the permutation test is also controlled by taking the simplified form of the permutation test statistic. Moreover, for balanced sample size in each group, the Fisher-Pitman permutation test statistic further simplified to

$$\mathbf{T} = \sum_{g=1}^{S} (\bar{Y}_g)^2 \tag{2.2}$$

since the sample sizes for balanced design are equal. Besides, Fisher- Pitman permutation test is a perfect solution for Behrens-Fisher statistical problem.

2.2.8 Permutation Test for Multi-sample Behrens-Fisher Problem

[42] proposed a solution for the multivariate Behrens-Fisher problem of two-way MANOVA. He used the ANOVA type statistic ATS to test the equality of population mean vectors under heteroscedastic error variances. The asymptotic null distribution of ATS is a mixture of central chi-square distributions. This method is robust against skewed distributions, unequal variances, and nested factorial designs. They proved that the permutation version of ATS is consistent. However, although the ATS works well for finite sample sizes and small groups, they noted that the ATS is affected by the number of groups, suggesting the standardized form. In addition, they derived the confidence region for the interval estimation of mean vectors. Besides, it was more powerful than its competitor WTS and the bootstrap method regardless of the distributions' variance structure, sample sizes, and skewness.

Similarly, [83] developed a family of permutation tests for hypothesis testing of the equality of treatment effects on multivariate outcomes when the standard parametric MANOVA tests are not applicable due to the stringent assumptions. The proposed solution used squared Euclidean distance as a permutation test statistic. This method is crucial for testing the significance of both factors' main effects and interaction effects. Moreover, they developed permutation solutions for general factorial designs and complex models like nested or hierarchical models based on a distance measure. One advantage of this test is that it considers the possible dependency among the components of multivariate outcomes in multifactorial designs, which is problematic to cope with using traditional parametric MANOVA.

Nonetheless, attention must be considered when testing the interaction effects since error terms are not always exchangeable under the null hypothesis to get exact p-values. In addition, the distance-based statistic requires the transformation of the data. In this case, the interpretation of the results may change with the data transformation. Moreover, if the number of groups is more than two, a posterior hypothesis testing can be applied to identify which groups are significantly contributing to significant mean differences using the permutation version of the t-test .

[43] extended the application of permutation tests to general factorial designs with two or more factors, nested and hierarchical design in the presence of heteroscedasticity. The Wald type parametric test cannot provide a valid result for the hetroscedastic case and small sample sizes. The distribution of the Wald type test hardly approximates the chi-square distribution with small sample sizes.

Furthermore, when the number of factors and levels increase indefinitely in factorial designs, the Wald type parametric test could not efficiently cope with the hypothesis testing of treatments effects. In contrast, the proposed Wald type permutation test statistic works well under small sample sizes, heteroscedastic errors, and many factors. In addition, they considered the standardized Wald type permuted statistic using the estimated variance obtained from permuted data. However, the computation of the error variance may slow down the calculation of the permutation test, and it needs simplification. Finally, they proved that the distribution of Wald type permutation test statistic conditional on the sample data weakly converged in probability to the central chi-square distribution.

Moreover, their simulation study revealed that the chi-square test, the parametric ANOVA, and the Wald type parametric test provided a liberal type I error, especially under small sample sizes and skewed error distributions. In contrast, the Wald type permutation test respects the nominal α level irrespective of the sample size, covariance structure, and other factors except for a little inaccurate nature for skewed log-normal distribution for unequal variances.

2.2.9 Permutation Test for Categorical Data

Hypothesis testing problems for categorical data, especially in the multivariate case, are not easy to solve. With multidimensional variables, a crucial and complex aspect concerns the dependence structure of the response components. For this reason, following a distribution-free approach for such testing problems in the presence of the multivariate dependent variables is crucial.

Permutation solutions for multivariate stochastic ordering are investigated by [114, 115, 116, 81, 117, 118]. Statistical ordering is typical of various complex problems such as tests for restricted (directional) alternatives [119], multiple comparisons [120], ranking populations[114].

For instance, [118] applied multivariate permutation tests for comparing two populations in the presence of a multivariate categorical response. She focused on the so-called case-control study and used the odds ratio as the test statistic. If we consider the contingency table where rows correspond to treatments and columns to response categories (or vice versa), permutations can be applied by tables with different joint absolute frequencies but the same marginal absolute frequencies. Investigating the treatment effect is equivalent to testing the association between responses and treatments. Unlike the classical chi-square test, there are no restrictions on the minimum frequency value in the single cells in the permutation test. [118] proved that the permutation test is robust even for small sample sizes and, in some cases, minimal frequencies.

In practice, parametric tests such as the likelihood ratio test could not be appropriate for testing the stochastic dominance for ordered categorical variables. On the other hand, a fascinating application of permutation tests for univariate and multivariate ordered categorical variables under restricted alternatives was proposed by [116]. The proposed method is based on transformingg the categorical response variables into numeric variables. The global null hypothesis is based on the treatment and control group equality. First, H_o is broken down into k-1 sub null hypotheses, where k is the categories of the ordinal variable. Then, the partial permutation test statistics were used to test each sub-null hypothesis based on the sample moments. Next, they controlled the possible incorrect rejection of the null hypotheses or the so-called family-wise error rate. Finally, the simulated data from an ordinal multinomial distribution with four categories using the GenOrd R package to investigate the power behavior of permutation test (Fisher and Liptak combining function) compared to the competitor rank based Wilcoxon test and the Brunnel–Munzel test after ranking the units or categories. The empirical power comparison provided similar results, especially under the null hypothesis for different nominal levels.

Whereas under the alternative, the permutation test based on the sample moments is the most powerful by applying score transformation. They noted that the power of their permutation test is affected by the weights or score transformation, for example, symmetric or asymmetric scores. For this reason, Liptak combining function is relatively more powerful than the Fisher combining function.

Furthermore, when the nature of the response variables is ordinal, for example, Likert-scale questionnaires, tests for group comparisons can be carried out by replacing the ordered categories with numeric scores[79]. They applied the Wilks lambda type permutation statistic based on the rank transformations of units, and this test statistic works for skewed data such as ordinal dependent multivariate outcomes. They carried out a Monte Carlo simulation study to investigate power behavior and type I error rate of the permutation test for multivariate analysis of ordinal variables by generating data from the multivariate ordinal distributions. The competitor tests were the traditional MONAOVA test based on Wilks lambda, the nonparametric Wilks, structural equation models for MANOVA tests, the test based on spatial signs with inner centering and outer centering, and others. Except for the structural equation models for MANOVA tests, all the tests respect the nominal α level under H_o. The structural equations' system seems powerful, but this is due to the inflated type I error rate because it is anti-conservative under H_o.

One main interest in hypothesis testing of categorical data is to rank several groups based on some characteristics. For instance, in a biomedical study, several groups of patients can be ordered based on dose levels. Hence, such ranking of groups can be broken down into several the stochastic dominance or directional of pairwise comparisons. [114] proposed a metric-free permutation test for comparing and ranking multivariate populations. When the null hypothesis of equality of the multivariate populations is rejected, the need arises to determine which groups show significant differences to determine a final ranking. The post hoc comparison is the solution for such kind of problem. Moreover, their permutation solution for stochastic dominance and ranking can apply to mixed response variables.

Using Anderson- Darling type permutation statistic, they simulated multivariate ordinal categorical data to investigate the ranking of populations for the different number of groups, small sample sizes, and different skewed and normal distributions. The simulation results revealed that the estimate of correct global and correct individual ranking is close to the actual rank under the homogeneity assumptions irrespective of the sample size and choice of error distributions. In addition, the bias was fewer under the null hypothesis.

2.3 Permutation Approach for Tests on the Parameters of the Regression Model

The classical test of significance of the model coefficients in the parametric approach requires the estimation of model coefficients and other parameters to construct the test statistic. However, estimating the model coefficients might be problematic in high dimensionality, few degrees of freedom, and violation of assumptions (such as non-normal data, not IID errors, heteroscedastic variance). For instance, when the assumed distribution of errors is implausible, estimate the model coefficients using popular techniques such as the maximum likelihood estimator.

Moreover, the ordinary least square estimation method cannot be applied when the assumptions of linearity and IID are violated. For this reason, the classical significance test of model coefficients based on the parametric approach might not be suitable. Whereas, some nonparametric solutions, in particular the permutation approach, do not require the estimation of nuisance parameters such as the covariance structure to derive the permutation test statistic. Moreover, the permutation test doesn't require specifying the underlying distribution to test the significance of the model parameters in linear or nonlinear models.

In the literature of nonparametric regression analysis, the vast majority of research was dedicated to investigating the nonparametric estimation of the coefficients of linear regression and panel regression/ longitudinal models when the underlying distribution of the error terms are unknown, See for more details in the literature[121, 122, 123, 124].

One of the popular methods of nonparametric estimation is the smoother kernel, which does not require the functional form of the regression model to be known [124]. However, the nonparametric regression mainly relies on the bandwidth; the determination of the bandwidth is computationally intensive. There is a variety of bandwidth calculation methods such as the rule of thumb, plug-in methods, cross-validation, and bootstrap method [124]. There are many R packages dedicated to computing the bandwidth in order to reduce the computationally intensive nature of bandwidth computation such as np package [124], kernel smoother package [125] and npregfast package [126].

Furthermore, many authors only considered single independent variables while [124] extended to continuous multiple independent variables and the mixed type of multiple independent variables. However, extending the nonparametric regression to multivariate models might be difficult because of the computational demand.

Some contributions are dedicated to permutation tests on the significance of model parameters in the literature. A non exhaustive list of contributions on this topic includes [1, 127, 128, 129, 6].

However, some distinctions about the adopted approach must be made. For instance, some authors used permutations of raw data [95, 6] while other researchers permuting the residuals of the full model [130, 131]. Moreover, some of the researchers permuted the residuals of the reduced model, and they found approximate permutation p-values [132, 133, 66, 134]. A frequent method is permuting the dependent variables but permuting the design matrix is also considered as we see in Table 2.1. Under the null hypothesis, some authors also considered the permutation of both the dependent variables and the design matrix (the explanatory variables).

Consider the regression model:

$$Y = X\beta + Z\gamma + \epsilon \tag{2.3}$$

Where Y is the response variable, X is the main explanatory variable with regression coefficients β, Z is nuisance variables with nuisance parameter γ, and ϵ is the error terms. Then, we summarize different methods of permutations from the previous studies in Table 2.1, where P is the permutation matrix, $\tilde{H}$ is ride estimates, λ is regularization parameters, R is the residuals, $\tilde{R}$ is ridge residuals. Finally, we refer the reader for detailed definition, explanation, and notation used for the list of considered permutation methods [1, 128, 6].

Table 2.1: Linear model after permutation

Method	permuted Model
Draper–Stoneman	$Y = PX\beta + Z\gamma + \epsilon$
Still–White	$PRZY = X\beta + \epsilon$
Freedman–Lane	$(PRZ + HZ)Y = X\beta + Z\gamma + \epsilon$
Manly	$PY = X\beta + Z\gamma + \epsilon$
ter Braake	$(PRM + HM)Y = X\beta + Z\gamma + \epsilon$
Kennedy	$PRZY = RZX\beta + \epsilon$
Huh–Jhun	$PQ'RZY = Q'RZX\beta + \epsilon$
Smith	$Y = PRZX\beta + Z\gamma + \epsilon$
Freedman–Lane HD	$(P\tilde{R}\lambda + \tilde{H}\lambda)Y = X\beta + Z\gamma + \epsilon$
Double residualization	$(P\tilde{R}\lambda + \tilde{H}\lambda)Y = \tilde{R}\lambda X\beta + \epsilon$

The inferential results based on permuting of raw data, residuals of the full model, and residuals of the reduced model more or less provided the same conclusions except for skewed data and presence of outliers [135, 6, 136]. Nonetheless, extensive simulation studies were performed to compare the powerfulness of different permutation methods of the data. Hence, among permutation methods of Table 2.1, the most accurate and reliable permutation test, which controls the type I error rate, was obtained using the Freedman-Lane permutation method [1, 128, 6].

2.3.1 Permutation Test for Testing Partial Regression Coefficients

In many practical applications, testing the significance of partial regression coefficients is quite common since the nature of the study leads to focus on a sub-set of explanatory variables. However, in multiple linear regression, the test on the significance of partial regression coefficients is a complex task also for the permutation approach since under H_o exchangeability does not apply unless the effect of associated nuisance variables is eliminated.

If the test concerns all the explanatory variables, residuals are exchangeable under the null hypothesis, and the test on the whole regression coefficients is possible. In this case, we can obtain an exact permutation test. However, computing an exact p-value might be difficult in the case of the significance test of the partial regression coefficients. An approximate permutation solution is possible by estimating the nuisance parameters and permuting the residuals of the reduced model similarly to Freedman-Lane[6].

[6] proved that the permutation distribution of the square of Pearson partial correlation coefficient obtained from different permutation principles by permuting raw data, residuals of the reduced model, and residuals of full model approximately follows the normal distribution under the null hypothesis. One advantage of their permutation test is that it can also be applied for a one-sided significance test of partial regression coefficients. They simulated data to compare power behavior of permutation test based on the Freedman-Lane and Kennedy permutation method with t-test under the null hypothesis and alternative hypothesis for different sample sizes, the number of predictors, and different error distributions. The simulation results revealed that the Kennedy permutation method inflated the type I error rate, especially for a small sample size. While, the Freedman-Lane permutation method provided a rejection rate under null hypothesis similar to α regardless of the number of predictor variables, sample sizes, and error distributions. However, their methods work only for linear models. In our opinion, the extension to the general mixed model is possible by considering Kendall correlation coefficients instead of Pearson partial correlation coefficients as a permutation

test statistic.

One of the most exciting applications of permutation tests in general liner models concerns problems with confounding factors [128]. When the significance of regression coefficients is tested, the confounding effect of covariates must be eliminated through a suitable permutation strategy of statistical units. Hence, the permutation test statistic is computed as a function of the observed data of the predictors. However, since the parameters of the confounding factors are unknown the permutation solution provided an approximate result.

In what follows, the design matrix is partitioned into two parts, such as predictors and confounding factors. Similarly, the vector of regression coefficients is also partitioned. They explained how the exchangeability condition is considered under the null hypothesis (some variables have zero effects) for different error term structures such as independent symmetric errors, exchangeable errors, and block dependent error terms. In case of dependency of the error terms due to blocks, the exchangeability could be applied by permuting units within blocks or exchanging the whole blocks (see the formal reasoning in [128]).

The permutation test statistic could be based on the estimated coefficients [128]. However, they recommended a pivotal test statistic similar to Anderson's, such as F statistic, t-test statistic, Pearson's correlation coefficient, and coefficient of determination. Furthermore, they considered an extensive simulation study to compare the power behavior of different permutation methods (see Table 2.1) with the parametric method based on the F test under different scenarios such as skewed distributions, balanced and unbalanced sample sizes, equal and unequal variances, and even in the presence of outliers. The permutation approach to account for the outliers is interesting. Under the null hypothesis that the model coefficients are zero, the Still–White and Kennedy solutions are conservative, especially for small sample sizes. On the other hand, Freedman–Lane and Smith's proposals seem to perform well under the null hypothesis respecting the nominal level α.

Moreover, the power behavior of the permutation test based on Freedman–Lane under the alternative hypothesis outperforms those of other permutation methods and the parametric F test regardless of sample sizes, error distributions, and other conditions.

Apart from the violation of assumptions, in some cases, the parametric tests are not feasible for hypothesis testing of parameters in a linear model, for example, the presence of outliers with small sample sizes, mixed error distributions, and nonrandom sampling. To overcome these problems[129], proposed an R package called lmPerm, which is based on the permutation approach, for computing the p-values of the tests on the significance of regression coefficients. Also, this package is suitable for analyzing polynomial models and multivariate responses. They noted that permuting rows of observations is possible even in the presence of interaction effects. They considered the simulation study to investigate the power behavior of the permutation test based on the lmp function and the standard F test. The simulations revealed that the F test is conservative and less powerful for small sample sizes, especially under non-non-normal distributions. Regarding the outliers, the power of the parametric F test decreases as the percentage of outliers increases and the sample sizes decrease.

The permutation test is performant even in a high percentage of outliers and small sample sizes. In the case of heterogeneous variances, the two methods have the same power performance. However, a limitation of this method is that it is computationally intensive when an exact permutation strategy is considered (all permutations of the data are taken into account). They proposed other methods based on stopping sampling upon meeting given criteria to overcome this problem. For example, the estimated standard deviation of p-values is less than 0.1. However, estimation of standard deviation may not be convenient for some reasons, and it is better to approximate the permutations distribution via the conditional Monte Carlo procedure.

2.3.2 Permutation Tests for Longitudinal Model Coefficients

Due to heteroscedastic and autocorrelated errors, estimating and testing the significance of coefficients of the panel, longitudinal, and mixed-effects models using standard parametric methods may be problematic. In contrast, the permutation approach has at least an approximate solution for such problems[137]. However, sometimes exact results based on the permutation approach for autocorrelated error terms might be challenging since the exchangeability condition may not work in this circumstance.

On the other hand, in the case of non-autocorrelated error terms, for instance, for the general fixed-effect model, an exact permutation solution can be obtained, even in the presence of heteroscedastic error variance. [138] proposed an exact multivariate permutation solution for mixed variables within the fixed effect framework even in heteroscedastic variance. Their permutation solution could account for the model misspecification. However, this cannot be generalized in the case of unbalanced sample sizes.

One of the advantages of the permutation tests is their robustness to departure from error normality and their capacity to consider the multivariate responses' dependence structure. However, their theoretical proposal needs validation through an extension of the simulation study to know how the power of permutation test is performant to test the fixed effect model coefficients in comparison with the standard parametric test under different scenarios by considering different levels of multicollinearity, correlated responses, small sample sizes, unequal variances, and unbalanced sample sizes. Moreover, an extension to test the significance of general mixed effects is possible.

Furthermore, in the linear mixed model, one concern is the inclusion and exclusion of the random component, which accounts for the individual heterogeneity in the longitudinal study. In a linear mixed model, the test on the random effects is crucial. In what follows, the standard parametric tests such as score test, Wald test, and likelihood test might not be applicable under some conditions, for instance, small sample sizes [139]. In addition, as Lee and Braun explained, the parametric tests do not have some approximate distributions under the null hypothesis, such as a chi-square distribution, since the variance component of the random effect is zero under the null hypothesis. Moreover, the significance test of random effect coefficients, using a parametric test, is not robust to the distribution of the random component.

However, the permutation approach can be applied to test the significance of random effect parameters. [139] proposed permutation test statistic based on the best linear unbiased predictions (BLUPs) and the restricted likelihood ratio test for testing the random effect coefficients in the linear mixed model (LMM). When they are concerned with single random effect parameters, the testing procedure is equivalent to testing whether the variance component is zero or not. Under null variance, they permuted the estimated residuals. The estimated residuals are obtained by subtracting the fixed part from the response. Whereas, when they are concerned with the significance of some random effect parameters, the effect of nuisance components was eliminated by taking the appropriate weight matrix. The simulation results revealed that the permutation test based on BLUP and the restricted likelihood ratio test respected α under the null hypothesis. In contrast, the parametric test based on the asymptotic likelihood ratio test in some cases, especially with small sample sizes, is anti-conservative because of the inflated type I error rate. The power behavior under the non-zero variance component or non-zero random effect revealed that the permutation test is better than the counterpart of the asymptotic likelihood ratio test regardless of error distributions and the sample sizes.

Nonetheless, the BLUP permutation test statistic is not a reasonable solution when we have two or more random coefficients. Moreover, a simultaneous testing method was proposed based on the restricted likelihood ratio test approach that requires estimating the variance component at each permutation.

In addition, hypothesis testing on generally mixed model coefficients in the case of multivariate response is also another challenging problem within the standard parametric approach based on the likelihood ratio test [137]. The challenge might arise due to the dependency among the response variables. They proposed a permutation strategy for testing the significance of vector generalized linear mixed models (VGLM) parameters in the presence of factors regardless of the known multivariate distributions. In particular, they proposed a permutation method to test the significance of the interaction effect between units for the family of VGLM. The proposed method can be applied for testing coefficients of linear models, nonlinear models, general linear models, mixed models, and general mixed models, beyond as mentioned VGLM whereas, the permutation solution of[6] couldn't be applied to VGLM parameters.

The permutation solution is flexible and applicable to any model (linear or non-linear model, with continuous or binary response). For example, a combined permutation test could test the significance of coefficients of both fixed effects and random effects of a general(linear) mixed model.

Testing for the interaction effects with the permutation approach, some of the coefficients are considered nuisance parameters, and their corresponding variables are eliminated in the analysis through residualization. In this case, the exact permutation test is possible when the exchangeability condition is fulfilled, and this might not be the case if we have two or more factors so that an approximate significance test is possible. [137] simulated data from two samples for both univariate and multivariate cases under different distributions to compare the power behavior of the exact permutation test, the asymptotic permutation test, and the likelihood ratio test. They found that both the exact and asymptotic permutation test provided rejection rates close to α under H_o for balanced sample sizes. In comparison, the likelihood test is conservative. The power of the permutation test is greater than the likelihood test regardless of constant variance and normality. However, further simulation studies may be designed to investigate how mild multicollinearity affects the type I error rate and the power of the permutation test.

2.4 Permutation Tests for Dependent Samples

This section is dedicated to the permutation approach for the comparative study of dependent samples. We discuss the permutation (M)ANOVA for longitudinal datasets and two-sample or multi-sample tests for paired data or repeated measures. We start from the simple repeated measures experimental design or tests for paired samples, and we extend the review to more complex designs. The typical parametric solution for the two-sample test on location for dependent samples is the so-called paired t-test. However, this solution is not appropriate for small sample sizes and non-normal data [55].

