Fully Nonparametric Methods for Partially Complete Data in Repeated Measures Design

MERGA BELINA FEYASA

2018

ADDIS ABABA UNIVERSITY

DEPARTMENT OF STATISTICS

DOCTORAL THESIS

Fully Nonparametric Methods for Partially Complete Data in Repeated Measures Design

---

Author:
Merga Belina FEYASA

Supervisors:
Dr. Solomon W. HARRAR
Prof. Eshetu WENCHEKO

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics

in the

Department of Statistics
Addis Ababa University

July 26, 2018

ADDIS ABABA UNIVERSITY

GRADUATE PROGRAMS

Fully Nonparametric Methods for Partially Complete Data in Repeated Measures Design

By

Merga Belina FEYASA

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics

Approved by Examining Board:

Dr. Solomon W. HARRAR		
Supervisor	Signature	Date
Prof. Eshetu WENCHEKO		
Supervisor	Signature	Date
Prof. Asheber ABEBE		
Examiner	Signature	Date
Dr. Emmanuel GABREYOHANNES		
Examiner	Signature	Date
Mekonnen TADESSE		
Chairman	Signature	Date

July 26, 2018

Declaration of Authorship

I, Merga Belina FEYASA, declare that this thesis titled, "Fully Nonparametric Methods for Partially Complete Data in Repeated Measures Design" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has not previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"It is important to know the appropriate theory behind Statistical analysis in order to do it right!"

Author Unknown

Addis Ababa University

College of Natural and Computational Sciences

Department of Statistics

Doctor of Philosophy in Statistics

Fully Nonparametric Methods for Partially Complete Data in Repeated Measures Design

by Merga Belina FEYASA

Abstract

In this dissertation, two related but distinct problems are studied. The first one is a fully nonparametric rank-based method for comparing samples with partially paired data. Partially-paired (correlated) data naturally arise, for example, as a result of missing values, in incomplete block designs or meta analysis. In the nonparametric setup, treatment effects are characterized in terms of functionals of distribution functions and the only assumption needed is that the marginal distributions to be non-degenerate. The setup accommodates binary, ordered categorical, discrete and continuous data in a seamless fashion. The use of nonparametric effects also addresses the Behrens-Fisher problem from the nonparametric point of view and allows construction of confidence intervals. Although, the nonparametric methods are mainly asymptotic, methods for small sample approximations are also proposed. The second problem studied is also a fully nonparametric rank-based method but for partially repeated measures data. Here a vector of nonparametric relative effect measures are defined and linear hypotheses on these effects are considered. A multitude of tests are available for hypothesis related to a vector of relative effects. We focus on asymptotic results and finite sample performance for Wald-type statistic (WTS), ANOVA-type statistic (ATS) and Multiple Comparison Test Procedure (MCTP). Notwithstanding the limitation that the theory is thoroughly investigated for the three time point case, the results can formally be extended to the more general set up but the involved expressions will be much more complicated. The finite sample behavior of the tests are investigated via simulation studies. The results provide numerical evidence of favorable performance of the nonparametric method. The new methods have overwhelming power advantage when treatment effects are reflected in the shape of the distribution while they perform comparably better with parametric methods for location-type alternatives. Data from a therapeutic-drug clinical trial and a randomized controlled epidemiological study are used to illustrate the application of the methods.

Acknowledgements

First and foremost, I would like to thank God Almighty for his numerous graces, help in all aspects and the successful completion of this study. I would also like to acknowledge my supervisors for their many helpful suggestions and ideas that went into this dissertation. I wish to thank Dr. Solomon W. Harrar for his insight, encouragement and friendship throughout this process. His confidence in my skills along with his guidance and support have helped me to make it through both the ups and downs of this process. Professor Eshetu Wencheko has been instrumental in the development of this research and I would like to thank him for his invaluable input, ideas and thoughtful guidance to make this research work happen. I could not have completed this without the love and support of my family. In particular, I would like to acknowledge my uncle Mr. Wakgari Feyasa who offered advice and assistance at every stage of the process. His belief in me has and will continue to be instrumental in my future success. In addition he has lived his life in such a way that his career has served to help me in the development of my personal and professional goals. I would also like to thank my father and my mother for their countless words of encouragement and support. My special thanks goes to Mr. Mekonnen Tadesse, chairman of our department. My friends have offered unconditional love and support which has been invaluable during this process. I am very much grateful to Addis Ababa University for allowing me to pursue my study, and the University of Kentucky for providing me the opportunity to take PhD level courses at the Department of Statistics, University of Kentucky. I cannot thank enough the faculty at the two Departments of Statistics at Addis Ababa University and University of Kentucky. This has been a difficult but rewarding process and the generosity, patience, acceptance and assistance of the friends I have made within these departments have been responsible for much of my success. I would like to thank them all.