Furthermore, when the error distribution is unspecified, the nonparametric approach based on the Wilcoxon signed-rank test is used for this problem. Nonetheless, the nonparametric rank test also requires continuous distribution and homoscedasticity to units [4]. The re-sampling method, in particular, permutation test for such problem can also be applied to discrete data, assuming unequal variances between units, and it is more powerful[140, 84, 141, 142].

[55] studied the permutation solution for paired data, comparing it with the nonparametric Bootstrap and the parametric paired t-test in case of small sample sizes and skewed distributions. Under the null hypothesis, the paired sample problem, the observations are exchangeable within the statistical units. Therefore, the permutation inference requires the mild condition of exchangeability under the null hypothesis [52, 39, 54]. However, the permutation test can be applied under some conditions for hypothesis testing of location or scale parameters under exchangeable and non-exchangeable error terms. See, for example, the permutation approach for paired samples studied by [143, 55, 109, 144].

They assumed three simple conditions in which the permutation principle can work to obtain an asymptotic p-value. Those are the distribution-free nature of the test, the convergent limiting distribution of the test, and consistency of the test [145, 146, 147]. A simulation study revealed that the standard paired t-test was conservative for skewed distributions. The bootstrap method provided a liberal type I error rate for small sample sizes. The permutation test respects α under the null hypothesis regardless of the dependency among paired samples, sample sizes, and skewness of the distributions. The permutation test is more powerful than the paired t-test, especially for skewed distributions.

Analogous to the comparative study of mean and variance for independent samples, hypothesis testing of mean and variance simultaneously is also a fascinating area of research in the paired sample. In other words, apart from the average shift between paired samples, the change in variance that arises due to individual heterogeneity and treatment in paired data might be the focus of the research. In this regard, [145] introduced the multi-aspect permutation procedure for testing mean and variance for paired data. Furthermore, they are interested in testing the effect of treatment on the first and second moments of paired data. Hence, the null hypothesis is the equality of moments (first and second) of pre-treatment and post-treatment.

Accordingly, different partial permutation test statistics for each aspect were considered based on the sum of observed data differences and squared of observed data, respectively. Finally, they combine the p-values of the partial permutation tests to obtain a global p-value using the nonparametric combination method [39, 140, 84]. Finally, they used different resampling tests to investigate the comparative performance as a function of the dependency of paired responses: Pitman's test based on Pearson's product-moment correlation coefficient, jackknife test based on the ratio log of the sample variances, and Spearman's rank correlation coefficient. The simulation results revealed that the permutation test has a good power behavior to study the effect of treatment on the mean and variance of paired data in the presence of correlation among pre-treatment and post-treatment observations. Moreover, the simulation supported that the power of the test is also good under normality and for small sample sizes.

Furthermore, parametric ANOVA for repeated measure designs also requires certain assumptions (such as normality, constant variance, and no correlation between units) to test the mean difference among groups over time. For this reason, a robust and flexible test about the mean difference among groups in repeated measure designs requires a distribution-free test such as a permutation test. [148] deeply explained the split-plot analysis for repeated measure experiment when the assumption of constant correlation is satisfied. They noted that in case of violation of parametric ANOVA assumptions, obtaining an exact p-value is not possible, and a nonparametric test is crucial.

However, the nonparametric hypothesis testing using the permutation approach also needs careful consideration of what to permute [148]. According to them, in repeated measure designs, "*...The permutations should correspond to the randomization*". In other words, for experimental design, the permutation is possible under the null hypothesis in which observations are exchangeable due to randomization. For instance, in a case-control study of repeated measure designs, row vectors of the observations are permuted among case and control groups. However, permutation of the measurements within statistical units is not possible since the randomization does not apply[148].

The null hypothesis may be equality of means for different follow-up times in some repeated measure designs. Hence, in this case, under the null hypothesis, the observations within statistical units are observed randomly, and permutation of measurements within statistical units is possible. Although the principle of permutations, corresponding to randomization, is restricted to the experimental design, for observational or survey study, some mild conditions are sufficient such as independence of observations to perform the permutations under exchangeability.

In the literature, if the permutation test is exact, the probability of rejecting the true null hypothesis does not exceed the preassigned level of significance under the null hypothesis of no treatment effect [39, 12]. The permutation test could provide an exact and approximate p-value depending on the exchangeability condition under the null hypothesis in repeated (M)ANOVA design or general mixed models [57]. They used different permutations in repeated measure designs to obtain exact and approximate results, such as permuting reduced residuals (balanced and unbalanced design) and modified residuals. According to them, for a balanced design, the reduced residuals could be permuted under the null hypothesis if it results in zero expectation after removing the effect of nuisance variables. In other words, the observations of the residuals are permuted in a row-wise fashion rather than permuting within individuals. They provided a necessary and sufficient condition to define the reduced residuals for a mixed linear model or repeated measures.

In addition, in the case of the non-spherical distribution of residuals, an exact p-value is obtained by modifying the reduced residuals and removing the correlation within subjects. In repeated measure design or mixed linear model, the random effect does not affect the principle of permuting the observations. In other words, under the null hypothesis, the outcome variable is equal to the measurement error term and the random error term. As a result, since the random error term has zero expectation under the null hypothesis, it does not affect the permutations. They designed a simulation study to compare three types of permutations (raw data, reduced residuals, and modified residuals) with the counterpart of the F test for repeated measures ANOVA under different distributions.

As a result, the simulations provided an approximate type I error rate close to the nominal level for all permutation methods. Furthermore, the power behavior of the permutation test based on permuting the reduced residuals, the raw data, and modified residuals provided similar power performance regardless of non-normal distributions. Moreover, the power behavior of the permutation test was performant over the standard F test for repeated ANOVA, especially for non-normal distributions. However, the modified residuals permutation method needs to estimate the Eigenvalues at each permutation to obtain an exact p-value, and it might increase the computation time.

In a comparative study of dependent samples for multivariate outcomes, apart from the dependency among the response variables, there is also dependency among the observations over the follow-up time, known as intraclass correlation. Consequently, the power of the parametric test is not optimal if we are not controlling the intraclass correlation [149]. In other words, besides normality and constant variance, hypothesis testing of repeated measures or longitudinal analysis using the parametric test for two or more groups requires the equality of correlations. However, a steady correlation in follow-up studies is unlikely to hold since observations within statistical units are highly related over time. For this reason, assuming constant correlation within individuals seems impossible. For instance, [150] proposed permutation test regardless of the dependencies among the observations for testing treatment effects in multivariate repeated measures or longitudinal MANOVA setting. The permutation procedure in their study is just taking place by permuting the pooled data regardless of the dependency within subjects. Then they considered the permutation version of Wald test to test the mean difference among groups over time for the multivariate longitudinal dataset. They proved that the Wald type and studentized Wald type permutation statistic approximates the same distribution under the null hypothesis. Since the parametric Wald type test requires a large sample, they considered a simulation study to investigate the power improvement of Wald type permutation test statistic and ANOVA type statistic for small sample sizes under various distributions and covariance structures.

Consequently, the simulation results revealed that the Wald type statistic provided a liberal type I error rate for small sample sizes and high follow-up time points. Besides, the type I error rate exceeded the nominal level for testing the interaction effects. The ANOVA type statistic provided an approximate significance level for small sample size and normal distribution, and it provided a conservative type I error rate for high follow-up time and non-normal data. The proposed Wald type permutation statistic provided an exact nominal level regardless of the sample size and error distribution despite the dependencies of the observations. However, the Wald type permutation statistic is also quite conservative for high follow-up periods and testing interaction effect in a longitudinal study.

2.5 Permutation Approach for Missing Data

The most challenging problem in statistical data analysis, especially hypothesis testing, is missing observations. In particular, the missing value is the most frequent problem in follow-up studies, such as in longitudinal data, panel data, survival data, and time-series data. Moreover, the missingness issue is also quite common in the cross-sectional study. Missing observation is a general term that comprises censored as well as truncated values [151, 152]. The missingness means that the researcher could not record the observations for some statistical units for various reasons, such as the available value, dropouts, detection limit, and survey nonresponse. Therefore, most parametric methods are not feasible for hypothesis testing in the presence of one or more missing values in the dataset.

In the literature, the parametric solution for missing data problems is highly relying on the stringent assumption of missing completely at random (MCAR) missing mechanism[39]. In other words, the probability of a missingness value is unrelated to treatments, censoring process, and observed values of the outcome variables [153, 154, 155]. Moreover, when the missing data process is missing not at random (MNAR), the parametric methods are not flexible to account for the missingness of the data. Whereas distribution-free testing methods, such as the permutation test overcome the problem of missing data (including censoring and truncation) conveniently and efficiently [156, 35, 157, 39, 158].

In paired designs with missing data, the commonly used nonparametric log-rank test might not be the general solution in some situations, such as small sample sizes. For this situation, some researchers propose the use of a permutation approach[159, 157].

In particular, [157] use a permutation test for a comparative study of pre-treatment and post-treatment effect for incomplete paired data. First, they consider two permutation test statistics computed from paired and incomplete data using the mean differences. Then, they take the linear combination of the paired and unpaired statistics to construct the overall permutation test statistic. Finally, they simulate data under normal and skewed distributions to investigate the power behavior of the permutation test by comparing it with standard paired t-test and Bhoj t-test based on paired and unpaired data [160]. The simulation results revealed that the parametric paired t-test doesn't control the type I error rate. On the contrary, the permutation test and the Bhoj t-test controlled the type I error rate and provided higher power regardless of the underlying distributions and sample sizes.

However, their method requires estimating the combining parameter of the two permuted test statistics. In our opinion, after considering two different partial permutation tests for paired and unpaired data, the partial information can be combined according to the nonparametric combination approach [39].

The censoring problem, specifically right censoring, is quite common in survival analysis. Hence, a nonparametric approach such as permutation test is useful [156, 161, 35]. In particular, there is an interesting study about the permutation solution for comparing the survival curves of two independent samples in the presence of right censoring data [158]. As they explained, the distribution of right censoring might be equal or unequal in the two independent samples. In what follows, if the distributions of the right censoring are equal under the null hypothesis, the censoring happens according to a random process, and it can be ignorable [158].

However, unequal censoring distributions under the null hypothesis may affect the estimation and hypothesis testing. [158] considered the composite hypothesis of equality of survival curves for time to the event and censored value. They broke down the null hypothesis into sub hypotheses one for each aspect, and the null distribution of the partial permutation test statistics are derived by permuting the sample data. Their solution can be extended to the missing data problem following the nonparametric combination procedure [39]. The inference about the observed failure time is considered conditional on the censoring value[158]. They designed a simulation study to investigate the power performance of the permutation test compared to the log-rank test and weighted Kaplan-Meier test for both equal censoring and unequal censoring. The simulation results prove that the permutation test is reliable and robust for equal censoring, especially for the small sample sizes. Overall the permutation test is performant to account for the censoring problem without specifying the underlying distribution and modeling the dependency structure between survival time and censoring.

2.6 Some Remarks

The use of permutation tests for data analysis in many fields has increased. It is a vital statistical tool for analyzing complex datasets when parametric tests are not applicable or not powerful due to the violation of assumptions. Permutation test tackles Behrens-Fisher problem in hypothesis testing. The use of pivotal statistics is neither necessary nor sufficient for a permutation test. Moreover, exchangeability is a sufficient condition in the permutation principle. The permutation strategy can be applied in the paired sample by permuting the signs. Whereas, in repeated measure designs, the permutations should correspond to the randomization[148]. Besides, permutation tests can be applied for testing coefficients of the linear models, nonlinear models, general linear models, mixed models, and general mixed models. Last but not least, the permutation test can solve hypothesis testing problems of directional alternatives, umbrella alternatives, location-scale parameters, missing data problems, and others.

CHAPTER 3

ADVANCES ON POWER AND ASYMPTOTIC PROPERTIES OF COMBINED PERMUTATION TESTS

3.1 Combined Permutation Tests for Two-sample Problems

The parametric tests cannot be applied when the number of variables is greater than sample sizes since the degrees of freedom must be positive[39]. Furthermore, many stringent assumptions make parametric tests less flexible and less powerful[5, 4]. On the other hand, combined permutation tests are a general tool for testing several multivariate problems under mild condition[39].

Many researchers' studies combined permutation tests regarding their assumption, applications, properties, and robustness[4, 39]. However, the literature lacks evidence on the power performance of combining functions concerning the percentage of true partial alternative hypotheses for two-sample problems. Hence, we study the power behavior of CPT for different percentages.

Consider the dataset:

$$\mathbf{X} = \{\mathbf{X}_j, j = 1, 2\} = \{X_{qji}, i = 1, \cdots n_j, j = 1, 2, q = 1, \cdots Q\} \qquad (3.1)$$

Which takes value on the Q-dimensional space χ, for which σ algebra $\mathcal{A}$ and a nonparametric family $F_j \in P$ (non-degenerate unknown distribution) is defined. Where $\mathbf{X}_j$ is the dataset of the j^{th} group.

The null hypothesis is equality of multivariate distributions of responses on two groups:

$$H_o : \{F_1 = F_2\} = \{\mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2\} \tag{3.2}$$

Let the null hypothesis H_o of a testing procedure is broken down into k sub-null hypotheses

$$H_{o1}, \ldots, H_{ok} \tag{3.3}$$

and the global null hypothesis

$$H_o : \cap_{i=1}^{k} H_{0i} \tag{3.4}$$

is true if all the k partial hypotheses are jointly true. Likewise, the alternative hypothesis H_1 is broken down into k partial alternatives and

$$H_1 : \cup_{i=1}^{k} H_{1i} \tag{3.5}$$

is true if at least one partial null hypothesis is false.

Besides, let

$$\mathbf{T} = \mathbf{T}(\mathbf{X}) \tag{3.6}$$

is k-dimensional vector of test statistics and each component of $\mathbf{T}_i(\mathbf{X})$ be suitable to test each partial hypothesis H_{oi} against H_{1i}. Hence, the global multivariate test to test the significance of global null hypothesis is obtained by combining the k-partial tests, by using a suitable combining function. The most widely used combining functions are the Tippett the Fisher and the Liptak [5, 39].

We are performing simulation studies about the power performance of different CPTs in the case of two samples. Hence, the preliminary results show that the most powerful CPT is based on the Tippett combination when the percentage of true partial alternative hypotheses is $\leq 30\%$, that based on the Fisher combination when the percentage is $> 30\%$ and $< 100\%$, and that based on the Liptak combination when the percentage is 100% (see Figure A.1 in the appendix).

3.1.1 Application of Permutation Approach for Sustainability

The sustainability agenda has been among the hot topic around the globe. Initially, the idea of sustainability seems only environmental protection and limited to developed nations[162, 163]. However, it comprises the environmental, economical and social aspects[164, 163, 165].

Since sustainable development goals (SDGs) are ongoing issues, obtaining a reliable dataset is problematic. The united nation developed the 17 standard indicators with 169 targets in 2015[166]. The substantiality indicators are a vital tool to quantify the SDGs[167].

Due to high conflicts in the neighboring countries in the last two decades, Ethiopia is the second-largest refugee-hosting country next to Uganda in sub-Saharan Africa countries [168]. The refugees from Eritrea, Somalis, South Sudanese, and Sudanese have an impact on the source of conflicts in Ethiopia. Hence, we investigate the SDGs (social, environmental, and economic sustainability) by assessing the living condition of refugees and the hosting community households in Ethiopia.

The dataset is obtained from the World Bank[168], Ethiopian skills profile survey 2017. There are 5 separated household refugees and 5 host communities within a 5kms radius, in the Somalia region of Ethiopia, in Buramino camp (code 33). There are multidimensional sustainability measures in terms of the standard of living of households. We have Q=30 variables (partial tests) in the study.

Parametric tests are not suitable due to correlated responses, one-sided alternative hypotheses, mixed variables, and the presence of a high number of variables with small sample sizes. We want to test:$H_o : \{F_1(X) = F_2(X)\}$ and $H_1 : F_1(X) \leq F_2(X)$, where H_1: the standard of living for separated refugees is a worse than the host community households in Ethiopia.

The Fisher combining function provides the p-value of 0.013, which is less than $\alpha = 0.05$, and we carry out a posthoc analysis in Table A.11 (see appendix). Thus, the separated refugees living in the Buramino camp have worse standard of living than the host community households in Ethiopia. This result is consistent with the World Bank report[168].

3.2 Advances on Permutation Multivariate Analysis of Variance for Big Data

In many multivariate analyses of variance (MANOVA) applications, the classic parametric solutions for testing equality hypotheses in population means or multisample and multivariate location problems might not be suitable for various reasons. For instance, the stringent and implausible assumptions of iid observations and multivariate normality are the main reasons for considering parametric methods neither flexible nor robust and consequently often unsuitable. Moreover, in the presence of big data with a high number of response variables, great attention should be paid to when the number of response variables is larger than the sample sizes because of the loss of degrees of freedom. Even if there is not a unique definition, in statistics, a dataset is usually classified as "big data" if it represents a collection of informative data, extensive in terms of volume, velocity, and variety, such that specific analytical technologies and methods are required for the extraction of value or knowledge [169]. Big data are typical of many empirical disciplines such as biomedicine, economics, biology, ICT, education and research, financial services, social media, automotive industries, e.tc[170]. Frequently, the high volume of big data depends on the multivariate nature of the dataset due to a large number of variables. In addition, the variety of big data, due to the presence of different types of variables (quantitative and qualitative) and to the variability and heterogeneity of data, makes inferential problems more complex and requires robust and valid techniques to make inferences. For instance, in studies focused on social media, text, video, audio, and image data are jointly analyzed. Hence, tests of hypotheses for big data must be addressed with appropriate methods that lead to reliable decisions in short times and take into account the variability and heterogeneity of the information.

A typical approach to variable-oriented multivariate problems consists in the application of exploratory methods based on dimensionality reduction such as principal component analysis (PCA) or factor analysis (FA)[171, 172]. For two-sample multivariate testing problems, a typical solution is the Hotelling T-square test in the presence of numeric data. These methods are based on stringent assumptions such as the linearity of the relationships between variables or normality.

Linearity is an extreme and often unrealistic assumption. On the other hand, normality is a reasonable assumption only with large sample sizes due to the asymptotic properties of the statistics. Nevertheless, even in cases where linearity and normality are reasonable assumptions, especially in inferential problems, in many variables, the estimation of a large number of unknown parameters, such as covariances or correlations, is required. Moreover, when the sample size is less than the number of variables, a problem related to the degrees of freedom arises and some typical parametric methods, such as the Hotelling T-square test, are not applicable.

In such problems, nonparametric methods are preferable because they don't require that the underlying probability law belongs to a given family of distributions and no parameters need to be estimated. In particular, permutation tests follow a distribution-free approach and are almost as powerful as parametric methods based on normality when this assumption is true but much more powerful when the true underlying distribution deviates from the Gaussian [5, 83].

Solutions for multivariate tests within the family of permutation methods consider the dependence between response variables without modeling it explicitly, and consequently without the need of estimating parameters or assuming linearity [173, 4, 105]. Permutation solutions for multivariate location problems have been proposed and studied mainly in terms of power and robustness to the underlying distribution, especially comparing their performance with that of the classic parametric tests[80, 83, 5]. An exciting proposal is based on the combination of the univariate permutation tests of the marginal variables [5]. [173, 39] proved that the power of the most commonly used combined permutation tests, with fixed sample size and a divergent number of variables under the alternative hypothesis, tends to one in the two-sample problem.

A different combined test is obtained according to the combining function used. Hence a deep study to compare different combined tests, especially for big data with a large number of variables, is essential and suitable in order to find the most powerful test under different scenarios. However, for the multivariate multisample location problem, a comparative study of the power behavior of the most crucial combined permutation tests as the number of variables diverges is missing. In particular, it is useful to know under which conditions each of the different tests is preferable in terms of power, how the power of each test increases when the number of variables under the alternative hypothesis diverges, and the power behavior of each test as a function of the proportion of true alternative hypotheses.

This chapter aims to fill this gap in the literature about combined permutation tests. The paper is organized as follows. Section 2 is dedicated to a literature review of the MANOVA problem. The method of the combined permutation test is described in section 3. In section 4, comparative simulation study results are reported and discussed. In section 5, the application of the method to a real case study is presented. Finally, the conclusions are in section 6.

3.3 Literature Review

The goal of several empirical studies is the comparison of two or more populations in the presence of multivariate response variables. Often, regardless of the number of factors, the problem consists in the testing significance of treatment effects or the presence of a shift in some location parameters. The variation of population means was investigated using multivariate analysis of variance (MANOVA). Various parametric multivariate tests based on strong assumptions have been proposed to test whether there is a significant difference between group means. The most commonly used are Hotelling T-square test [100], the test of[101] and the proposal of [102]. The main assumptions of these tests are normality, constant variances, and continuous responses. Moreover, these methods cannot be applied for big datasets when the number of response variables exceeds the sample size.

Nonparametric solutions have been proposed to overcome the limits of the parametric tests due to the lack of robustness [39, 4, 80, 174]. For instance, [83] introduced a nonparametric solution based on the permutation test for an ecological problem. The permutation test statistic was the Fisher F ratio obtained from a distance matrix, and the simulation results proved the appropriateness of the permutation test for both one-way and two-way MANOVA. [80] studied the accuracy and power of permutation tests for MANOVA based on different test statistics. According to his study, the sum of squares between groups with the Euclidean distance was preferable to the Chord distance and the sum of Fs of univariate ANOVA. Moreover, the simulation study revealed that the permutation test was powerful also under heteroscedastic and with unbalanced samples.

Several works concerning applications of permutation tests for one-way and two-way MANOVA have been published in the literature. A non-exhaustive list includes the following papers[99, 175, 176, 113, 177, 111, 79, 105]. However, the extension of the permutation test for two-way MANOVA requires great attention in permuting the statistical units between groups because the exchangeability condition is guaranteed only within the levels of one factor by considering the second factor as a block. Thus, constrained permutations are essential[83]. The two-sample multivariate problem has been frequently considered[39, 178]. Instead, the multisample case has been addressed by fewer authors (see [174]). In some cases, permutation solutions for complex problems such as for multi-aspect tests are introduced[179]. In addition, for directional alternatives [180, 116], and tests for categorical data [62, 180] have been developed. This paper focuses on multisample location problems for numeric variables and non-directional alternative hypotheses.

3.4 Methods

3.4.1 Multivariate Permutation Test

The permutation test is a distribution-free test based on the assumption of exchangeability under the null hypothesis [5]. To apply the permutation principle, the sample data are partitioned into groups based on the treatment levels in an experimental study and pseudogroups in an observational study. To this end, the structure of the dataset for $S \geq 2$ independent samples and V-dimensional response is represented by:

$$\mathbf{Y} = \{\mathbf{Y}_{igq}, i = 1 \cdots n_g, g = 1 \cdots S, q = 1 \cdots V\} \tag{3.7}$$

The dataset $\mathbf{Y}$ takes values on the V-dimensional sample space Ω for which a σ-algebra A and a nonparametric family P (non-degenerate unknown distribution) are defined and supposed to be exchangeable.

Hypothesis testing based on the permutation approach requires a clear formulation of the null hypothesis. The null hypothesis in the MANOVA problem is defined as the equality of S multivariate (unknown) distributions:

$$H_o : \{P_1 = \cdots = P_S\} = \{Y_1 \overset{d}{=} \cdots \overset{d}{=} Y_S\}. \tag{3.8}$$

Under homoscedasticity, the difference between the groups is due to a shift in location. Thus, the null hypothesis could be formulated as equality of group means for each response variable. Let Y_g be a V-variate numeric random variable such that $Y_g = \mu + \delta_g + \epsilon_g$, with μ vector of unknown parameters, $\delta_g, g = 1 \cdots S$, vectors of treatment effects and $\epsilon_g, g = 1 \cdots S$, exchangeable random vectors that follow an unknown probability distribution with equal variance-covariance matrix Σ and such that $E(\epsilon_g) = 0$.

The null hypothesis is:

$$H_o : \{\delta_1 = \cdots = \delta_S = 0\} \tag{3.9}$$

A further decomposition of the null hypothesis with respect to the marginal distributions of the multivariate response can be considered. The multivariate hypothesis can be broken down into V partial null hypotheses:

$$H_o : \cap_{q=1}^{V} (\delta_{1q} = \cdots = \delta_{Sq} = 0) \equiv \cap_{q=1}^{V} H_{oq} \tag{3.10}$$

The intersection symbol means that the null hypothesis of the overall problem is true if all the V partial null hypotheses are true. Accordingly, with a similar approach, the alternative multivariate hypothesis H_1 of inequality in distribution may be represented as follows:

$$H_1 : \cup_{q=1}^{V} \bar{H}_{oq} \tag{3.11}$$

Where the union symbol indicates that the alternative hypothesis is true if at least one partial null hypothesis is false and $\bar{H}_{oq}$ denote the negation of the q^{th} partial null hypothesis. It is worth noting that directional alternatives are also possible, but this paper focuses on two-tailed multisample multivariate problems.

When the overall null hypothesis is true and equality in distribution holds, the vector of V observations concerning a generic statistical unit comes from any of the S populations with equal probability. In other words, exchangeability of the units with respect to the populations/samples is satisfied. In order to determine the null distribution of the test statistic, all the possible assignments of the n units to the S samples can be considered. By assuming that the n_1 units of the first sample correspond to the first n_1 rows of the observed dataset $\mathbf{Y}$, the n_2 units of the second sample correspond to the next n_2 rows of the dataset, and so on, until the n_S units of the S^{th} sample that correspond to the last n_S rows of the dataset. Each possible assignment is equivalent to a permutation of the rows of the dataset or to resampling without replacement the n units with $n = n_1 + n_2 + \cdots + n_S$.

Frequently, for computational convenience, instead of considering the exact test, based on all the $n!/(\Pi_{g=1}^{S} n_g!)$ possible assignments of the n units to the S groups, a random sample of permutations is used according to the Conditional Monte Carlo method.

3.4.2 Partial Tests

The method of Combined Permutation Test (CPT) to the permutation MANOVA presented above consists of one univariate permutation test for each partial hypothesis and in combination the p-values of the univariate tests. According to the permutation distribution, the dependence between the univariate partial test statistics is taken into account in the resampling strategy by permuting the rows of the observed dataset instead of permuting the elements of the columns independently.

A suitable test statistic for each partial permutation test is the so called Treatment Sum of Squares (S_{Treat}) that depends on the deviations of the within group sample means from the total sample mean. Hence, the q^{th} partial test statistic or test statistic of the q^{th} partial test, with $q = 1 \cdots V$, is

$$T_q = \sum_{g=1}^{S} n_g (\bar{Y}_{gq} - \bar{Y}_{.q})^2 \tag{3.12}$$

With $\bar{Y}_{.q} = \sum_g n_g \bar{Y}_{gq} / \sum_g n_g = \sum_g n_g \bar{Y}_{gq}/n$, where $\bar{Y}_{gq}$ represents the mean of the values of the q^{th} variable observed in the g^{th} sample. The multivariate permutation distribution of the test statistic $T = (T_1 \cdots T_V)$ under the null hypothesis is obtained through the following procedure:

1. Compute the vector of observed values of $\mathbf{T}$ from the dataset Y:
   $\mathbf{T}_{obs} = \mathbf{T}(Y) = (T_{(1,obs)} \cdots T_{(V,obs)})$

2. Randomly permute the rows of the dataset (or reassign statistical units to groups) and compute the values of the test statistics as a function of the permuted dataset: $T^p = T(Y^p)$

3. Repeat step (2) R times independently and compute the permutation test statistics. Let $T^p_{(q,r)}$ be the value of the q^{th} partial test statistic related to the r^{th} permutation of the dataset Y^p_r. Hence $T^p_r = T(Y^p_r) = (T^p_{(1,r)} \cdots T^p_{(V,r)})$

4. Estimate the significance level function of the partial tests function:$\lambda^p_{q(r)} = \frac{\{\frac{1}{2}+\sum_r I[T^p_{q(r)} \geq T_{q(obs)}]\}}{R+1}, q = 1 \cdots V$. Where I(E) is indicator function of E that takes value 1 if E is true and 0 otherwise.The p-value of the q^{th} partial test is $\hat{\lambda}^p_{(q,r)} = \lambda(T^p_{(q,r)})$.

3.4.3 Combination

According to the method based on the combination of partial permutation tests, the test statistic for the overall problem is obtained by combining the p-values of the partial tests. The synthesis of the information provided by the partial tests regarding the marginal variables is provided by applying a suitable combining function ψ. Hence, the test statistic useful for the overall test, the multivariate analysis of variance, is $T_{comb} = \psi(\lambda_1 \cdots \lambda_V)$.

The proposal of combining p-values of partial tests in order to solve multivariate, multi-aspect, multi-strata tests, or other complex testing problems that can be broken down into partial univariate tests, appeared for the first time in the literature twenty years ago in [5] and was later studied and developed by several authors. For extended but not exhaustive reviews, see [173, 4]. Since, for the combination of the partial tests, $\psi(.)$ must satisfy some simple, mild, and easily attainable conditions, several different functions can be used, and each of them corresponds to a different solution with specific properties within the family of combined permutation tests.

A suitable combining function $\psi : (0,1)^V \rightarrow \Re$ must satisfy the following properties:

1. $\forall \lambda'_q, \lambda''_q in(0,1), \lambda'_q \leq \lambda''_q \leftrightarrow \psi(\ldots \lambda'_q \cdots) \geq \psi(\cdots \lambda''_q \cdots)$ ceteris paribus (non-increasing monotony)

2. $\exists \lambda_q \in \{\lambda_1 \cdots \lambda_V\} s.t. \lambda_q \to 0 \leftrightarrow \psi(\lambda_1 \cdots \lambda_V) \to \bar{\psi} \leq \infty$ (finite supremum)

3. $\forall \alpha \in (0,1), \exists T_{(comb,\alpha)} \leq \bar{\psi}$ where $T_{(comb,\alpha)}$ is the test critical value (finite critical value)

The most popular combining functions in the literature of combined permutation tests are Fisher, Liptak and Tippett functions. The Fisher omnibus combining function is

$$T_F = -2 \sum_q log(\lambda_q) \tag{3.13}$$

Where log(x) denotes the natural logarithm of x. Liptak's combining function is based on the transformation of the complement to one of the p-values through the inverse of the cumulative distribution function (or the quantile function) of the standard normal distribution:

$$T_L = \sum_q \Phi^{-1}(1 - \lambda_q) \tag{3.14}$$

Where $\phi(x) = P(X \leq x)$ with $X \sim N(0,1)$. Tippett combination is based on an order statistic and considers an observed value of the combined test statistic, the complement to one of the most significant p-value:

$$T_T = max_{(1 \leq q \leq V)}(1 - \lambda_q) \tag{3.15}$$

Under the null distribution, if the V partial tests are independent and continuous, the Tippett function follows the uniform distribution in (0,1). Without loss of generality, let us assume that the null hypothesis of the overall problem is rejected for large values of the combined test statistic T_{comb}. It is trivial to show that all three combination rules defined above satisfy this condition. Given that the observed value of the combined test statistic $T_{comb,obs}$.The p-value of the permutation MANOVA with the combined permutation test is given by

$$\hat{\lambda}_{(comb,obs)} = \lambda(T_{(comb,obs)}) \tag{3.16}$$

The three presented tests can have much different power behaviors under different conditions; hence a comparative analysis to deepen their properties, advantages, and limits is vital to support the analyst in deciding which test to use based on the power.

3.5 Simulation Study

A Monte Carlo simulation study investigated the power behavior of the three combined permutation tests defined in the previous section for the MANOVA problem. Different scenarios, under the null and the alternative hypothesis, were considered in order to compare the power of the three proposals as a function of the sample sizes, the number of samples, of the number of components of the multivariate response, and the proportion of true partial alternative hypothesis when H_0 is false.

Data were simulated according to the one-way MANOVA model. We considered multivariate datasets with two different sizes from the point of view of the number of responses: $V = 50$ and $V = 100$. With regard to the number of compared samples, $S = 3$ and $S = 5$ are the cases taken into account. Simulated data were generated from q-variate normal random variables (probabilistic condition most favorable to the classic parametric tests) under homoscedasticity. For all the S populations, the variance of each of the V components of the multivariate response and the correlation between any pair of variables was set equal to 1 and to 0.3 respectively. Hence, the $V \times V$ covariance matrix of each population is $\Sigma = [\sigma_{qj}]$ with $\sigma_{qq} = 1, q = 1 \cdots V$, and $\sigma_{qj} = 0.3, i \neq j \in \{1 \cdots V\}$.

Table 3.1: Rejection rates of CPTs for V=100 and α=0.05.

S	n	τ	ψ	*Proportion of true partial alternative hypotheses (p)*									
				0	*0.06*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.7*	*0.9*	*1*
3	10	0.5	F	0.047	0.082	0.108	0.25	0.462	0.642	0.776	0.87	0.918	0.924
			L	0.045	0.078	0.074	0.156	0.308	0.466	0.596	0.792	0.882	0.914
			T	0.05	0.426	0.546	0.64	0.752	0.798	0.846	0.858	0.898	0.928
		1	F	0.036	0.106	0.31	0.918	1	1	1	1	1	1
			L	0.034	0.078	0.138	0.418	0.806	0.882	0.916	0.93	0.986	1
			T	0.054	0.988	1	1	1	1	1	1	1	1
	30	0.5	F	0.056	0.104	0.24	0.89	1	1	1	1	1	1
			L	0.056	0.08	0.132	0.34	0.822	0.884	0.902	0.938	0.988	1
			T	0.058	0.94	0.99	1	1	1	1	1	1	1
		1	F	0.046	0.124	0.342	0.996	1	1	1	1	1	1
			L	0.046	0.086	0.16	0.432	0.872	0.87	0.878	0.956	0.984	1
			T	0.052	1	1	1	1	1	1	1	1	1
5	10	0.5	F	0.046	0.136	0.374	0.968	1	1	1	1	1	1
			L	0.042	0.1	0.18	0.486	0.812	0.904	0.956	0.965	0.986	1
			T	0.052	0.984	1	1	1	1	1	1	1	1
		1	F	0.052	0.144	0.359	0.988	1	1	1	1	1	1
			L	0.054	0.104	0.168	0.502	0.838	0.862	0.924	0.934	0.984	1
			T	0.056	1	1	1	1	1	1	1	1	1
	30	0.5	F	0.05	0.13	0.37	0.994	1	1	1	1	1	1
			L	0.052	0.076	0.178	0.442	0.802	0.89	0.892	0.922	0.98	1
			T	0.054	1	1	1	1	1	1	1	1	1
		1	F	0.044	0.124	0.352	0.996	1	1	1	1	1	1
			L	0.034	0.072	0.156	0.5	0.84	0.898	0.904	0.944	0.98	1
			T	0.05	1	1	1	1	1	1	1	1	1

Source: author computations. F: Fisher, L: Liptak, T:Tippett, τ: location shift, ψ:combining function

The number of simulated datasets and the number R of permutations were both equal to 1000. In the simulations, we considered the balanced design with size $n_1 = \cdots = n_S = n$. The two sample sizes taken into account are $n = 10$ and $n = 30$. In the simulations, $\mu = 0$. Let p be the proportion of true partial alternative hypothesis. Then, the V-variate normal distribution of the random variable that simulates data for the g^{th} sample ($g = 1 \cdots S$) has a vector of means with (p-1)V zeros and pV values equal to $\tau(g-1)$. Formally $\delta_g = \tau(g-1)(\mathbf{1}, \mathbf{0})^T$. Where $\mathbf{1}$ is a vector of pq elements equal to 1 and $\mathbf{0}$ is a vector of (p-1)V elements equal to 0 and T is the transpose. To consider different shifts in the population locations, the simulations were carried out with τ=0.5 and τ=1.0. Moreover, the different proportions p of true alternative hypotheses used in the scenarios are 0.00,0.05/0.06,0.10,0.20,0.30,0.40,0.50,0.70,0.90,1. The first positive proportion in the list is 0.05 if V=100 (5 true partial alternative hypotheses) and 0.06 if V=50 (3 true partial alternative hypotheses). The significance level chosen in all the scenarios is α=0.05. All simulations were carried out with the `R` programming software version 4.1.0. The authors created specific scripts for this purpose.

Table 3.1 shows the rejection rates of the tests under all different cases when the number of variables V is equal to 100. The performance of the tests under H_0 can be evaluated from the column corresponding to p=0.00 (no true partial alternative hypotheses). It is evident that, in most cases, the rejection rates are either less than or very close to the nominal α level 0.05. The test based on the Tippet combination exceeds α more frequently than the others, but the probability of wrong rejection of H_0 seems to be not far from 0.05; hence we can say that all the tests are well approximated.

When $p > 0$, the power behavior of the tests can be assessed under H_1. Unbiasedness of all the tests is demonstrated because the rejection rates are greater under the alternative hypothesis than under the null hypothesis. Moreover, the greater the sample size, the higher the power, thanks to the consistency of the tests. As expected, the power is an increasing function of the shift of the population locations that depend on τ. Finally, the greater the number of samples, the higher the rejection rates of the tests. Focusing on the effect of p on the estimated probability of rejecting H_0 when it is false, the increasing monotonic relationship is evident for all the tests. The growth rate of the power concerning p is high, and when 100% of the partial alternative hypotheses is true, the rejection of H_0 is sure or almost sure.

From the comparative analysis, the Liptak test is always the worst, except in the case in which all the partial alternative hypotheses are true. In this scenario, the power of all the combined tests tends to be one, and the tests are equivalent. In general, the lower performance of the test based on the Liptak combination is evident, and it is uniformly less powerful than the other permutation MANOVAs. In contrast, the Fisher combining function provided moderate power performance. This is consistent with [39] statement about the preferability of other tests than Liptak, except for p=1. However, when the proportion of true partial alternative hypotheses is low, the combined test based on Tippett's rule is by far the best. Also, this conclusion is not surprising, according to[5] but, in our simulation study, the extent of the difference in the performance of the test based on Tippet's function can be evaluated. Moreover, according to these results, Tippet's combination is never less performant than the others, except in the first set, when S=3, n=10 and τ=0.5 when $p \geq 0.90$, where the differences in the rejection rates of the various tests are negligible.

In Table 3.2, the rejection rates of the tests when the number of variables is V=50 are reported. Again the good performance of the test under the null hypothesis (p=0.00) is proved by the values of the estimated power. These values are usually not greater than α=0.05 even if sometimes, especially in the case of Tippet's combination, they exceed the significance level. Nevertheless, when greater than 0.05, the rejection rates under H_0 are not far from α, and the tests are well approximated. Hence this conclusion is valid regardless of the number of variables V.

Table 3.2 also confirms that the probability of right rejection of the null hypothesis of MANOVA by the combined permutation tests increases with the sample size n, with the number of samples S, with the shift parameter τ and with the proportion of true partial alternative hypotheses p. Another empirical evidence of the simulation study is that the power is generally greater with 100 variables than with 50 variables. This statement seems clear thinking to the tendency of the power to one when the number of variables diverges in the two-sample problem proved by [39]. They focus on the relationship between the power of the overall test and non-centrality parameter in case 100% of the variables are under the alternative hypothesis. According to our results, the power of the multisample tests in the case V=100 is much greater than in the case V=50 only when the percentage of true partial alternative hypotheses is low; otherwise, the difference seems not evident and always in the same direction. Hence, in our opinion, the proportion of true partial alternative hypotheses matters, and it is more important than the absolute number of true partial alternatives. For instance, when V=50 and p=0.40, the number of true partial alternative hypotheses is 20, exactly as when V=100 and p=0.20. Nevertheless, in the former case, when S=3, n=10 and τ=0.5 the rejection rates of the tests based on Fisher, Liptak, and Tippett combination are 0.626, 0.490, and 0.636 respectively; instead in the latter case, under the same scenario, 0.250, 0.156 and 0.640 respectively. Hence, even if the number of true alternative hypotheses is the same, the power of the tests based on Fisher and Liptak combinations is much lower when the proportion of true partial alternative hypotheses is smaller. Tippett represents an exception. Consider, under the same scenario, the case V=50 and p=0.20 (rejection rate 0.466) and V=100 and p=0.10 (rejection rate 0.546). Hence, with the same proportion p, the power increases with V only in the case of Tippett's combination.

In general, the case V=50 confirms that the Liptak combination is the best choice only when p=1, but in this situation, the power of the other tests is very similar. In most of the considered settings, the Tippett combination is preferable because the power quickly tends to 1 as the proportion of true alternative hypotheses diverges. When S=3, n=10, and τ=0.5, this is the most powerful test up to p=0.40. For larger values of p, it becomes the less powerful test.

Table 3.2: Rejection rates of CPTs for V=50 and α=0.05.

S	n	τ	ψ	*Proportion of true partial alternative hypotheses (p)*									
				0	*0.06*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.7*	*0.9*	*1*
			F	0.05	0.082	0.096	0.23	0.414	0.626	0.766	0.87	0.924	0.886
		0.5	L	0.054	0.066	0.074	0.142	0.258	0.49	0.668	0.828	0.912	0.888
			T	0.057	0.268	0.316	0.466	0.556	0.636	0.726	0.768	0.818	0.81
	10		F	0.042	0.054	0.26	0.892	0.996	1	1	1	1	1
3			L	0.038	0.068	0.12	0.446	0.812	0.938	0.95	0.986	0.988	1
		1	T	0.05	0.094	0.966	1	1	1	1	1	1	1
			F	0.052	0.1	0.23	0.824	1	1	1	1	1	1
		0.5	L	0.054	0.07	0.138	0.348	0.756	0.936	0.954	0.974	0.992	1
			T	0.056	0.876	0.96	0.994	0.998	0.998	1	1	0.998	1
	30		F	0.038	0.128	0.278	0.996	1	1	1	1	1	1
		1	L	0.04	0.092	0.14	0.424	0.846	0.938	0.948	0.974	0.98	1
			T	0.05	1	1		1	1	1	1	1	1
			F	0.038	0.164	0.31	0.904	0.998	1	1	1	1	1
		0.5	L	0.044	0.132	0.162	0.408	0.798	0.944	0.952	0.97	0.988	1
			T	0.051	0.946	0.994	0.996	1	1	1	1	1	1
	10		F	0.048	0.182	0.174	0.978	1	1	1	1	1	1
		1	L	0.052	0.114	0.356	0.452	0.826	0.95	0.948	0.97	0.994	1
			T	0.054	1	1	1	1	1	1	1	1	1
5			F	0.052	0.126	0.316	0.99	1	1	1	1	1	1
		0.5	L	0.048	0.076	0.156	0.458	0.836	0.94	0.956	0.976	0.996	1
			T	0.054	1	1	1	1	1	1	1	1	1
	30		F	0.051	0.136	0.348	0.986	1	1	1	1	1	1
		1	L	0.049	0.072	0.156	0.468	0.852	0.938	0.954	0.976	0.99	1
			T	0.053	1	1	1	1	1	1	1	1	1

3.6 Case Study about Organizational Well-being of University Workers

Organizational wellbeing is the first element that influences a public organization's effectiveness, efficiency, productivity, and development. Therefore, as part of objective 3 of the 2014-2016 Positive Action Plan proposed by the Equality Opportunities Office of the University of Ferrara (UNIFE), the Rector's Delegate for Equal Opportunities presented a project to promote the improvement of the working well-being of the administrative-technical staff. This project consists of the definition of interventions aimed at improving the quality of working life based on findings deriving from empirical surveys.