Merga Belina FEYASA
July 26, 2018

Contents

List of Figures

List of Tables

List of Abbreviations

MAR Missing At Random

MCAR Missing Completely At Random

EM Expectation-Maximization

MI Multiple Imputation

MP Missing Percentage

WTS Wald-Type Statistic

ATS ANOVA-Type Statistic

MCTP Multiple Comparison Test Procedure

GLMM General Linear Mixed Model

Dedicated to Atsede Geleta.

Chapter 1

Introduction

In this chapter, we provide an overview of repeated measures in general. The primary focus will be nonparametric methods for data from repeated measures designs. Later in the chapter we describe missing data methodology in general and specialized nonparametric approaches with missing data analysis in particular. We close this chapter with a discussion of the objectives of this dissertation. In particular, we identify the gap in methodological development that this research fills.

1.1 Repeated Measures

A repeated measures model refers to a model where randomly chosen subjects are observed repeatedly under the same or different treatments. They include growth curves, longitudinal data and others. Many other scenarios can result in repeated measures, like spatial observations, not just in time (Diggle et al., 2013). The important feature is that multiple measurements are being made on the same experimental unit. Reasons for performing repeated measurement experiments/studies are intent to reduce the between-subject variation of the measurements, decrease the number of subjects needed to be recruited/trained for the study and the need to studying changes over time (Diggle et al., 2013).

There are various methods for analyzing repeated measures data in a parametric setup. Although our interest rests on the development of nonparametric methods for repeated measures data with missing observations, we highlight one of the existing standard tools, namely the linear mixed-effects model. Let $\boldsymbol{Y}_i$ denote the n_i-dimensional vector of measurements available for subject $i = 1, ..., n$. A General Linear Mixed Model (GLMM) for the i^{th} subject is commonly expressed by the model

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{1.1}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are $(n_i \times p)$ and $(n_i \times q)$ matrices of known covariates, $\boldsymbol{\beta}$ is a $p \times 1$ column vector of the fixed-effects regression coefficients. The random effects $\boldsymbol{b}_i$ and residual components $\boldsymbol{\epsilon}_i$ are assumed to be independent with distributions $N(0, \Sigma_{\boldsymbol{b}_i})$, and $N(0, \Sigma_{\boldsymbol{\epsilon}_i})$, respectively. The linear mixed-effects model in equation (1.1) is a commonly used tool for, among others, variance component models and for analyzing repeated measures like longitudinal data (Diggle et al., 2013).

Longitudinal Data Analysis is statistical data analysis where measurements on each subject/unit are taken repeatedly over time. Designs in longitudinal studies, are one of the most widely used repeated measures designs. The primary merit of a longitudinal study is its effectiveness for studying change over time. Longitudinal data analysis requires special statistical methods because the set of observations on one subject tends to be intercorrelated, hence the assumption of independence is no longer tenable and the correlation must be taken into account to come up with valid scientific inference. Generally data collected from different experimental units will be assumed to be independent whereas from the same unit may be, to some degree, dependent (Brunner et al., 2002). Longitudinal study areas include clinical trials, observational studies in humans, animals; studies of growth and decay in chemistry and so on (Diggle et al., 2013).

Longitudinal data often arise in different factorial settings, for instance, when one homogeneous group of subjects is observed repeatedly at *t* time points, where each subject is observed several times at each time point; one homogeneous group of subjects is observed repeatedly under different conditions, where each subject is observed under each condition at *t* time points or more than one homogeneous group of subjects are observed repeatedly at *t* time points. In the latter case, a common objective in the analysis of longitudinal data is to assess whether the way in which the response changes over time is different across treatment groups. In the repeated measurements, the test for group by time interaction is sometimes called the test for parallelism. Our discussion in Chapter 3, which is limited to two groups and two time points, focuses on a group-time interaction effect in repeated measures design. Basically downsides of using GLMM or other parametric or semi parametric methods for analysis of repeated measures is that when the underlying assumptions for the models are grossly violated their usability becomes questionable. In addition, they suffer from presence of outlier observations in the dataset. So the need for other models which averts this problem arises. The methods we are developing are fully nonparametric and robust to the issues mentioned above.