A questionnaire was administered to a sample of 120 employees of UNIFE in order to assess the degree of work-related stress, to detect the opinions of employees with respect to the organization and the working environment and identify possible actions to improve the general conditions of the public employees at UNIFE. One goal of the survey was also to test the existence of possible differences in organizational well-being among sub-groups of employees defined by gender and age.

The 120 respondents represent a random sample of the population of the technical-administrative staff. In order to test for the joint effect of gender and age on the organizational wellbeing at UNIFE, a simple random sample of 30 employees was selected from each of the following four groups: FU50: 50 years old or younger females, FO50: over 50 years old females, MU50: 50 years old or younger males and MO50: over 50 years old males.

The questionnaire, consisting of 79 questions, was administered to the respondents from the 4^{th} to the 11^{th} of December 2014. The Italian National Anti-Corruption Authority designed the questionnaire (ANAC) and the National Institute for Occupational Accident Insurance (INAIL) that decided to adopt a Likert scale, based on the first six integer values representing the level of agreement concerning the 79 statements (1= not at all,$\cdots$, 6=completely). The 79 statements are reported in Appendix A.10.

Let Y_{qg} be the random variable that represents the response concerning the q^{th} statement of an employee belonging to group g, with $q = 1 \cdots 79$ and $g \in S = FU50, FO50, MU50, FO50$. The following hypotheses can represent the testing problem:

$$H_0 : \cap_{q=1}^{79} [Y_{(q,FU50)} \stackrel{d}{=} Y_{(q,FO50)} \stackrel{d}{=} Y_{(q,MU50)} \stackrel{d}{=} Y_{(v,MO50)}] \tag{3.17}$$

$$H_1 : \cup_{q=1}^{79} [\exists g', g'' \in S s.t. Y_{(q,g')} \stackrel{d}{\neq} Y_{(q,g'')}] \tag{3.18}$$

The significance level is α=0.05. According to the simulation study, the most suitable testing method seems to be the combined permutation test based on the Tippett combining function. The application of this test provides a p-value of 0.755, much greater than α. Hence the null hypothesis cannot be rejected. At the significance level of 0.05, there is no empirical evidence to reject the null hypothesis of no difference in the organizational wellbeing between groups in favor of the hypothesis that the organizational wellbeing of the groups is not the same. In other words, we cannot conclude that there is a significant effect of gender and age on the employees' wellbeing. The authors carried out the analysis by creating specific `R` scripts for the implementation of the methodology.

3.7 Conclusions

The work aimed to deepen the study of the power behavior of combined permutation tests for two-samples and MANOVA problems with big data. The assessment of the convergence rate of the power to one as the number of variables increases and a comparison between the three most commonly used members within this family of tests represent the primary scientific added value of the paper.

For the two-sample test, the most powerful CPT is that based on the Tippett combination when the percentage of true partial alternative hypotheses is $\leq$ 30%, that based on the Fisher combination when the percentage is $>$ 30% and $<$ 100%, and that based on the Liptak combination when the percentage is 100%.

These nonparametric multisample location tests are well approximated, consistent, unbiased, and robust for small sample sizes. However, the power is also an increasing function of the number of samples and the number of dataset variables. The asymptotic behavior of the tests when the number of variables diverges was studied. The simulations proved that the proportion of true partial alternative hypotheses is more vital than the absolute number of dataset variables in explaining the power increase. The test based on the Tippett combination represents an exception to this general rule.

This test seems to be much more powerful than the others when the proportion of true partial alternative hypotheses is not large but competitive when the proportions are close to one. This is the only condition in which the Liptak combination test is competitive, but this test is by far the least powerful for small proportions of true alternatives.

Definitely, it seems that, among the distribution free solutions to the multivariate analysis of variance in the family of combined permutation tests, the method based on the Tippet combination is in general preferable, especially if there are no preventive information about the possible percentage of variables (or marginal distributions) under the alternative hypothesis. Instead of the Tippett combination, the Fisher rule can be applied when the percentage is close to 100%. On the other hand, the Liptak combination seems to be non-convenient in general.

This methodological tool is an essential and valuable solution to testing problems for big data, especially when the number of variables is enormous and the sample sizes are small.

CHAPTER 4

APPLICATION OF COMBINED PERMUTATION TESTS FOR TESTING MULTIVARIATE REGRESSION COEFFICIENTS

4.1 Introduction

In recent decades, following the rapid growth of efficient statistical software and the high demand for efficient data analysis methods, by industries, firms, hospitals, companies, and institutions, new statistical techniques for multivariate and complex datasets have been created all over the world by data scientists, researchers, scholars, and practitioners.

Due to the complex nature of the datasets about environmental and economic sustainability, the usual parametric testing procedure may not be a suitable method. For instance, there are datasets about firm performance, where performance is multidimensional with a high number of components [181]. Even though some researchers use unidimensional measurements of firm performance, it is unrealistic to investigate a firms' performance using a single indicator[182]. Hence, many researchers use multidimensional measures [183, 181]. For instance, [181] introduces the subjective model with nine determinants/dimensions of performance, such as profitability, growth, the market value of the firm, customer satisfaction, employee satisfaction, environmental audit, corporate governance, and social aspects.

In many studies concerning firms' performance, the scholars use different multivariate models to investigate the relationship between performance and other variables, such as linear and nonlinear models and subjective models [184, 182, 185]. To investigate the association between potential predictors and the multivariate responses representing firms' performance, one may follow a different parametric approach such as Wilk's lambda test, Pillai-Trace multivariate test, Roy's largest root, or other statistics for the MANCOVA problem.

In this chapter, we focus on the multivariate multiple linear regression model, which is popular in many fields but characterized by stringent assumptions [186, 187, 188]. In case of violating the parametric assumptions, researchers, scholars, and academicians must consider suitable remedies such as logarithmic transformation, square root, variables' removing, increase of sample sizes, and others. However, the data transformation may not be appropriate.

In some empirical studies, a typical practice is to carry out estimates and test the model by considering the regression equations separately or by assuming independence of the components of multivariate response. When the dependence is taken into account, according to the parametric approach, it must be specified and represented by a set of parameters (for example, covariance or correlation), and such parameters must be estimated. In the presence of non-normal errors, covariances or correlations are not suitable for representing the dependence.

Often, in many fields such as in clinical trials, engineering, economics, and others, researchers are willing to use a small number of sampling units or experimental units to reduce cost, time, and resources in hypothesis testing. However, they might be interested in studying many variables, such as independent variables in a multivariate regression analysis or explanatory variables in multiple regression. As a result, the statistical power of the classic parametric t and F tests on the regression coefficients is low due to small sample sizes and consequently inflated variance and loss of degrees of freedom. Moreover, the statistical inference might give misleading results.

The other more challenging problem in studying multivariate models is related to the dependence structure of the errors (and of the dependent variables), given that the assumption of independence is extreme and unrealistic. For this reason, some typical parametric approaches are not suitable for hypothesis testing of multivariate models. One of these approaches is the likelihood ratio test [186, 188].

We develop a solution within the family of combined permutation tests because nonparametric and suitable for complex testing problems that can be broken down into partial (independent tests)[81, 116, 39, 4].

The typical application of permutation tests consists of two-samples or multi-sample tests, where two or more groups are compared in terms of location or scale. For instance, many authors regarding permutation tests are frequently interested in testing the effect of treatments on the univariate or multivariate responses[189, 116, 190, 39]. On the other hand, some authors introduce the application of permutation tests on the coefficients of simple regression, partial regression, and multiple linear regression using different test statistics[6, 191, 192, 97, 21, 193]. However, to the best of our knowledge, testing the impact of explanatory variables on the multivariate response of linear models within the permutation approach has not been considered. Hence, to fill the gap in the literature, we propose a method based on the combined permutation tests for testing the validity of the model as a whole. In other words, we focus on the joint significance of the regression coefficients estimates. Furthermore, we study the power behavior of the method under different settings and compare it with the main parametric competitor, the Pillai-Trace test.

The chapter is structured as follows: section 2 is dedicated to the literature review about the multivariate regression. Section 3 is dedicated to the presentation of the new method based on combined permutation tests. Section four is about the simulation study. In section five, we present the application of combined permutation tests about private firm performance in Ethiopia. Finally, section six is dedicated to the conclusions.

4.2 Literature Review on Multivariate Regression Models

4.2.1 Basic Notions on Multivariate Regression Model

The main goal of regression analysis is to investigate the relationship between inputs (predictors, explanatory variables, regressors, and independent variables) with one or more outputs (responses or dependent variables). Regression analysis includes suitable inferential techniques to use sample data for estimation of parameters, tests of hypotheses, and predictions [186, 187, 194]. The validity of the regression analysis results depends on the satisfaction of some assumptions about the model.

The multiple regression model with more than one dependent variable is called multivariate multiple regression model (MMR) [186, 187, 188].

The multivariate regression model is widely applicable in many areas of fields such as in medicine, genetics, economics, finance, engineering, politics, economics, businesses, and others [186, 195, 187, 188].

Consider the multivariate multiple linear regression model in the equation form:

$$Y_{id} = \beta_{0d} + \sum \beta_{jd} X_{ij} + \epsilon_{id}, \quad i = 1, \cdots, n, j = 1, \cdots, p, d = 1, \cdots, V \quad (4.1)$$

That corresponds to the following matrix algebra representation:

$$Y = XB + E \tag{4.2}$$

$$\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,V} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,V} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ 1 & x_{3,1} & \cdots & x_{3,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \beta_{0,1} & \beta_{0,2} & \cdots & \beta_{0,V} \\ \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,V} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p,1} & \beta_{p,2} & \cdots & \beta_{p,V} \end{bmatrix}$$

$$+ \begin{bmatrix} \epsilon_{1,1} & \epsilon_{1,2} & \cdots & \epsilon_{1,V} \\ \epsilon_{2,1} & \epsilon_{2,2} & \cdots & \epsilon_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n,1} & \epsilon_{n,2} & \cdots & \epsilon_{n,V} \end{bmatrix}$$

Where **Y** is the nxV matrix of responses, **X** is the nx(p+1) design matrix of predictors, **B** is the (p+1)xV matrix of regression coefficient's, **E** is the nxV matrix of error terms, n is the number of statistical units, V is the number of responses and p is the number of regressors. The corresponding dth response variable has the regression model represented by:

$$Y_d = XB_{(d)} + E_{(d)} \tag{4.3}$$

Where $Y_d, \beta_{(d)}$ and $E_{(d)}$ represent the d^{th} column of **Y**, **B** and **E** respectively.

4.2.2 Assumptions of Multivariate Multiple Linear Regression

The dependent variables in linear regression analysis must be continuous. In contrast, the independent variables can be continuous, categorical, and mixed. Assumptions, estimation methods, tests of hypotheses, and diagnostic analysis of MMR are similar to those of the classic multiple linear models. The goodness of fit and model diagnostics in MMR is mainly carried out for one regression model at a time [196, 197].

The assumptions of the multivariate multiple regression model are an extension of those of the univariate multiple regression model: linearity in parameters, multivariate normality of the errors, homoscedasticity, independence of the vectors of error terms with respect to units, and others. If one of these assumptions is violated, the resulting statistical inference will give misleading conclusions[196, 194]. Consequently, we need a robust method that works under mild assumptions.

Some useful R functions in R statistical environment for data manipulation in multivariate regression are pairs and summary [198].

4.2.3 Estimation of Multivariate Multiple Linear Regression

In the model and consequently, in the estimation procedures, the design matrix or matrix of predictors is fixed and unique for all the response variables. Thus, to estimate the matrix of coefficients **B**, the OLS estimation method can be used considering the full column rank of **X**. OLS does not require any distributional assumption of the error terms[195, 194]. Manually, we can estimate **B** by minimizing the squared cross-product matrix of residuals as in a univariate case and considering all explanatory variables equally important. Under the null hypothesis $\mathbf{B} = \mathbf{0}$, the estimated $\hat{B}$ is given by:

$$\hat{B} = (X'X)^{-1}X'Y \tag{4.4}$$

The estimated dth vector of regression coefficients is given by:

$$\hat{B}_d = (X'X)^{-1}X'Y_d \tag{4.5}$$

However, when the interest is focused on a subset of the explanatory variables, the reduced form of the model is estimated using Lagrangian method under the restricted model according to constraints $\mathbf{CB} = \mathbf{0}$ [195]. Where the contrast matrix **C** is Sx(p+1), has a full row rank $S \leq (p+1)$. The estimated **B** under the restriction $\mathbf{CB} = \mathbf{0}$ becomes:

$$\tilde{B} = \hat{\mathbf{B}} - (X'X)^{-1}C'[(C(X'X)^{-1}C')^{-1}C\hat{\mathbf{B}}] \tag{4.6}$$

The R statistical software has many options to perform the estimation of a matrix of coefficients using OLS [198]. The most typical is the lm function.

4.2.4 Tests for the Multivariate Multiple Linear Regression Model

When one or more assumptions of the classic parametric approach are violated, testing the significance of the matrix of regression coefficients in multivariate regression analysis is a challenging problem. In this case, a more robust statistical method is required. For example, testing the statistical significance of the regression coefficient of the multivariate multiple regression (MMR) is frequently done using the likelihood ratio test (LRT) or Hotelling

T^2 test [188]. However, those tests require stringent assumptions not always plausible or reasonable.

The necessary R functions for the significance test are those of the (M)ANOVA tests such as anova or manova in the car package and linearHypothesis() [199, 194]. To test the significance of some (not all) regression coefficients the update function can be used to fit the reduced model and then the anova for the comparison of the full model with the reduced one. Alternatively, for the Hoteling T^2 test, the hoteling.test function is available in the MASS standard package [198]. The linearHypothesis is a special function to perform the four parametric tests for the MANOVA, such as Wilks' lambda, Pillai's trace test, Hotelling-Lawley trace, and Roy's greatest root test.

The Wilks lambda statistic $\wedge$ is given by: $\wedge = \frac{det(\hat{\Sigma})}{det(\tilde{\Sigma})}$, ratio between the sum of squares cross products of the residuals of the full model and the sum of squares cross products of the reduced model (under null hypothesis), and it approximates to a chi-square distribution χ^2 with Vp degrees of freedom[200].

$$-klog\wedge \approx \chi^2 \tag{4.7}$$

where $k = (n - (V + p + 1)/2)$, log is natural logarithm, V and p are number of response and explanatory variables respectively. Moreover, Let $\tilde{E} = n\hat{\Sigma}$ and $\tilde{H} = n(\tilde{\Sigma} - \hat{\Sigma})$ then the Pillai's trace test is $\sum \frac{\lambda_d}{1+\lambda_d}$, where λ_d is the nonzero eigenvalues of $\tilde{H}\tilde{E}^{-1}$.

4.3 Methodology of the Permutation MANOVA for the Multivariate Regression Model

4.3.1 Permutation Test

In the parametric approach, the probability that the test statistic takes values more extreme than the observed one under the null hypothesis is computed assuming a known and completely specified probability distribution. In other words, the p-value is computed according to the chi-square or Hotelling T^2 distribution. However, if the null hypothesis is false, the analysis is not valid because the distribution in H_1 is not equal to the assumed one. Hence, obtaining the quantiles is impossible unless we know the non-central parameters. In contrast, the permutation test is distribution-free and calculates the rejection probabilities or p-values under the null hypothesis regardless of the distribution of the error terms. By resampling (shuffling) the rows of the matrix of the explanatory variables many times, the p-values can be computed as the fraction of permuted datasets for which the value of the test statistic is greater than or equal to the observed one.

The permutation test is nonparametric; it does not rely on assumptions about the underlying distribution and requires exchangeability under H_o. In the null hypothesis that the whole set of explanatory variables does not affect the response, exchangeability is satisfied. In fact, in H_o, all the regression coefficients are null, and each vector of observed responses can relate to any vector of observed values of the predictors with the same probability.

In the previous studies concerning permutation tests for linear regression models, researchers use three different permutation strategies[95, 130, 131, 133, 66, 134]. For instance, by considering nuisance variables in the equation, testing the significance of the partial regression coefficients is made by using constrained permutation of residuals under the reduced model. On the other hand, by considering no nuisance explanatory variables in the study, some authors permute the observations of the dependent variable $\mathbf{Y}$. When the test concerns all the explanatory variables, no constrain is needed, and exchangeability permits the permutations of residuals obtained by the OLS estimates of parameters. Finally, some authors obtain the permutation distribution of the test statistics by permuting the rows of the matrix of explanatory variables [191, 192, 6, 97].

To calculate the null permutation distribution of the test statistic and its corresponding p-value, we should evaluate the possible permutation sample space and its cardinality $n!$ [5, 39, 4]. However, when the sample size n is large, the sample space's determination is complicated and time-consuming. For this reason, we apply the conditional Monte Carlo Procedure (MCP) by considering a random sample of $R \leq n!$ permutations. Then, according to the Glivenko-Cantelli theorem, the approximation of the permutation test obtained with the conditional Monte Carlo is very high, and such test can be considered almost equivalent to the exact test.

4.3.2 Combined Permutation Test

When a testing problem can be broken down into partial tests, the methodology on the combined permutation test is suitable and effective. It requires applying a univariate permutation test for each partial problem and obtaining a multivariate test statistic. To compute the p-value of the original global test, a suitable combining function may be used to reduce the multivariate test statistic into one (combined) univariate test statistic. The arguments of the combining function are the p-values of the partial tests.

4.3.3 The null hypothesis

Let us consider problems where the null hypothesis to be tested is that the components of the matrix of regression coefficients are equal to zero except for the constraints (the predictors do not affect the multivariate responses). The global alternative is that at least one predictor has a significant effect. Hence, the null hypothesis H_o can be broken down into null sub-hypotheses such that the global null hypothesis is true if all the partial hypotheses are jointly true. Likewise, the alternative hypothesis H_1 is broken down into partial alternatives such that H_1 is true if at least one partial alternative hypothesis is true. Formally:

$$H_{od(j)} : \{\beta_{jd} = 0, j = 1, \cdots, p, d = 1, \cdots, V\} \tag{4.8}$$

$$H_o : \{\bigcap_{d,j} H_{od(j)}\} \tag{4.9}$$

$$H_{1d(j)} : \{\beta_{jd} \neq 0\} \tag{4.10}$$

$$H_1 : \{\bigcup_{d,j} H_{1d(j)}\} \tag{4.11}$$

Under the null hypothesis, exchangeability holds, and by permuting observed $\mathbf{Y}$, it is possible to determine the null distribution of the test statistics. Thus, the permuted dataset is:

$$Z^* = (Y^*_{u^*(i)d}, X_{ij}), i = 1, \cdots, n, d = 1, \cdots, V, j = 1, \cdots, p\} \tag{4.12}$$

where

$$Y^*_{u^*(i)d} = XB + E_{u^*(i)d} \tag{4.13}$$

$u^*(i)$ is a permutation of units/labels $(i = 1, \cdots, n)$.

4.3.4 Permutation Test Statistic

Let us denote the corresponding (p.V)-dimensional vector of test statistics by $\mathbf{T} = \mathbf{T(Z)}$ and each component of the $T_{dj}(Z)$ be suitable to test the partial null hypothesis $H_{od(j)}$ against the partial alternatives $H_{1d(j)}$.

The partial permutation tests must be analyzed jointly. If all partial tests $T_{dj}(Z)$ are marginally unbiased and consistent, then the combined test is unbiased and consistent. Hence, they are stochastically larger in H_1 than H_o. As the sample sizes tend to infinity, the probability that the permutation p-values are less than α under H_1 tends to 1 [39].

We propose using the regression coefficients' estimators or other permutationally equivalent statistics (such as the t- statistic of the classic parametric t-test) as univariate test statistics of the partial test on a single coefficient. The OLS estimators represent a suitable choice under the assumption of homoscedasticity of errors and uncorrelated errors. Hence, $T_{dj}(Z) = \hat{\beta_{jd}}$ with $\hat{\beta_{jd}} = [(X'X)^{-1}X'Y]_{jd}$. The widely used parametric maximum likelihood estimators (MLE) require the more stringent assumption of multivariate normality for the joint distribution of the error terms.

The general procedure for performing such a combined test is the following:

1. Calculate the matrix of observed values of $\mathbf{T}$ from the dataset $\mathbf{Z}$ under the full model. $\mathbf{T} : \mathbf{T}_{obs} = T(\mathbf{Z})$ and combine using suitable combining function Ψ.

2. Consider a permutation of the rows of the matrix of response variables $\mathbf{Y}$.

3. Compute the corresponding values of the test statistics from the permuted dataset $\mathbf{T}^*_{(1)} = \mathbf{T}(Z^*_{(1)})$.

4. Repeat step$(2-3)$ R times independently to get $\mathbf{T}^*_{(r)} = \mathbf{T}(Z^*_{(r)}), r = 1, \cdots, R$, according to the conditional Monte Carlo method .

5. Consider the estimate of the significance level function for each partial test $\hat{L}_{jd}(t) = \frac{0.5+\sum_r I[T^*_{jd(r)} \geq t]}{R+1} = \hat{P}(T^*_{jd} \geq t)$, with $j = 1, \cdots, p, d = 1, \cdots, V$, where I(A) is the indicator function of A.

6. For each r compute $\hat{\lambda}^*_{jd(r)} = \hat{L}_{jd}(t_{jd(r)})$ and $\hat{\lambda}_{jd(0)} = \hat{L}_{jd}(t_{jd,obs})$, Where $t_{jd(r)}$ and $t_{jd,obs}$ are the value taken by T_{jd} in the r^{th} permuted dataset and in the observed one respectively. $\hat{\lambda}_{jd(0)}$ is the estimated p-value of the partial test corresponding to the d^{th} response and the j^{th}predictor according to the null permutation distribution of the multivariate test statistic.

7. For each r calculate the combined test statistic $T''^*_{(r)} = \Psi(\hat{\lambda}^*_{11(r)}, \hat{\lambda}^*_{21(r)}, \cdots, \hat{\lambda}^*_{pV(r)})$ and compute $T''_{(0)} = \Psi(\hat{\lambda}_{11(0)}, \hat{\lambda}_{21(0)}, \cdots, \hat{\lambda}_{pV(0)}$ using suitable combining function Ψ .

8. Calculate the estimate of the p-value of the global (combined) test $\hat{\lambda''} = \sum_r \frac{I[T''^*_{(r)} \geq T''^*_{(0)}]+0.5}{R+1}$

Provided that a suitable estimation method for single regression coefficients is available, the proposed method can be easily extended to nonlinear models, general linear models, mixed models, general mixed models, and vector generalized linear mixed models (VGLM).

4.3.5 Combination Functions

As said the univariate test statistic T'' for the global problem is obtained by combining the p-values of the partial tests on the regression coefficients $\lambda_{11}, \lambda_{21}, \cdots, \lambda_{pV}$ with a suitable function $\Psi : T'' = \Psi(\lambda_{11}, \lambda_{21}, \cdots, \lambda_{pV})$. Ψ is continuous and measurable for all p-values.

The most widely used combining functions are the following:

Tippett Combining function :

$$T''_T = max_{j,d}(1 - \lambda_{jd}) \tag{4.14}$$

Fisher omnibus combining function:

$$T''_F = -2 \sum_{j,d} ln(\lambda_{jd}) \tag{4.15}$$

Liptak combining function :

$$T''_L = \sum_{j,d} \Phi^{-1}(1 - \lambda_{jd}) \tag{4.16}$$

Where Φ is the cumulative distribution function of the standard normal random variable and ln is the natural logarithm. Without loss of generality, let us assume that all combining functions are significant for large values[5, 39]. The combining functions must satisfy the following condition to get valid results:

1. It must be non-increasing in each p-values: $\Psi(..., \lambda_{jd}, ...) \geq \Psi(..., \lambda'_{jd}, ...) \lambda_{jd} < \lambda'_{jd}$.
2. It attains its supremum value $\bar{\Psi}$, when at least one partial p-value attains zero: $min_{j,d}\lambda_{jd} \longrightarrow 0 \implies \Psi(\cdots, \lambda_{jd}, \cdots) \longrightarrow \bar{\Psi}$
3. $\forall \alpha > 0$, the critical value $T''_\alpha < \bar{\Psi}$.

4.3.6 Power of CPT

Let

$$\Pi(\mathbf{B}, \alpha, \mathbf{T}, P, n) = \frac{\sum (\lambda''(T_o) <= \alpha)}{R} \tag{4.17}$$

Where $\lambda''(T_o)$ is the probability that the test statistic takes values greater than or equal to observed T_o. In this study, we investigate the power of permutation test and compared it to a parametric competitor using simulation studies.

The R environment has a package called lmPerm useful to analyze the multivariate regression model using the permuted Y^* for tests and for estimates OLS through the lmp function[129]. However, for this study, we use the classic lm function, and we develop a specific R script for the simulation study and data analysis.

4.4 Simulation Study

In this simulation study, multivariate datasets were generated from multivariate normal distributions. In addition, specific R scripts were created to perform the simulation study.

The power behavior of the combined permutation test on the significance of regression coefficients (permutation MANOVA) is compared with that of the classic parametric competitor, the Pillai-Trace test. The simulation process is divided into three steps.

1. Definition of the setting parameters
2. Random generation of the predictors data, that is the $\mathbf{X}$ matrix of independent variables
3. Random generation of errors' values in order to compute the simulated data of the dependent variables. That is $\mathbf{Y}$ matrix of responses.

We consider p the case of two explanatory variables, hence p=2. To simulate the data the bivariate predictor is assumed to follow a normal distribution with null vector of means: $(X_1, X_2)' \sim \mathcal{N}_2(0, \Sigma_x)$. $\Sigma_x = [\sigma_{x,lj}]$ with $\sigma_{x,jj} = 1, j = 1, 2$. In order to consider both the cases of uncorrelated and correlated predictors, the considered values of the covariance parameters $\sigma_{x,12} = \sigma_{x,21}\rho_x$ are 0 and 0.3. The choice of considering only two cases of ρ_x, that is $\rho_x = 0$ and $\rho_x = 0.3$ is due to the belief that the level of multicollinearity doesn't affect the power of the permutation test. High multicollinearity is not an interesting case for the simulation study. Hence, we includes in the settings only the situations of uncorrelated orthogonal predictors and moderate collinearity (quite realistic in real applications with two explanatory variables). The errors are simulated from V-variate normal distributions with null vector of means and covariance matrix $\Sigma_\epsilon = [\sigma_{\epsilon,ds}]$. We set up $\sigma_{\epsilon,dd} = 1, \forall d \in \{1 \cdots V\}$, and $\sigma_{\epsilon,ds}$ constant with respect to the subjects d and s, with $d \neq s \in \{1, 2, \cdots, V\}$. To simulate null, weak and strong correlation between the V-components, three values are considered: $\sigma_{\epsilon,ds} = \rho_\epsilon = 0.0$, $\sigma_{\epsilon,ds} = \rho_\epsilon = 0.3$ and $\sigma_{\epsilon,ds} = \rho_\epsilon = 0.8$.

It is worth noting that under the null hypothesis, the V-variate errors (and the corresponding V-variate responses) are exchangeable with respect to the statistical units. The constants of the V model equations are set up at zero that is $\beta_{o,jd} = 0, j = 1, \cdots, p, d = 1, \cdots, V$. Hence, to compute the simulated values of the dependent variables, for simplicity of representation we can consider, as **B** matrix of regression coefficients for the $2x2$ matrix that includes only the model slopes. The simulation take into account the following **B** matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. The considered simulation sample sizes are $n = 10, 20, 30, 40, 70, 100$. Finally, for investigating the asymptotic power behavior when the number of model equations (response) diverges, keeping the sample size fixed, different values of V are considered. We considered 1000 simulated datasets and 1000 permutations.

4.4.1 Simulation Results and Discussion

When $B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, that is under H_o, as shown in Table 4.1-Table 4.4 and figure 4.1d, combined permutation tests have rejection rates close to the significance level $\alpha = 0.05$. There are some exceptions to this general rule, but this is true also for the parametric Pillai-Trace test, and it seems not correlated to the sample size. Whereas the Pillai-Trace test seems to be conservative for small sample sizes. Hence, under H_o, the power of the combined permutation tests tends to respect the nominal α level, and we can say that these tests are well approximated.

As shown in Fig. 4.1a-c, Table (4.5, 4.6, 4.7, 4.8) and Table A3 (in the appendix) the power of the combined permutation tests under H_1, that is when $\mathbf{B} \neq 0$ is greater than the power under H_o. This proves that the tests are unbiased, and the parametric Pillai test. This is true regardless of the correlation of the model equations. Fig.4.3 and Table 4.8 show that this is true regardless of collinearity among predictors. In general, it is evident that when the alternative hypothesis is true, the power increases as a function of the sample sizes, and it tends to one as n diverges. Hence, the tests under study are consistent.

Table 4.1: The Power of the tests for $\rho_\epsilon = 0.3$, $\rho_x = 0$, $\alpha = 0.05$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.05	0.05	0.049	0.057	0.049	0.055
CPT Tippett	0.056	0.052	0.047	0.051	0.047	0.057
Parmetric Pillai	0.024	0.041	0.05	0.048	0.05	0.06

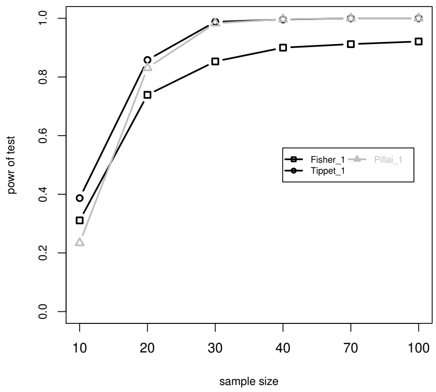
Table 4.2: The Power of the tests for $\rho_\epsilon = 0.8$, $\rho_x = 0$, $\alpha = 0.05$ and B=$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.041	0.041	0.052	0.051	0.053	0.04
CPT Tippett	0.042	0.052	0.052	0.051	0.051	0.054
Parametric Pillai	0.027	0.041	0.042	0.058	0.062	0.049

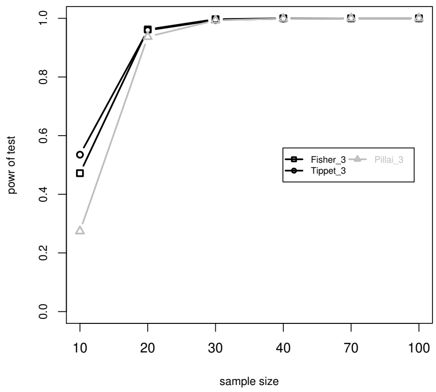
Table 4.3: The Power of the tests for $\rho_x = 0.0$, $\rho_\epsilon = 0$, $\alpha = 0.05$ and B=$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.051	0.05	0.045	0.044	0.049	0.039
CPT Tippett	0.048	0.048	0.044	0.042	0.045	0.044
Parametric Pillai	0.023	0.039	0.036	0.04	0.044	0.04

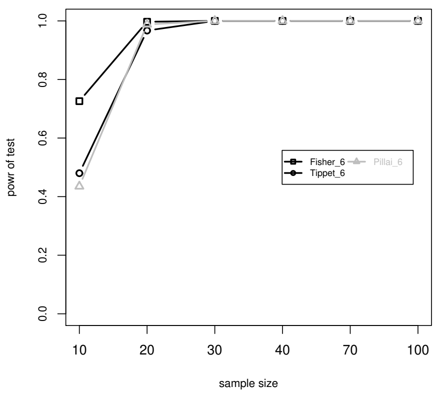
For small sample sizes, the power of the combined permutation test using Tippett for $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is more powerful than the competitor multivariate parametric test such as Pillai's-Trace test regardless of the correlation setup. In other words, the CPT based on the Tippett combination is the most powerful of the compared tests when only one coefficient is not null see (Fig.4.1a, 4.3a, 4.3c). When the number of coefficients not equal to zero is two, the



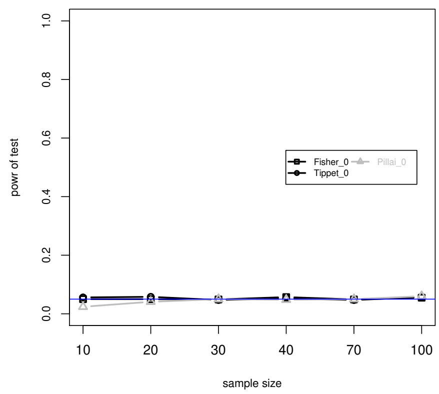
(a) B= $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$



(b) B= $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$



(c) B= $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$



(d) B= $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Figure 4.1: The power behaviour for different value of B,n, $\rho_x = 0.0$, $\rho_\epsilon = 0.3$

Table 4.4: The Power of the tests for $\rho_x = 0.3$, $\rho_\epsilon = 0$, $\alpha = 0.05$ and B= $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.058	0.059	0.05	0.054	0.052	0.039
CPT Tippett	0.051	0.052	0.052	0.049	0.052	0.039
Parametric Pillai	0.028	0.049	0.048	0.046	0.05	0.035

most powerful test is the CPT based on Fisher combination regardless of the values of the correlations mainly when the sample size is n=10 see (Fig. 4.1b, 4.2a, 4.3b, Table 4.5, Table 4.6 and TableA.6 (in the appendix)). For small sample sizes, the superiority of CPT based on Fisher's is much more evident when $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, even if when $n \geq 20$ the performance of the Pillai- Trace test is similar see Fig. 4.1c, 4.2b and Table 4.7. However, in some cases, such as when the sample size is large, the parametric Pillai-Trace test is somewhat performant. For instance, when $B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, without loss of generality, the Pillai multivariate test seems performant, as shown in Table 4.8.

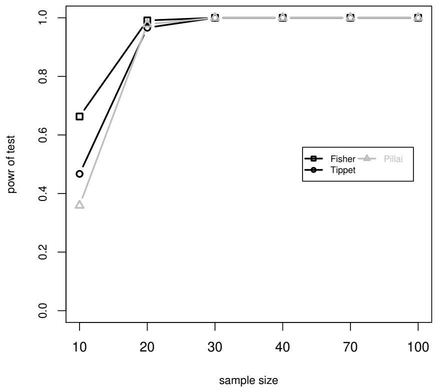
Table 4.5: The Power of the tests for $\rho_\epsilon = 0.3$, $\rho_x = 0$, $\alpha = 0.05$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.4	0.941	0.998	1	1	1
CPT Tippett	0.269	0.877	0.997	1	1	1
Parametric Pillai	0.349	0.98	1	1	1	1

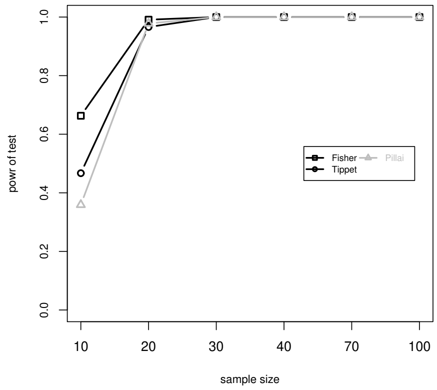
Table 4.6: The Power of test for $\rho_x = 0.0$, $\rho_\epsilon = 0$, $\alpha = 0.05$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.403	0.931	0.999	1	1	1
CPT Tippett	0.328	0.873	0.995	1	1	1
Parametric Pillai	0.304	0.965	1	1	1	1

As expected the power of the tests is non increasing function of the correlation between equations ρ_ϵ. For instance, this is evident in Table 4.9,



(a) B=$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$



(b) B=$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

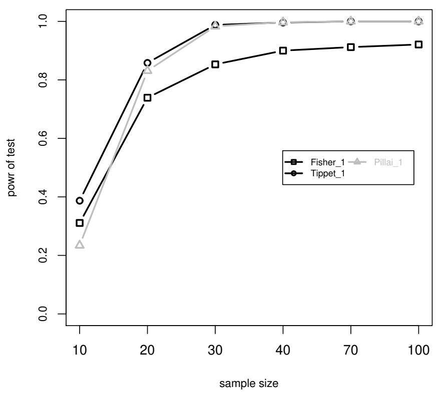
Figure 4.2: The power behaviour for different value of B, n, $\rho_x = 0.0$, $\rho_\epsilon = 0.8$

Table 4.7: The Power of the tests for $\rho_x = 0.0$, $\rho_\epsilon = 0$, $\alpha = 0.05$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

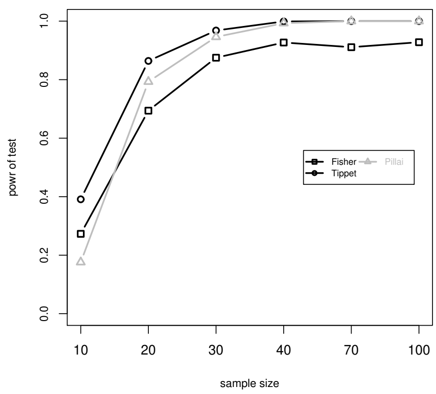
Sample size	10	20	30	40	70	100
CPT Fisher	0.748	1	1	1	1	1
CPT Tippett	0.504	0.979	1	1	1	1
Parametric Pillai	0.538	1	1	1	1	1

where the rejection rates for n=30, $\rho_x = 0.0$ and B=$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ are reported. If the equations are uncorrelated the global information provided by both of them is higher than in the case of non-zero correlation. This is reflected in the dependence of partial tests and therefore on the power of the regression MANOVA. The higher the correlation, the greater the redundant information provided by partial tests of different equations which results in fewer additional gain in power contribution for each equation in the global test.

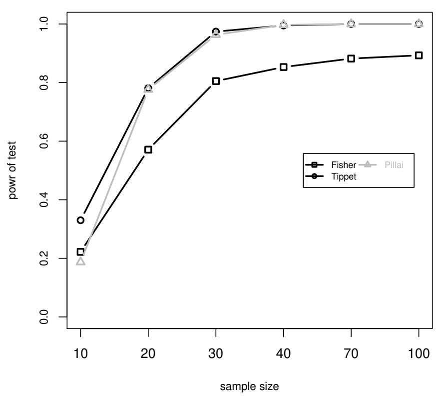
The second part of the simulation is about the small sample problem with a divergent number of dependent variables, namely equations, typical of some big data problems of genetics, marketing, psychology, and other fields. Data are simulated for a given sample size $n = 10$, correlation among responses



(a) B= $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ $\rho_x = 0.0,\ \rho_\epsilon = 0$



(b) B= $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $\rho_x = 0.0,\ \rho_\epsilon = 0$



(c) B= $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\rho_x = 0.3,\ \rho_\epsilon = 0$

Figure 4.3: The power behaviour for different value of B, n

Table 4.8: The power of the tests for different values of ρ_ϵ, ρ_x, n and B

ρ	B	Tests	Sample size(n)					
			10	20	30	40	70	100
	B3	Tippett	0.269	0.877	0.997	1	1	1
$\rho_\epsilon = 0.3$		Fisher	0.4	0.941	0.998	1	1	1
$\rho_x = 0$		Pillai	0.349	0.98	1	1	1	1
	B5	Tippett	0.511	0.962	1	1	1	1
		Fisher	0.652	0.986	1	1	1	1
		Pillai	0.71	0.997	1	1	1	1
	B2	Tippett	0.385	0.838	0.978	0.998	1	1
		Fisher	0.256	0.675	0.682	0.699	0.708	0.9
		Pillai	0.393	0.985	1	1	1	1
$\rho_\epsilon = 0.8$	B3	Tippett	0.32	0.855	0.994	1	1	1
$\rho_x = 0$		Fisher	0.363	0.927	0.995	1	1	1
		Pillai	0.349	0.98	1	1	1	1
	B5	Tippett	0.535	0.968	1	1	1	1
		Fisher	0.613	0.992	1	1	1	1
		Pillai	0.885	1		1	1	1
	B3	Tippett	0.328	0.873	0.995	1	1	1
		Fisher	0.403	0.931	0.999	1	1	1
		Pillai	0.304	0.965	1	1	1	1
	B5	Tippett	0.542	0.973	1	1	1	1
		Fisher	0.7	0.991	1	1	1	1
$\rho_x = 0$		Pillai	0.672	0.994	1	1	1	1
$\rho_\epsilon = 0$	B6	Tippett	0.504	0.979	1	1	1	1
		Fisher	0.538	1	1	1	1	1
		Pillai	0.43	0.98	1	1	1	1
	B3	Tippett	0.191	0.662	0.957	0.997	1	1
		Fisher	0.237	0.759	0.969	0.999	1	1
		Pillai	0.376	0.991	0.999	1	1	1
	B4	Tippett	0.494	0.93	0.995	1	1	1
$\rho_x = 0.3$		Fisher	0.448	0.923	0.997	1	1	1
$\rho\epsilon = 0$		Pillai	0.326	0.973	1	1	1	1
	B5	Tippett	0.419	0.92	0.998	1	1	1
		Fisher	0.495	0.996	0.999	1	1	1
		Pillai	0.695	0.997	1	1	1	1
	B6	Tippett	0.275	0.843	0.989	1	1	1
		Fisher	0.45	0.95	0.999	1	1	1
		Pillai	0.611	0.999	1	1	1	1

$B2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B3 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $B4 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $B5 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $B6 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

Table 4.9: The power of the compared tests as a function of ρ_ϵ with $\rho_x = 0$

ρ_ϵ	0	0.3	0.8
CPT Fisher	0.999	0.997	0.977
CPT Tippett	0.998	0.994	0.987
Parametric Pillai	1	0.993	0.966

($\rho_\epsilon = 0.3$), fixed number of explanatory variables ($p = 2$), multicollinearity ($\rho_X = 0.3$), and all elements of the matrix B, that is all the regression coefficients, non-zero. When $p * V > n$, the Pillai-Trace multivariate test cannot be applied see Table 4.10. The loss of degrees of freedom due to the increased number of dependent variables implies that the parametric approach is not applicable for big data problems with small sample sizes. Moreover, the power of the Pillai-Trace test decreases as the number of dependent variables increases. On the other hand, the power of the proposed nonparametric multivariate tests based on the permutation approach increases as the number of components of the multivariate response increases.

Table 4.10: The power of test for divergent number of responses, $\rho_\epsilon = \rho_x = 0.3$, B=$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$

V	2	4	8	12
CPT Fisher	0.434	0.477	0.497	0.519
CPT Tippett	0.284	0.3	0.312	0.326
Prametric Pillai	0.783	0.258	NA	NA

The power of the combined permutation test is increasing function of the number of non-zero coefficients as shown in Table 4.11, where the rejection rates of the two proposed permutation solutions are reported as a function of the **B** matrix, for n=20, $\rho_\epsilon = \rho_x = 0.3$. Tippett combination is preferable when only one coefficient is not equal to zero. When the number of non-zero coefficients is greater than two, the Fisher combination is preferable.

Table 4.11: The power of test as a function of non-zero coefficients, $\rho_\epsilon = \rho_x = 0.3$, n=20

B	$\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 0\\1 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$
CPT Fisher	0.05	0.739	0.962	0.963	0.986	0.997
CPT Tippett	0.052	0.858	0.959	0.961	0.962	0.967

The third part of the simulation is about a large number of response variables $V = 3$, large explanatory $p = 14$ (some continuous and others binary), large sample size $n = 848$. We perform this simulation to examine whether the proposed combined permutation test is suitable for the firm performance dataset or not. Hence, we consider the correlation matrix of firm performance as to measure dependency of equations (see Table A.8 in the appendix). As we see in Table A.7 in the appendix, the power of both parametric and nonparametric methods converges to one. Due to the Central limit theorem, for the large enough sample sizes, the parametric method provides high power performance of the test. Similarly, the CPT provides a performant power for a large sample size due to consistent property regardless of the dimension of variables. Hence, CPT is robust for divergent variables with small sample sizes and a divergent number of variables with large sample sizes. Moreover, the proposed combined permutation tests are a suitable method for analyzing firm performance datasets with 3 dimensional responses and 11 explanatory variables for 848 private firms, and we use a firm dataset for the case study.

4.5 CPT for Private Firm Performance Analysis

Firms' financial and business performance has an increasingly significant role in promoting economic growth and sustainable development around the globe. Similarly, the firm performance has a remarkable contribution to economic growth and sustainable economic development of fast-growing countries like Ethiopia by eradicating poverty, creating job opportunities, enhancing the GDP value, and enhancing the welfare of the society.

Firm performance could be considered the firm's effectiveness, success, and efficiency. Firm performance such as growth in sales, profitability, labor productivity, employment growth, and customer satisfaction are significant indicators to determine economic growth and sustainable development[201]. For instance, firstly, the higher profitability for the investor, the better the return of the investment, which in turn incentive investors to invest more capital, and the demand for labor increases as well. Again the increase in demand for labor increases the employees' wage, reducing income inequality to some extent. Secondly, it has a vital role in promoting economic growth and achieving sustainable development. The greater the firm's performance, the higher the attitude towards reinvestment the need for capital and skillful labor. The combination of the labor force, saving, and investment increases production and promotes economic growth[202, 203, 204].

Efficient and specialized workers can increase the production of goods and improve the quality of products and services, also introducing (innovation and improved technology)[205]. Efficient labors promote productivity and, in turn, contributes to increased profitability.

The growth of performance is one of the goals of the growth and transformation plan (GTP I &II) of Ethiopia[206]. The reformed government of Ethiopia emphasizes the importance of private enterprises. Therefore, Ethiopia is promoting investments infrastructures, reducing crime and corruption, tax regulation, fostering innovation, improving the contribution of industrial sectors to the GDP, and exporting more products to bring the rapid growth[207].

Many researchers use the financial perspective to measure firms' performance[208, 209]. They considered profitability in terms of financial ratios such as return on investment, return on equity, and asset turnover; growth in terms of increase in total sales and revenues; liquidity in terms of cash flow; solvency in terms of leverage; and risk in terms of financial leverage.

Although there are many financial indicators and other nonfinancial measures that represent firms' performance [181], other researchers in the literature use business perspective to measure the firm performance in terms of productivity, flexibility, and adaptability[210, 181, 211, 201].

On the other hand, in the literature, there are many empirical and theoretical contributions about the impact of sustainability indicators on firm performance[212, 213, 214, 215, 216]. For instance, [214] studied the association between environmental sustainability and multidimensional measures of firm performance including financial, and non-financial aspects transformed using factor analysis in India. In addition, [212] investigated the relationship between innovation, firm performance, and the three pillars of sustainability such as social, environmental, and economic sustainability. Moreover, many obstacles hinder the private firms' performance. Enterprise has been vulnerable to many factors to achieve its performance [207, 217] such as crime, corruption, weak access to finance, outage of power supply, and others.

However, there is a lack of empirical analysis about private firms' performance in Ethiopia and their association with sustainability indicators to the best of our knowledge. Thus, we conduct a multivariate regression analysis with multidimensional variables representing private firms' performance. This high number of regression coefficients suggests the advisability of adopting the combined permutation test for the regression MANOVA of the model.

4.5.1 Data and Method

The dataset for this study is obtained from the World Bank 2015 Ethiopian enterprise survey[218]. The dataset has 848 private firms, and it does not include firms from the agricultural sector and public firms.

The components of the multidimensional response are profitability, growth, and labor productivity represented by the indicators real annual total sales growth (GS), employment growth (GE), and labor productivity growth (GLP), respectively, expressed as percentages. For details about the calculations of those dimensions, we refer the reader to the World Bank 2015 report [218]. Hence, $GS = (1/t).(\frac{S-S'}{(S+S')/2}).100$, where S is the total annual sales for all products and services, t is the number of years, and S' is the total annual sales for all products and services t years ago. $GE = (1/t).(\frac{L-L'}{(L+L')/2}).100$, where L and L' are permanent full-time workers of the current and previous period respectively. Moreover, $GLP = (1/t).(\frac{(S/L)-(S'/L')}{((S/L)+(S'/L'))/2}).100$.

The selected business environment or sustainability indicators as explanatory variables in this study are access to finance, corruption, crime, infrastructure (electricity outage), ICT, unskilled labor force, the informality of firms, tax rates, and regulation, innovation, firm size, trade, percentage of firms ownership, and female owners.

The ownership status of the firms has a crucial role in determining the firms' performance. That is, the higher percentage of ownership by private firms, the higher firms' expected performance, and inspiring the government towards a privatization strategy.

Corruption is the abuse of managers or officials of public administration for private gain, and it is supposed to reduce firms' performance[219]. For example, in dead in the corrupted system, private firms are asked to pay illegally for import licenses, water connection, electricity connection, and construction permits, and these practices might affect firm performance.

Innovation is one of the distinguishing features of the world's economy in developed and developing countries. According to Schumpeter, innovation is the means to allocate resources effectively, and it boosts production growth[220]. Similarly, when firms are capable of using the advanced or latest technology and innovation, they could improve their production process and increase productivity[221].

In developing countries, access to finance has a significant role in determining the private firm performance[222]. Getting loans or credit allows firms to employ more workers and make more investments.

Infrastructure is the cornerstone to facilitate private and public firms' economic growth and development. For instance, without sufficient electric power and ICT, the sustainable development goals and the private firm's performance are distant goals. In Ethiopia, power outages are a typical situation, and they might hinder firms' performance[207]. On the other hand, the use of ICT and the internet helps private firms to reduce the communication costs[223, 224].

Firm size is defined as a categorical variable, and the categories are micro (< 5 employees), small firms ($5 - 19$ employees), medium firms ($20 - 99$ employees), and large firms (100 or more employees). The size of firms has a significant role in the performance. There are some arguments about the effect of the size of firms on private firm performance. Some of the researchers argue that the larger firm size, the higher the firm performance because they find those big companies have better performance than small companies [225, 226].

4.5.2 Results of the Case Study

The global p-value of the test on the hypothesis that the explanatory variables significantly affect the multidimensional firms' performance measures, using CPT with Tippett combining function, is 0.001. We use Tippett combining function since we suspect that only a few predictors are significantly associated with firms' performance. Thus, the global p-value is less than

$\alpha = 0.05$, and we reject the null hypothesis.

To determine which explanatory variables significantly contribute to private firm performance, we conduct post hoc analysis by controlling the family-wise error rate (FWER) and displaying the adjusted p- values in Table 4.12. Hence, small firm size, percentage of ownership, finance, power outage (hours), and ICT significantly contribute to the firm performance at a 0.05 level of significance. In particular, small firm size and power outage (hours) are related to real annual labor productivity growth. In addition, the percentage of ownership and ICT positively affects the real annual employment growth. Moreover, finance significantly contributes to growth in total sales.

To sum up, the contributions of firms' private ownership status on percentage annual employment growth in Ethiopia has a crucial role in supporting the privatization strategy of firms by policymakers. This result provides evidence in favor of the privatization of firms in Ethiopia. Furthermore, the firms' ownership is crucial for job creation and reducing the unemployment rate since it contributes to the annual growth in employment. Hence, the firms' ownership status has important policy implications in determining the Ethiopian transformation into a middle-income country.

Access to loans or credit from the financial system contributes to annual growth in sales of private firms in Ethiopia. Hence, it improves the total output of the private firms produced in the fiscal year. Therefore, a feasible financial system could be vital for facilitating the economic growth of Ethiopia by providing loans or credit to private firms. This result is consistent with the study about the role of finance in promoting firms' growth by [217].

Some of the researchers argue that the firm size is increasing function of performance [225, 226]. In contrast, we obtain that small firms significantly contribute to firms' performance. Moreover, the electric power outages negatively affect the annual sales growth of private firms in Ethiopia. Therefore, to tackle this problem, the Ethiopian government must finish the Great Ethiopian Renaissance Dam (GERD) and start generating hydroelectric power as soon as possible to promote rapid economic development. Finally, ICT is positively significant with the growth in employment of private firms in Ethiopia.

Table 4.12: The significance test of multivariate regression coefficients

Variables	GS(%)	GE(%)	GLP(%)
Medium firm size	0.955(0.988)	0.755(0.988)	0.847(0.988)
Micro firm size	0.791(0.991)	0.826(0.991)	0.806(0.991)
Small firm size	0.049(0.089)	0.089(0.846)	0.007(**0.016***)
Firms ownership(%)	0.524(0.756)	0.001(**0.0001***)	0.630(0.756)
Females owner	0.342(0.591)	0.272(0.591)	0.830(0.830)
Regulation and tax	0.587(0.801)	0.919(0.919)	0.466(0.779)
Corruption	0.146 (0.266)	0.299(0.299)	0.069(0.178)
Firms informality	0.936(0.936)	0.187(0.431)	0.180(0.431)
Finance	0.001(**0.0001***)	0.498(0.498)	0.058(0.111)
Power outage(hours)	0.579(0.807)	0.846(0.846)	0.014(**0.037***)
ICT	0.507(0.743)	0.013(**0.034***)	0.765(0.765)
Innovation	0.774(0.799)	0.434(0.779)	0.511(0.799)
Trade	0.600(0.844)	0.576(0.844)	0.531(0.844)
Unskilled labour	0.612(0.751)	0.052(0.116)	0.691(0.751)

Note:the table contains raw p-value and adjusted p-value in side the parenthesis,the bold value with star is significant at $1\% or 5\%$

4.6 Conclusions

The chapter aims to propose a nonparametric solution within the family of combined permutation tests for testing the joint significance of the multivariate regression coefficients and investigate its power behavior through a comparative simulation study. The proposed test is exact, unbiased, and consistent regardless of sample sizes, dependency among responses, and multicollinearity. Moreover, the type I error rate of the main parametric competitor for small sample sizes is distant from α (conservative).

In conclusion, the power behavior of the proposed solution is better than the Pillai multivariate parametric test even under multivariate normality and homoscedastic error terms. Furthermore, the simulation results reveal that the power of the Pillai-Trace test decreases much faster than the proposed permutation solution in the presence of dependency among response variables.

In general, the power of the parametric Pillai's-Trace test decreases as the number of non-zero coefficients (i.e., the number of responses and relevant predictors) increases and cannot be applied when the number of coefficients is larger than the sample sizes. In contrast, the power of the proposed permutation test increases as the number of dependent variables increases with fixed sample size.

In addition, as the number of non-zero coefficients (the percentage of true alternative hypotheses) is large, the Fisher combination is more powerful than Tippett and vice versa.

Finally, the case study shows that small firm size, percentage of ownership, finance, power outage (hours), and ICT significantly affect the firms' performance at 5% level of significance. Therefore, a feasible financial system could be vital for facilitating the economic growth of Ethiopia by providing loans or credit to private firms. Therefore, we suggest that the Ethiopian government should finish the Great Ethiopian Renaissance Dam (GERD) and start generating hydroelectric power as soon as possible to promote rapid economic development.

CHAPTER 5

GENERAL CONCLUSION

Often, in empirical studies, the features of datasets have a complexity that makes them difficult to manage and analyze, such as multivariate responses, small sample sizes, unknown dependence structure of variables, non-normal distributions, and big data characteristics. Therefore, the parametric approach may not be suitable, flexible, and robust. In general, the parametric methods require stringent assumptions (often unrealistic) such that the inferential results are not reliable. Hence, a nonparametric solution is often preferable if not necessary. The thesis focuses on complex testing problems, typical of empirical economic, social, or environmental sustainability studies. After a review of permutation methods, for tests of hypothesis, we take into account two common but complex problems such as multivariate tests for comparing groups in the presence of numeric variables and tests on the significance of the regression coefficients (jointly considered) in a multivariate linear regression analysis (regression MANOVA). Finally, through Monte Carlo simulation, we propose solutions within the family of combined permutation tests and deepen the investigation on their properties, especially in particular conditions such as small sample sizes and a large number of responses.

5.1 Final Remarks

According to the simulation studies concerning the location problem for numeric variables with independent samples, CPT is unbiased, consistent, and powerful for small sample sizes. The test is not only consistent in the classic sense; when the number of responses diverges under the alternative hypothesis, keeping fixed the sample sizes, the power tends to one. We prove that the most vital parameter determining the convergence to one of the power is the percentage of true partial alternative hypotheses rather than the absolute number of the true partial alternative hypothesis.

The comparative study of the different CPT provides evidence that the test based on the Tippett combination is much more powerful until a given percentage of the true partial alternative hypothesis. After this threshold, that depends on some factors such as sample sizes, number of samples, and others. The Fisher combination is preferable when this percentage is greater than this threshold (even if the Tippett combination remains performant). The Liptak combination is the most powerful only when 100% of the partial alternative hypothesis is true, but the rejection rates of all the tests are similar in this situation. This property makes CPT an essential solution for tests of hypotheses for big data, with a large number of responses and (but not only) small sample sizes.

For the regression MANOVA, the proposed permutation solution is robust and suitable for many predictors and (or) responses, competitive to the parametric Pillai test, and valid with the parametric approach is not applicable (sample size less than a number of variables). An essential property of the proposed solution, based on the combination of the partial permutation tests on the single coefficients, is that it appropriately considers the regression MANOVA as a multivariate test with possibility, in case significance of the global test on the whole set of regression coefficients, of attributing the global significance to some specific regression coefficients. This can be done by considering adjusted partial p-values to control the family-wise error rate and avoid the inflation of the type I error rate of the global test. Therefore, linking the tests on the significance of the single regression coefficients to the general MANOVA problem is more appropriate than considering the single tests separately, as usually done in practice.

BIBLIOGRAPHY

[1] J. Hemerik, M. Thoresen, and L. Finos, "Permutation testing in high-dimensional linear models: an empirical investigation," *Journal of Statistical Computation and Simulation*, vol. 91, no. 5, pp. 897–914, 2021.

[2] J. Frossard and O. Renaud, "Permutation tests for regression, anova and comparison of signals: The permuco package," 2018.

[3] P. Ding, A. Feller, and L. Miratrix, "Randomization inference for treatment effect variation," *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pp. 655–671, 2016.

[4] S. Bonnini, L. Corain, M. Marozzi, and L. Salmaso, *Nonparametric hypothesis testing: rank and permutation methods with applications in R.* John Wiley & Sons, 2014.

[5] F. Pesarin, *Multivariate permutation tests: with applications in biostatistics.* Wiley Chichester, 2001, vol. 240.

[6] M. J. Anderson and J. Robinson, "Permutation tests for linear models," *Australian & New Zealand Journal of Statistics*, vol. 43, no. 1, pp. 75–88, 2001.

[7] R. A. Fisher, "Design of experiments," *Br Med J*, vol. 1, no. 3923, pp. 554–554, 1936.

[8] G. Flandin and K. J. Friston, "Analysis of family-wise error rates in statistical parametric mapping using random field theory," *Human brain mapping*, vol. 40, no. 7, pp. 2052–2054, 2019.

[9] J. P. Romano and E. Lehmann, "Testing statistical hypotheses," 2005.

[10] P. H. Westfall and S. S. Young, *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons, 1993, vol. 279.

[11] E. J. Pitman, "Significance tests which may be applied to samples from any populations," *Supplement to the Journal of the Royal Statistical Society*, vol. 4, no. 1, pp. 119–130, 1937.

[12] M. Anderson and C. T. Braak, "Permutation tests for multi-factorial analysis of variance," *Journal of statistical computation and simulation*, vol. 73, no. 2, pp. 85–113, 2003.

[13] W. F. Christensen and B. N. Zabriskie, "When your permutation test is doomed to fail," *The American Statistician*, pp. 1–11, 2021.

[14] B. M. de Andrade, J. S. de Gois, V. L. Xavier, and A. S. Luna, "Comparison of the performance of multiclass classifiers in chemical data: Addressing the problem of overfitting with the permutation test," *Chemometrics and Intelligent Laboratory Systems*, vol. 201, p. 104013, 2020.

[15] C. Cai and Y. Wang, "Understanding the power of persistence pairing via permutation test," *arXiv preprint arXiv:2001.06058*, 2020.

[16] T. B. Berrett, Y. Wang, R. F. Barber, and R. J. Samworth, "The conditional permutation test for independence while controlling for confounders," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 82, no. 1, pp. 175–197, 2020.

[17] C. Chang, T. Jaki, M. S. Sadiq, A. A. Kuhlemeier, D. Feaster, N. Cole, A. Lamont, D. Oberski, Y. Desai, and M. L. Van Horn, "A permutation test for assessing the presence of individual differences in treatment effects," *arXiv preprint arXiv:1911.07248*, 2019.

[18] K. Rao, R. Drikvandi, and B. Saville, "Permutation and bayesian tests for testing random effects in linear mixed-effects models," *Statistics in medicine*, vol. 38, no. 25, pp. 5034–5047, 2019.

[19] T. B. Berrett, Y. Wang, R. F. Barber, and R. J. Samworth, "The conditional permutation test for independence while controlling for confounders," *arXiv preprint arXiv:1807.05405*, 2018.

[20] M. H. Samuh and F. Pesarin, "Applications of conditional power function of two-sample permutation test," *Computational Statistics*, vol. 33, no. 4, pp. 1847–1862, 2018.

[21] C. J. DiCiccio and J. P. Romano, "Robust permutation tests for correlation and regression coefficients," *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 1211–1220, 2017.

[22] J. Gagnon-Bartsch and Y. Shem-Tov, "The classification permutation test: A nonparametric test for equality of multivariate distributions," *arXiv preprint arXiv:1611.06408*, 2016.

[23] B. T. Files, V. J. Lawhern, A. J. Ries, and A. R. Marathe, "A permutation test for unbalanced paired comparisons of global field power," *Brain topography*, vol. 29, no. 3, pp. 345–357, 2016.

[24] P. R. Filgueiras, J. C. L. Alves, C. M. Sad, E. V. Castro, J. C. Dias, and R. J. Poppi, "Evaluation of trends in residuals of multivariate calibration models by permutation test," *Chemometrics and Intelligent Laboratory Systems*, vol. 133, pp. 33–41, 2014.

[25] J.-D. Gómez-Mantilla, V. G. Casabó, U. F. Schaefer, and C.-M. Lehr, "Permutation test (pt) and tolerated difference test (tdt): Two new, robust and powerful nonparametric tests for statistical comparison of dissolution profiles," *International journal of pharmaceutics*, vol. 441, no. 1-2, pp. 458–467, 2013.

[26] U. Hahn, "A studentized permutation test for the comparison of spatial point patterns," *Journal of the American Statistical Association*, vol. 107, no. 498, pp. 754–764, 2012.

[27] H. Nguyen, I. Wood, and M. Hill, "A robust permutation test for quantitative silac proteomics experiments," *Journal of Integrated OMICS*, vol. 2, no. 2, pp. 80–93, 2012.

[28] D. Basso and L. Salmaso, "A permutation test for umbrella alternatives," *Statistics and Computing*, vol. 21, no. 1, pp. 45–54, 2011.

[29] B. Phipson and G. K. Smyth, "Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn," *Statistical applications in genetics and molecular biology*, vol. 9, no. 1, 2010.

[30] E. McIntee, E. Viglino, S. Kumor, C. Rinke, L. Ni, and M. E. Sigman, "Non-parametric permutation test for the discrimination of float glass samples based on libs spectra," *Journal of Chemometrics*, vol. 24, no. 6, pp. 312–319, 2010.

[31] D. Manfredini, S. Bonnini, R. Arboretti, and L. Guarda-Nardini, "Temporomandibular joint osteoarthritis: an open label trial of 76 patients treated with arthrocentesis plus hyaluronic acid injections," *International journal of oral and maxillofacial surgery*, vol. 38, no. 8, pp. 827–834, 2009.

[32] D. Basso and L. Salmaso, "A discussion of permutation tests conditional to observed responses in unreplicated 2 m full factorial designs," *Communications in Statistics—Theory and Methods*, vol. 35, no. 1, pp. 83–97, 2006.

[33] M. Marozzi and L. Salmaso, "Multivariate bi-aspect testing for the two-sample location problem," *Communications in Statistics—Theory and Methods*, vol. 35, no. 3, pp. 477–488, 2006.

[34] S. Bonnini, L. Corain, F. Munaò, and L. Salmaso, "Neurocognitive effects in welders exposed to aluminium: an application of the npc test and npc ranking methods," *Statistical Methods and Applications*, vol. 15, no. 2, pp. 191–208, 2006.

[35] S. Bonnini, L. Salmaso, A. Solari *et al.*, "Multivariate permutation tests for evaluating effectiveness of universities through the analysis of student dropouts," *Statistica & Applicazioni*, vol. 3, pp. 37–44, 2005.

[36] L. Finos and L. Salmaso, "Nonparametric multi-focus analysis for categorical variables," *Communications in Statistics-Theory and Methods*, vol. 33, no. 8, pp. 1931–1941, 2004.

[37] L. Corain and L. Salmaso, "An empirical study on new product development process by nonparametric combination (npc) testing methodology and post-stratification," *Statistica*, vol. 63, no. 2, pp. 335–357, 2003.

[38] H. Scheffé, "The analysis of variance john wiley & sons," *Inc., New York*, pp. 25–50, 1959.

[39] F. Pesarin and L. Salmaso, *Permutation tests for complex data: theory, applications and software.* John Wiley & Sons, 2010.

[40] A. Janssen and T. Pauls, "A monte carlo comparison of studentized bootstrap and permutation tests for heteroscedastic two-sample problems," *Computational Statistics*, vol. 20, no. 3, pp. 369–383, 2005.

[41] R. Wang and W. Xu, "An approximate randomization test for high-dimensional two-sample behrens-fisher problem under arbitrary covariances," *arXiv preprint arXiv:2108.01860*, 2021.

[42] Ł. Smaga, "Inference for general manova based on anova-type statistic," in *Advanced Studies in Classification and Data Science.* Springer, 2020, pp. 241–254.

[43] M. Pauly, E. Brunner, and F. Konietschke, "Asymptotic permutation tests in general factorial designs," *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pp. 461–473, 2015.

[44] R. C. Francis and B. F. Manly, "Bootstrap calibration to improve the reliability of tests to compare sample means and variances," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 12, no. 8, pp. 713–729, 2001.

[45] P. W. Mielke Jr and K. J. Berry, "Permutation tests for common locations among samples with unequal variances," *Journal of Educational Statistics*, vol. 19, no. 3, pp. 217–236, 1994.

[46] B. Wheeler, M. Torchiano, and M. M. Torchiano, "Package 'lmperm'," *R package version*, vol. 2, no. 0, 2016.

[47] A. Zeileis, M. A. Wiel, K. Hornik, and T. Hothorn, "Implementing a class of permutation tests: the coin package," *Journal of statistical software*, vol. 28, no. 8, pp. 1–23, 2008.

[48] L. Finos, M. L. Finos, and I. Rcpp, "Package 'flip'," 2011.

[49] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric statistical methods.* John Wiley & Sons, 2013, vol. 751.

[50] P. R. Peres-Neto and J. D. Olden, "Assessing the robustness of randomization tests: examples from behavioural studies," *Animal Behaviour*, vol. 61, no. 1, pp. 79–86, 2001.

[51] T. W. O'Gorman, "An adaptive permutation test procedure for several common tests of significance," *Computational statistics & data analysis*, vol. 35, no. 3, pp. 335–350, 2001.

[52] P. I. Good, "Extensions of the concept of exchangeability and their applications," *Journal of Modern Applied Statistical Methods*, vol. 1, no. 2, p. 34, 2002.

[53] R. B. Anderson, "Conceptual distinction between the critical p value and the type i error rate in permutation testing," *Journal of modern applied statistical methods*, vol. 12, no. 1, p. 2, 2013.

[54] R. A. Giancristofaro and C. Brombin, "Overview of nonparametric combination-based permutation tests for multivariate multi-sample problems," *Statistica*, vol. 74, no. 3, pp. 233–246, 2014.

[55] F. Konietschke and M. Pauly, "Bootstrapping and permuting paired t-test type statistics," *Statistics and Computing*, vol. 24, no. 3, pp. 283–296, 2014.

[56] M. J. Anderson, "Permutational multivariate analysis of variance," *Department of Statistics, University of Auckland, Auckland*, vol. 26, pp. 32–46, 2005.

[57] S. Kherad-Pajouh and O. Renaud, "A general permutation approach for analyzing repeated measures anova and mixed-model designs," *Statistical Papers*, vol. 56, no. 4, pp. 947–967, 2015.

[58] D. Basso, F. Pesarin, L. Salmaso, and A. Solari, "Synchronized permutation tests in two-way anova," in *Permutation Tests for Stochastic Ordering and ANOVA*. Springer, 2009, pp. 133–172.

[59] E. Chung and M. Olivares, "Permutation test for heterogeneous treatment effects with a nuisance parameter," *Journal of Econometrics*, 2021.

[60] M. Giacalone, Z. Agata, P. C. Cozzucoli, and A. Alibrandi, "Bonferroni-holm and permutation tests to compare health data: methodological and applicative issues," *BMC medical research methodology*, vol. 18, no. 1, pp. 1–9, 2018.

[61] R. Arboretti Giancristofaro, S. Bonnini, L. Corain, and L. Salmaso, "Dependency and truncated forms of combinations in multivariate combination-based permutation tests and ordered categorical variables," *Journal of Statistical Computation and Simulation*, vol. 86, no. 18, pp. 3608–3619, 2016.

[62] R. A. Giancristofaro and S. Bonnini, "Moment-based multivariate permutation tests for ordinal categorical data," *Journal of Nonparametric Statistics*, vol. 20, no. 5, pp. 383–393, 2008.

[63] S. Hayasaka and T. E. Nichols, "Combining voxel intensity and cluster extent with permutation test framework," *Neuroimage*, vol. 23, no. 1, pp. 54–63, 2004.

[64] M. Marozzi *et al.*, "Some notes on nonparametric inferences and permutation tests," *Metron International Journal of Statistics*, vol. 60, no. 3-4, pp. 139–151, 2002.

[65] Y. Wan, J. Cohen, and R. Guerra, "A permutation test for the robust sib-pair linkage method," *Annals of human genetics*, vol. 61, no. 1, pp. 77–85, 1997.

[66] P. E. Kennedy and B. S. Cade, "Randomization tests for multiple regression," *Communications in Statistics-Simulation and Computation*, vol. 25, no. 4, pp. 923–936, 1996.

[67] M. Marozzi, "Some remarks about the number of permutations one should consider to perform a permutation test," *Statistica*, vol. 64, no. 1, pp. 193–201, 2004.

[68] E. Akkartal, M. Mendes, and E. Mendes, "Determination of suitable permutation numbers in comparing independent group means: a monte carlo simulation study," 2010.

[69] K. Mardia, "A non-parametric test for the bivariate two-sample location problem," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 29, no. 2, pp. 320–342, 1967.

[70] A. Janssen, "Studentized permutation tests for non-iid hypotheses and the generalized behrens-fisher problem," *Statistics & probability letters*, vol. 36, no. 1, pp. 9–21, 1997.

[71] G. Neuhaus and L.-X. Zhu, "Permutation tests for multivariate location problems," *Journal of Multivariate Analysis*, vol. 69, no. 2, pp. 167–192, 1999.

[72] M. Marozzi, "A bi-aspect nonparametric test for the two-sample location problem," *Computational statistics & data analysis*, vol. 44, no. 4, pp. 639–648, 2004.

[73] R. Fried and H. Dehling, "Robust nonparametric tests for the two-sample location problem," *Statistical Methods & Applications*, vol. 20, no. 4, pp. 409–422, 2011.

[74] M. Biswas and A. K. Ghosh, "A nonparametric two-sample test applicable to high dimensional data," *Journal of Multivariate Analysis*, vol. 123, pp. 160–171, 2014.

[75] F. Pesarin, L. Salmaso, E. Carrozzo, and R. Arboretti, "Union–intersection permutation solution for two-sample equivalence testing," *Statistics and Computing*, vol. 26, no. 3, pp. 693–701, 2016.

[76] S. Bonnini, M. Getnet *et al.*, "Permutation multivariate tests for treatment effect: theory and recent debelopments," in *SUSAN SSACAB 2019*. The Biostatistics Research Unit of the South African Medical Research Council, 2019, pp. 30–30.

[77] M. H. Samuh and R. A. Sanusi, "Permutation tests for two-sample location problem under extreme ranked set sampling," *Pakistan Journal of Statistics and Operation Research*, pp. 387–408, 2020.

[78] F. Bertoluzzo, F. Pesarin, and L. Salmaso, "On multi-sided permutation tests," *Communications in Statistics-Simulation and Computation*, vol. 42, no. 6, pp. 1380–1390, 2013.

[79] W. H. Finch, "Comparison of multivariate means across groups with ordinal dependent variables: a monte carlo simulation study," *Frontiers in Applied Mathematics and Statistics*, vol. 2, p. 2, 2016.

[80] V. Pillar, "How accurate and powerful are randomization tests in multivariate analysis of variance?" *Community Ecology*, vol. 14, no. 2, pp. 153–163, 2013.

[81] L. Finos, L. Salmaso, and A. Solari, "Conditional inference under simultaneous stochastic ordering constraints," *Journal of statistical planning and inference*, vol. 137, no. 8, pp. 2633–2641, 2007.

[82] G. J. Székely, M. L. Rizzo *et al.*, "Testing for equal distributions in high dimension," *InterStat*, vol. 5, no. 16.10, pp. 1249–1272, 2004.

[83] M. J. Anderson, "A new method for non-parametric multivariate analysis of variance," *Austral ecology*, vol. 26, no. 1, pp. 32–46, 2001.

[84] C. Brombin and L. Salmaso, "Multi-aspect permutation tests in shape analysis with small sample size," *Computational Statistics & Data Analysis*, vol. 53, no. 12, pp. 3921–3931, 2009.

[85] J. H. Friedman and L. C. Rafsky, "Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests," *The Annals of Statistics*, pp. 697–717, 1979.

[86] N. Henze, "A multivariate two-sample test based on the number of nearest neighbor type coincidences," *The Annals of Statistics*, vol. 16, no. 2, pp. 772–783, 1988.

[87] P. R. Rosenbaum, "An exact distribution-free test comparing two multivariate distributions based on adjacency," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 4, pp. 515–530, 2005.

[88] L. Corain and L. Salmaso, "Improving power of multivariate combination-based permutation tests," *Statistics and Computing*, vol. 25, no. 2, pp. 203–214, 2015.

[89] K. Neubert and E. Brunner, "A studentized permutation test for the non-parametric behrens–fisher problem," *Computational Statistics & Data Analysis*, vol. 51, no. 10, pp. 5192–5204, 2007.

[90] B. L. Welch, "The generalization of 'student's'problem when several different population varlances are involved," *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947.

[91] A. Welz, G. D. Ruxton, and M. Neuhäuser, "A non-parametric maximum test for the behrens–fisher problem," *Journal of Statistical Computation and Simulation*, vol. 88, no. 7, pp. 1336–1347, 2018.

[92] M.-A. Proschan *et al.*, "A permutation approach to the behrens-fisher problem," *Journal of the Korean Statistical Society*, vol. 33, no. 1, pp. 79–97, 2004.

[93] H. Scheffé, "Practical solutions of the behrens-fisher problem," *Journal of the American Statistical Association*, vol. 65, no. 332, pp. 1501–1508, 1970.

[94] M. J. Anderson, D. C. Walsh, K. Robert Clarke, R. N. Gorley, and E. Guerra-Castro, "Some solutions to the multivariate behrens–fisher problem for dissimilarity-based analyses," *Australian & New Zealand Journal of Statistics*, vol. 59, no. 1, pp. 57–79, 2017.

[95] B. F. Manly, *Randomization, bootstrap and Monte Carlo methods in biology.* chapman and hall/CRC, 2018.

[96] S. Hahn and L. Salmaso, “A comparison of different synchronized permutation approaches to testing effects in two-level two-factor unbalanced anova designs,” *Statistical Papers*, vol. 58, no. 1, pp. 123–146, 2017.

[97] K. J. Berry, P. W. Mielke Jr, and H. W. Mielke, “The fisher-pitman permutation test: an attractive alternative to the f test,” *Psychological reports*, vol. 90, no. 2, pp. 495–502, 2002.

[98] P. R. Peres-Neto and D. A. Jackson, “How well do multivariate data sets match? the advantages of a procrustean superimposition approach over the mantel test,” *Oecologia*, vol. 129, no. 2, pp. 169–178, 2001.

[99] N. Mantel and R. S. Valand, “A technique of nonparametric multivariate analysis,” *Biometrics*, pp. 547–558, 1970.

[100] H. Hotelling, “Relations between two sets of variates,” in *Breakthroughs in statistics*. Springer, 1992, pp. 162–190.

[101] S. S. Wilks, “Certain generalizations in the analysis of variance,” *Biometrika*, pp. 471–494, 1932.

[102] K. S. Pillai, “Some new test criteria in multivariate analysis,” *The Annals of Mathematical Statistics*, pp. 117–121, 1955.

[103] S. Hahn, F. Konietschke, and L. Salmaso, “A comparison of efficient permutation tests for unbalanced anova in two by two designs–and their behavior under heteroscedasticity,” *arXiv preprint arXiv:1309.7781*, 2013.

[104] D. Basso, M. Chiarandini, and L. Salmaso, “Synchronized permutation tests in replicated i× j designs,” *Journal of Statistical Planning and Inference*, vol. 137, no. 8, pp. 2564–2578, 2007.

[105] R. Arboretti, R. Ceccato, L. Corain, F. Ronchi, and L. Salmaso, “Multivariate small sample tests for two-way designs with applications to industrial statistics,” *Statistical Papers*, vol. 59, no. 4, pp. 1483–1503, 2018.

[106] L. Corain and L. Salmaso, “A critical review and a comparative study on conditional permutation tests for two-way anova,” *Communications in Statistics—Simulation and Computation®*, vol. 36, no. 4, pp. 791–805, 2007.

[107] D. I. Warton and H. M. Hudson, "A manova statistic is just as powerful as distance-based statistics, for multivariate abundances," *Ecology*, vol. 85, no. 3, pp. 858–874, 2004.

[108] E. Brunner and M. L. Puri, "Nonparametric methods in factorial designs," *Statistical papers*, vol. 42, no. 1, pp. 1–52, 2001.

[109] E. Chung and J. P. Romano, "Exact and asymptotically robust permutation tests," *The Annals of Statistics*, vol. 41, no. 2, pp. 484–507, 2013.

[110] G. Neuhaus, "Conditional rank tests for the two-sample problem under random censorship," *The Annals of Statistics*, pp. 1760–1779, 1993.

[111] B. H. McArdle and M. J. Anderson, "Fitting multivariate models to community data: a comment on distance-based redundancy analysis," *Ecology*, vol. 82, no. 1, pp. 290–297, 2001.

[112] J. C. Gower and P. Legendre, "Metric and euclidean properties of dissimilarity coefficients," *Journal of classification*, vol. 3, no. 1, pp. 5–48, 1986.

[113] P. Legendre and M. J. Anderson, "Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments," *Ecological monographs*, vol. 69, no. 1, pp. 1–24, 1999.

[114] R. Arboretti, S. Bonnini, L. Corain, and L. Salmaso, "A permutation approach for ranking of multivariate populations," *Journal of Multivariate Analysis*, vol. 132, pp. 39–57, 2014.

[115] S. Bonnini, "Testing for heterogeneity with categorical data: permutation solution vs. bootstrap method," *Communications in Statistics-Theory and Methods*, vol. 43, no. 4, pp. 906–917, 2014.

[116] R. A. Giancristofaro and S. Bonnini, "Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives," *Statistical Methods and Applications*, vol. 18, no. 2, pp. 221–236, 2009.

[117] ——, "Permutation tests for heterogeneity comparisons in presence of categorical variables with application to university evaluation," *Metodoloski Zvezki*, vol. 4, no. 1, p. 21, 2007.

[118] R. A. Giancristofaro, "Multivariate permutation tests in genetics," *Statistica*, vol. 62, no. 4, pp. 681–694, 2002.

[119] P. H. Westfall, R. D. Tobias, and R. D. Wolfinger, *Multiple comparisons and multiple tests using SAS.* SAS Institute, 2011.

[120] S. S. Gupta and S. Panchapakesan, *Multiple decision procedures: theory and methodology of selecting and ranking populations.* SIAM, 2002.

[121] A. M. Mami, M. M. Mansour, and A. M. Jaber, "Scholars journal of physics, mathematics and statistics," 2020.

[122] N. E. Helwig and M. N. E. Helwig, "Package 'npreg'," 2019.

[123] J. M. Rodriguez-Poo and A. Soberón, "Nonparametric estimation of fixed effects panel data varying coefficient models," *Journal of Multivariate Analysis*, vol. 133, pp. 95–122, 2015.

[124] J. S. Racine, *Nonparametric econometrics: A primer.* Now Publishers Inc, 2008, vol. 4.

[125] R. C. Team *et al.*, "R: A language and environment for statistical computing," 2013.

[126] M. Sestelo, N. M. Villanueva, L. Meira-Machado, and J. Roca-Pardiñas, "npregfast: An r package for nonparametric estimation and inference in life sciences," *Journal of Statistical Software*, vol. 82, no. 1, pp. 1–27, 2017.

[127] J. Nyblom, "Permutation tests in linear regression," in *Modern nonparametric, robust and multivariate methods.* Springer, 2015, pp. 69–90.

[128] A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith, and T. E. Nichols, "Permutation inference for the general linear model," *Neuroimage*, vol. 92, pp. 381–397, 2014.

[129] R. E. Wheeler and M. Torchiano, "Permutation tests for linear models in r," *The Comprehensive R Archive Network*, vol. 1, no. 2, 2010.

[130] C. J. Ter Braak, "Permutation versus bootstrap significance tests in multiple regression and anova," in *Bootstrapping and related techniques.* Springer, 1992, pp. 79–85.

[131] M. J. Anderson and P. Legendre, "An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model," *Journal of statistical computation and simulation*, vol. 62, no. 3, pp. 271–303, 1999.

[132] E. S. Edgington, "Randomization tests. new york: M," 1995.

[133] D. Freedman and D. Lane, "A nonstochastic interpretation of reported significance levels," *Journal of Business & Economic Statistics*, vol. 1, no. 4, pp. 292–298, 1983.

[134] S. Kherad-Pajouh and O. Renaud, "An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova," *Computational Statistics & Data Analysis*, vol. 54, no. 7, pp. 1881–1893, 2010.

[135] H. Önder and Z. Cebeci, "Use and comparison of permutation tests in linear models," *Anadolu Tarım Bilimleri Dergisi*, vol. 24, no. 2, pp. 93–97, 2009.

[136] P. Legendre, "Comparison of permutation methods for the partial correlation and partial mantel tests," *Journal of statistical computation and simulation*, vol. 67, no. 1, pp. 37–73, 2000.

[137] L. Finos and D. Basso, "Permutation tests for between-unit fixed effects in multivariate generalized linear mixed models," *Statistics and Computing*, vol. 24, no. 6, pp. 941–952, 2014.

[138] D. Basso and L. Finos, "Exact multivariate permutation tests for fixed effects in mixed-models," *Communications in Statistics-Theory and Methods*, vol. 41, no. 16-17, pp. 2991–3001, 2012.

[139] O. E. Lee and T. M. Braun, "Permutation tests for random effects in linear mixed models," *Biometrics*, vol. 68, no. 2, pp. 486–493, 2012.

[140] C. Brombin, "A nonparametric permutation approach to statistical shape analysis," 2009.

[141] F. B. Butar and A. Bandulasiri, "Comparison of the power of the paired samples using permutation tests," *J Math Sci Math Educ*, vol. 4, no. 2, pp. 19–31, 2009.

[142] J. Kaiser, "An exact and a monte carlo proposal to the fisher–pitman permutation tests for paired replicates and for independent samples," *The Stata Journal*, vol. 7, no. 3, pp. 402–412, 2007.

[143] F. Konietschke, A. C. Bathke, S. W. Harrar, and M. Pauly, "Parametric and nonparametric bootstrap methods for general manova," *Journal of Multivariate Analysis*, vol. 140, pp. 291–301, 2015.

[144] A. Janssen and T. Pauls, “How do bootstrap and permutation tests work?” *The Annals of statistics*, vol. 31, no. 3, pp. 768–806, 2003.

[145] C. Brombin, L. Salmaso, G. Ferronato, and P.-F. Galzignato, “Multi-aspect procedures for paired data with application to biometric morphing,” *Communications in Statistics—Simulation and Computation®*, vol. 40, no. 1, pp. 1–12, 2010.

[146] M. Pauly, “Discussion about the quality of f-ratio resampling tests for comparing variances,” *Test*, vol. 20, no. 1, pp. 163–179, 2011.

[147] M. Omelka and M. Pauly, “Testing equality of correlation coefficients in two populations via permutation methods,” *Journal of Statistical Planning and Inference*, vol. 142, no. 6, pp. 1396–1406, 2012.

[148] J. J. Higgins and W. Noble, “A permutation test for a repeated measures design,” 1993.

[149] I. G. Dragset, “Analysis of longitudinal data with missing values.: Methods and applications in medical statistics.” Master’s thesis, Institutt for matematiske fag, 2009.

[150] S. Friedrich, E. Brunner, and M. Pauly, “Permuting longitudinal data despite all the dependencies,” *arXiv preprint arXiv:1509.05570*, 2015.

[151] S. Prinja, N. Gupta, and R. Verma, “Censoring in clinical trials: review of survival analysis techniques,” *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, vol. 35, no. 2, p. 217, 2010.

[152] X. Liu, *Methods and applications of longitudinal data analysis.* Elsevier, 2015.

[153] R. J. Little, D. B. Rubin, and S. Z. Zangeneh, “Conditions for ignoring the missing-data mechanism in likelihood inferences for parameter subsets,” *Journal of the American Statistical Association*, vol. 112, no. 517, pp. 314–320, 2017.

[154] R. J. Little and D. B. Rubin, *Statistical analysis with missing data.* John Wiley & Sons, 2019, vol. 793.

[155] D. B. Rubin, “Inference and missing data,” *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[156] M. Ditzhaus, A. Janssen, and M. Pauly, "Permutation inference in factorial survival designs with the casanova," *arXiv preprint arXiv:2004.10818*, 2020.

[157] D. Yu, J. Lim, F. Liang, K. Kim, B. S. Kim, and W. Jang, "Permutation test for incomplete paired data with application to cdna microarray data," *Computational Statistics & Data Analysis*, vol. 56, no. 3, pp. 510–521, 2012.

[158] R. A. GIANCRISTOFARO, M. Bolzan, F. Campigotto, L. Corain, and L. Salmaso, "Combination-based permutation testing in survival analysis," *Quaderni di Statistica*, vol. 12, pp. 21–44, 2010.

[159] R. L. Einsporn and D. Habtzghi, "Combining paired and two-sample data using a permutation test," *Journal of Data Science*, vol. 11, no. 4, pp. 767–779, 2013.

[160] D. S. Bhoj, "Testing equality of means of correlated variates with missing observations on both responses," *Biometrika*, vol. 65, no. 1, pp. 225–228, 1978.

[161] M. Brendel, A. Janssen, C.-D. Mayer, and M. Pauly, "Weighted logrank permutation tests for randomly right censored life science data," *Scandinavian Journal of Statistics*, vol. 41, no. 3, pp. 742–761, 2014.

[162] B. D. Paul, "A history of the concept of sustainable development: literature review," *The Annals of the University of Oradea, Economic Sciences Series*, vol. 17, no. 2, pp. 576–580, 2008.

[163] S. Asefa, *The economics of sustainable development.* WE Upjohn Institute, 2005.

[164] O. Langhelle, "Sustainable development: exploring the ethics of our common future," *International Political Science Review*, vol. 20, no. 2, pp. 129–149, 1999.

[165] I. Burton, "Report on reports: Our common future: The world commission on environment and development," *Environment: Science and Policy for Sustainable Development*, vol. 29, no. 5, pp. 25–29, 1987.

[166] W. Colglazier, "Sustainable development agenda: 2030," *Science*, vol. 349, no. 6252, pp. 1048–1050, 2015.

[167] Y. Farsari and P. Prastacos, “Sustainable development indicators: an overview,” *Foundation for the Research and Technology Hellas*, vol. 24, pp. 197–208, 2002.

[168] T. W. Bank, “The world bank skills profile survey (es). 2017.” https://microdata.worldbank.org/index.php/catalog/3445.

[169] E. Baro, S. Degoul, R. Beuscart, and E. Chazard, “Toward a literature-driven definition of big data in healthcare,” *BioMed research international*, vol. 2015, 2015.

[170] H. Özköse, E. S. Arı, and C. Gencer, “Yesterday, today and tomorrow of big data,” *Procedia-Social and Behavioral Sciences*, vol. 195, pp. 1042–1050, 2015.

[171] R. Johnson, “Wichern. d.(2007). applied multivariate statistical analysis,” 1997.

[172] A. Farcomeni and L. Greco, *Robust methods for data reduction*. CRC press, 2016.

[173] F. Pesarin and L. Salmaso, “Finite-sample consistency of combination-based permutation tests with application to repeated measures designs,” *Journal of Nonparametric Statistics*, vol. 22, no. 5, pp. 669–684, 2010.

[174] S. Bonnini, “Multivariate approach for comparative evaluations of customer satisfaction with application to transport services,” *Communications in Statistics-Simulation and Computation*, vol. 45, no. 5, pp. 1554–1568, 2016.

[175] K. R. Clarke, “Non-parametric multivariate analyses of changes in community structure,” *Australian journal of ecology*, vol. 18, no. 1, pp. 117–143, 1993.

[176] V. D. P. Pillar and L. Orlóci, “On randomization testing in vegetation science: multifactor comparisons of relevé groups,” *Journal of Vegetation Science*, vol. 7, no. 4, pp. 585–592, 1996.

[177] P. W. Mielke Jr and K. J. Berry, “Multivariate tests for correlated data in completely randomized designs,” *Journal of Educational and Behavioral Statistics*, vol. 24, no. 2, pp. 109–131, 1999.

[178] D. Polko-Zajac *et al.*, "A comparative study of the power of parametric and permutation tests for a multidimensional two-sample location problem," *Argumenta Oeconomica Cracoviensia*, vol. 2, no. 23, pp. 69–79, 2020.

[179] ——, "On permutation location–scale tests," *Statistics in Transition. New Series*, vol. 20, no. 4, pp. 153–166, 2019.

[180] S. Bonnini, N. Prodi, L. Salmaso, and C. Visentin, "Permutation approaches for stochastic ordering," *Communications in Statistics-Theory and Methods*, vol. 43, no. 10-12, pp. 2227–2235, 2014.

[181] M. Selvam, J. Gayathri, V. Vasanth, K. Lingaraja, and S. Marxiaoli, "Determinants of firm performance: A subjective model," *Int'l J. Soc. Sci. Stud.*, vol. 4, p. 90, 2016.

[182] M. L. L. Sa, S. Choon-Yin, Y. K. Chai, and J. H. A. Joo, "Knowledge creation process, customer orientation and firm performance: Evidence from small hotels in malaysia," *Asia Pacific Management Review*, vol. 25, no. 2, pp. 65–74, 2020.

[183] P. J. Richard, T. M. Devinney, G. S. Yip, and G. Johnson, "Measuring organizational performance: Towards methodological best practice," *Journal of management*, vol. 35, no. 3, pp. 718–804, 2009.

[184] F. Memon, N. A. Bhutto, and G. Abbas, "Capital structure and firm performance: A case of textile sector of pakistan," *Asian Journal of Business and Management Sciences*, vol. 1, no. 9, pp. 9–15, 2012.

[185] H. Malik, "Determinants of insurance companies profitability: an analysis of insurance sector of pakistan," *Academic research international*, vol. 1, no. 3, pp. 315–321, 2011.

[186] T. W. Anderson, "An introduction to multivariate statistical analysis," Wiley New York, Tech. Rep., 1962.

[187] R. Davidson, J. G. MacKinnon *et al.*, *Econometric theory and methods*. Oxford University Press New York, 2004, vol. 5.

[188] R. Johnson, "Wichern. d.(2007). applied multivariate statistical analysis."

[189] M. Neuhäuser, "A note on the exact test based on the baumgartner–weiß–schindler statistic in the presence of ties," *Computational statistics & data analysis*, vol. 42, no. 4, pp. 561–568, 2003.

[190] L. Finos, F. Pesarin, L. Salmaso, and A. Solari, “Exact inference for multivariate ordered alternatives,” *Statistical Methods and Applications*, vol. 17, no. 2, pp. 195–208, 2008.

[191] F. E. Kennedy, “Randomization tests in econometrics,” *Journal of Business & Economic Statistics*, vol. 13, no. 1, pp. 85–94, 1995.

[192] B. S. Cade and J. D. Richards, “Permutation tests for least absolute deviation regression,” *Biometrics*, pp. 886–902, 1996.

[193] Q. Wu and P. Vos, “Permutation inference distribution for linear regression and related models,” *Journal of Nonparametric Statistics*, vol. 31, no. 3, pp. 722–742, 2019.

[194] N. E. Helwig, “Multivariate linear regression,” 2017.

[195] S. Searle, “Linear models john wiley & sons,” *Inc., New York-London-Sydney-Toronto*, 1971.

[196] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis.* Wiley, 2001.

[197] B. Kim, “Understanding diagnostic plots for linear regression analysis,” *University of Virginia website. Charlottesville, VA: University of Virginia.[Google Scholar]*, 2015.

[198] W. N. Venables and D. M. Smith, “The r development core team,” *An Introduction to R, Version*, vol. 1, no. 0, 2003.

[199] J. Fox and S. Weisberg, “Multivariate linear models in r,” *An R Companion to Applied Regression. Los Angeles: Thousand Oaks*, 2011.

[200] T. Pham-Gia, “Exact distribution of the generalized wilks's statistic and applications,” *Journal of Multivariate Analysis*, vol. 99, no. 8, pp. 1698–1716, 2008.

[201] O. Taouab and Z. Issor, “Firm performance: Definition and measurement models,” *European Scientific Journal*, vol. 15, no. 1, pp. 93–106, 2019.

[202] I. Jajri and R. Ismail, “Impact of labour quality on labour productivity and economic growth,” *African Journal of Business Management*, vol. 4, no. 4, pp. 486–495, 2010.

[203] R. Ayres and V. Voudouris, "The economic growth enigma: Capital, labour and useful energy?" *Energy Policy*, vol. 64, pp. 16–28, 2014.

[204] S. N. Durlauf, A. Kourtellos, and A. Minkin, "The local solow growth model," *European Economic Review*, vol. 45, no. 4-6, pp. 928–940, 2001.

[205] N. Rosenberg, "Adam smith on the division of labour: two views or one?" *Economica*, vol. 32, no. 126, pp. 127–139, 1965.

[206] N. P. Commission *et al.*, "Growth and transformation plan ii (gtp ii)(2015/16-2019/20)," *Addis Ababa: Federal Democratic Republic of Ethiopia*, 2016.

[207] L. T. Abdisa, "Power outages, economic cost, and firm performance: evidence from ethiopia," *Utilities Policy*, vol. 53, pp. 111–120, 2018.

[208] R. H. Chenhall and K. Langfield-Smith, "Multiple perspectives of performance measures," *European management journal*, vol. 25, no. 4, pp. 266–282, 2007.

[209] J. B. Santos and L. A. L. Brito, "Toward a subjective measurement model for firm performance," *BAR-Brazilian Administration Review*, vol. 9, no. SPE, pp. 95–117, 2012.

[210] A. Neely, M. Kennerley, and C. Adams, "Performance measurement frameworks: a review," *Business Performance Measurement*, pp. 143–162, 2007.

[211] S. Ding, A. Guariglia, and R. Harris, "The determinants of productivity in chinese large and medium-sized industrial firms, 1998–2007," *Journal of Productivity Analysis*, vol. 45, no. 2, pp. 131–155, 2016.

[212] J. Guerrero-Villegas, L. Sierra-García, and B. Palacios-Florencio, "The role of sustainable development and innovation on firm performance," *Corporate Social Responsibility and Environmental Management*, vol. 25, no. 6, pp. 1350–1362, 2018.

[213] G. Cainelli, M. Mazzanti, and R. Zoboli, "Environmentally oriented innovative strategies and firm performance in services. micro-evidence from italy," *International Review of Applied Economics*, vol. 25, no. 1, pp. 61–85, 2011.

[214] A. K. Gupta and N. Gupta, "Effect of corporate environmental sustainability on dimensions of firm performance–towards sustainable development: Evidence from india," *Journal of cleaner production*, vol. 253, p. 119948, 2020.

[215] K. Nayal, R. D. Raut, V. S. Yadav, P. Priyadarshinee, and B. E. Narkhede, "The impact of sustainable development strategy on sustainable supply chain firm performance in the digital transformation era," *Business Strategy and the Environment*, 2021.

[216] G. G. Noja, M. Cristea, C. N. Jurcut, A. Buglea, and I. Lala Popa, "Management financial incentives and firm performance in a sustainable development framework: Empirical evidence from european companies," *Sustainability*, vol. 12, no. 18, p. 7247, 2020.

[217] M. Ayyagari, A. D. Kunt, and V. Maksimovic, *How important are financing constraints?: the role of finance in the business environment*. World Bank Publications, 2006, vol. 3820.

[218] T. W. Bank, "The world bank.ethiopia enterprise survey (es). 2015." 10 2020, https://microdata.worldbank.org/index.php/catalog/2577.

[219] C. C. Williams and A. M. Kedir, "The impacts of corruption on firm performance: Some lessons from 40 african countries," *Journal of Developmental Entrepreneurship*, vol. 21, no. 04, p. 1650022, 2016.

[220] J. Schumpeter, "Creative destruction," *Capitalism, socialism and democracy*, vol. 825, pp. 82–85, 1942.

[221] X. Cirera and S. Muzi, *Measuring firm-level innovation using short questionnaires: evidence from an experiment*. The World Bank, 2016.

[222] B. Fowowe, "Access to finance and firm performance: Evidence from african countries," *Review of development finance*, vol. 7, no. 1, pp. 6–17, 2017.

[223] T. A. Byrd, J. P. Pitts, A. M. Adrian, and N. W. Davidson, "Examination of a path model relating information technology infrastructure with firm performance," *Journal of Business Logistics*, vol. 29, no. 2, pp. 161–187, 2008.

[224] S. Mithas, N. Ramasubbu, M. Krishnan, and V. Sambamurthy, "Information technology infrastructure capability and firm performance: An empirical analysis," *Ann Arbor*, vol. 1001, pp. 48 109–1234, 2005.

[225] L. Castany, E. López-Bazo, and R. Moreno, “Differences in total factor productivity across firm size-a distributional analysis,” 2005.

[226] J. Van Biesebroeck, “Exporting raises productivity in sub-saharan african manufacturing firms,” *Journal of International economics*, vol. 67, no. 2, pp. 373–391, 2005.

APPENDIX A

APPENDIX

Table A.1: The Power of tests for $\rho_x = 0.3$, $\rho_\epsilon = 0$, $\alpha = 0.05$ and B=$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.584	0.987	1	1	1	1
CPT Tippett	0.57	0.978	0.999	1	1	1
Parametric Pillai	0.696	0.995	1	1	1	1

Table A.2: .The Power of test for $\rho_x = 0.0$, $\rho_\epsilon = 0.8$,$\alpha = 0.05$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.567	0.985	0.999	1	1	1
CPT Tippett	0.55	0.969	0.998	1	1	1
Parametric Pillai	0.882	0.999	1	1	1	1

Table A.3: The Power of test for $\rho_x = 0.0$, $\rho_\epsilon = 0.3$,$\alpha = 0.05$ and B=$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.652	0.986	1	1	1	1
CPT Tippett	0.511	0.962	1	1	1	1
Parametric Pillai	0.71	0.997	1	1	1	1

Table A.4: The Power of test for $\rho_x = 0.0$, $\rho_\epsilon = 0.8$, $\alpha = 0.05$ and B=$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.613	0.992	1	1	1	1
CPT Tippett	0.535	0.968	1	1	1	1
Parametric Pillai	0.885	1	1	1	1	1

Table A.5: The Power of test for $\rho_x = 0.0$, $\rho_\epsilon = 0.8$,$\alpha = 0.05$ and B=$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.256	0.675	0.682	0.699	0.708	0.728
CPT Tippett	0.385	0.838	0.978	0.998	1	1
Parametric Pillai	0.393	0.985	1	1	1	1

Table A.6: The Power of test for $\rho_x = 0.3$, $\rho_\epsilon = 0$,$\alpha = 0.05$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$

Sample size	10	20	30	40	70	100
CPT Fisher	0.448	0.923	0.997	1	1	1
CPT Tippett	0.494	0.93	0.995	1	1	1
Prametric Pillai	0.326	0.973	1	1	1	1

Table A.7: The Power of test for large Y, large X, and $n = 848$, $\rho_x = 0.3$, $\rho_\epsilon = 0.3$

Sample size	848
CPT Fisher	1
CPT Tippett	1
Parametric Pillai	1

Table A.8: The correlation between Firm performance variables

Variables	Gs	GE	GLP
GS	1	0.2	0.5
GE	0.2	1	0.1
GLP	0.5	0.1	1

Table A.9: Descriptive statistics for continuous variables

Variables	Min	1stQ	Median	3rdQ	Max	Mean	SD
GS	0.00	0.00	9.42	24.55	66.73	18.68	23.21
GE	0.00	0.00	2.52	12.53	1000	17.65	73.29
GLP	0.00	0.00	20.09	21.96	683.33	20.08	44.31
Firm ownership	0.00	50.00	99.00	100.00	100.00	77.41	28.64
Power outage	0.00	6.00	12.00	15.00	365	14.97	18.88
trade	0.00	60.00	65.00	65.00	365	65.19	46.0.17
unskilled labour	0.00	0.00	30.00	30.00	5050.00	30.5	179.508

Table A.10: Organizational Well-being variables

Code	Statement
A.01	My working place is safe
A.02	I have been informed about the risks connected to my job
A.03	I am satisfied about the environment of my working place
A.04	I have suffered harassment
A.05	My dignity has been harmed at work
A.06	At work the smoking ban is respected
A.07	I usually take enough breaks
A.08	I can work hard
A.09	I am not confortable when I am working
A.10	The colleagues are not polite with me
A.11	I am allowed to take a break when I wish
A.12	I don't have the chance to take enough breaks
B.10	At work I have suffered bullying
B.01	In the workplace I am respected in my trade union membership
B.02	In the workplace I am respected in my political orientation
B.03	In the workplace I am respected in my religious faith
B.04	My gender identity is an obstacle to my enhancement at work
B.05	In the workplace I am respected in my ethnicity and race
B.06	In the workplace I am respected in relation to my mother tongue
B.07	My age is an obstacle to my enhancement at work
B.08	In the workplace I am respected in relation to my mother tongue
C.01	The workload is assigned with equity
C.02	The responsabilities are assigned with equity
C.03	My salary is proportional to the commitment
C.04	The pay is differentiated according to quantity and quality of work
C.05	My manager makes work decisions impartially
D.01	At UNIFE the profession path employee is well defined and clear
D.02	At UNIFE the career opportunities depend on merit
D.03	UNIFE gives the possibility to develop skills and aptitudes of individuals
D.04	My current role is appropriate to my professional profile
D.05	I am satisfied with my professional path within UNIFE
E.01	I know what is expected of my work
E.02	I have the skills to do my job
E.03	I have the resources and tools to do my job
E.04	I have an adequate level of autonomy in my work
E.05	My work gives me a sense of personal fulfillment
E.06	I know how to do my job
E.07	I understand what is expected of me at work
E.08	I have freedom of choice in deciding how to do my job
E.09	I have unattainable deadlines

Table A.11: Adjusted and raw p-values

variables	Raw p-values	Adjusted p-values
cons.ex.pricefood	0.97990201	1
cons.ex.nonfood.price	0.78092191	1
asset.price.sell.birr.	0.93850615	1
rooms_n	0.50044996	0.9961
water_home	1	1
tedu_h	1	1
thealth_h	1	1
tmarket_h	1	1
twater_h	1	1
electricity_phone	0.22437756	0.8616
electricity_hours 0.	0.83751625	1
electricity_hours_kdk	1	1
cook	0.50074993	0.9961
share_facility	0.21947805	0.8297
tmarket_kdk	1	1
acc_int	0.50174983	0.9961
phone_network	0.02649735	**0.01747**
share_num	0.98190181	1
land_access_yn	0.22567743	0.8688
acc_bank1	0.2229777	0.8616
acc_bank2	1	1
beh_risk_bus	0.50074993	0.9961
satisf_life	1	1
satisf_prev	0.92320768	1
rate_personal	0.72612739	1
rate_personal_future	0.0219978	**0.01423**
health_satisfaction	0.49455054	0.9961
school_satisfaction	0.14548545	0.6343
empl_satisfaction	0.49555044	0.9961
safe_violence	1	1

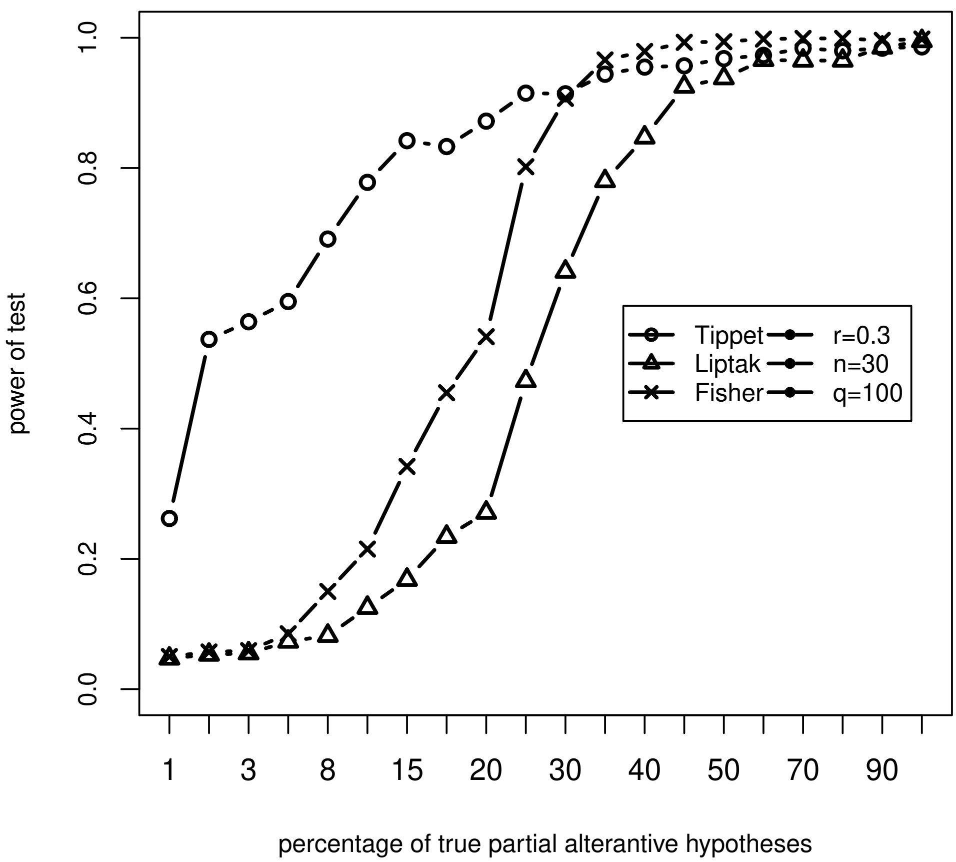


Figure A.1: Comparing the power of combing function for different percentage