1.2 Nonparametric Methods for Repeated Measures

Nonparametric methods for dependent data have been developed in a body of literature spanning over four decades. One of the earlier attempts was rank-based methods for repeated measures by Brunner and Neumann (1982), which was later generalized by Thompson (1990) and Thompson (1991) under a continuity assumption. *A nonparametric marginal model* arises when marginal distributions are used to describe *treatment effects* and to *formulate hypotheses*. The idea to formulate nonparametric hypotheses for the marginal models in terms of the distribution functions, where marginal distributions are used to describe treatment effects and to formulate hypotheses, was introduced by Akritas and Arnold (1994), and further developed in the mixed effects model context by Akritas and Brunner (1997). These ideas were elaborated for repeated measures design by Brunner et al. (1999). Motivated by the difficulty to interpreting alternatives when hypotheses are formulated in terms of distribution functions, Konietschke et al. (2012a) developed methods for repeated measures in one group for hypotheses formulated in terms of nonparmetric relative effects. These methods allow construction of confidence intervals and address the nonparametric Behrens-Fisher problem (Brunner and Munzel, 2000) in the sense that the distribution of the data in the various groups could still be different under the null hypothesis.

Methods in which the distribution functions governing the data are used to describe *effects* and to formulate hypotheses have several merits. For instance, they may be utilized for any data types (including non-numeric data), they are robust with respect to outliers and the results are invariant under strictly monotone transformations of the data (Brunner et al., 2002). Brunner and Puri (1996) introduce nonparametric ranking methods for testing the null hypothesis $H_o^F : F_1 = F_2$ in terms of marginal distribution functions. The parametric models of ANOVA were extended to a general framework by Brunner and Puri (2001), which includes arbitrary data types and also does not require the distribution functions to be continuous.

In nonparametric marginal models, the independence structure of the observations from the design is used to determine which marginal distributions of the observed random vectors are identical. They can also be used to see change over time (Brunner et al., 2002). Nonparametric hypothesis testing based on marginal distributions is not limited to measurements repeated over time, but can also be applied to spatial observations (simultaneous measurements on variables of interest at different locations).

Quite generally, in parametric or semi-parametric marginal models expectations or shift parameters of the marginal distributions are used to define effects. In fact, an effect may not be captured solely by a shift parameter. Thus a function which yields a summary of effect is needed. This is one of the strategies indicated by Diggle et al. (2013) to analyze longitudinal data. Consequently, sometimes it might be of interest to see “total differences" over time rather than differences between two time points only. In such a situation a function of repeated measurements at t time points $(X_{k1}, ..., X_{kt})$, for instance, functionals of the distribution function of the variables can be considered to quantify the total difference. As an option for evalutating longitudinal data with multiple time points, Brunner et al. (2002) suggest defining reasonable summary variables – either for a range of some time points or for the total time.

When data are collected over a period of time on subjects, effective analysis can be performed in a couple of steps: first summarizing the data for each subject into a single number like the mean, slope, Area Under Curve (AUC), etc. and then analyzing the summary scores (Matthews, 1993). It is not uncommon to see summary variables for quantitative data. For instance, the use of a summary statistic called AUC or linear combinations of observations as summary variables is limited to metric data only. There is an excellent discussion of summary variables by Brunner et al. (2002) which can work for any data type. Summary variables are widely applicable in the analysis of repeated measures. In medical research, sometimes the use of summary variables is the only applicable tool for the analysis of repeated measurements (Vossoughi et al., 2012). The simulation study by Vossoughi et al. (2012) shows that the summary measure approach dominated the traditional unstructured multivariate approach and its performance was close to the best-fitting linear mixed model in testing all the effects.

Modern nonparametric methods forego the parametric or semi-parametric assumptions by formulating hypotheses in terms of distribution functions or some suitable functionals of the distribution functions. These functionals, referred to as *nonparametric relative (summary) effects*, are also used to quantify the magnitude of the effects of interest (Brunner et al., 2002). The use of empirical distribution functions to estimate these functionals leads to rank-based methods in a natural way.

In this study we make use of the normalized distribution function introduced below as a summary statistic to define what we call relative summary effects of a treatment at different time points.

The usual nonparametric approaches assume that the data are continuous and the distributions are of the same shape. The assumption of continuous distribution functions can be dropped by using the so-called *normalized* version of the distribution function (Brunner et al., 2002). Suppose X is a random variable. Then, we define the left and right continuous distribution functions as

$$F^{-}(x) := P(X < x) \qquad \text{(left continuous)},$$

$$F^{+}(x) := P(X \leq x) \qquad \text{(right continuous)}.$$

The *normalized* distribution function that combines these two versions is given by

$$F(x) := \frac{1}{2}\{F^{-}(x) + F^{+}(x)\}. \tag{1.2}$$

This way of defining a distribution function allows a unified treatment of ties. The approach also enables the analysis of discrete data and in particular of ordered categorical data.

1.3 Missing Data Analysis

In the real world, missing data are ubiquitous. No matter how well our experiments are planned there will often be times when something goes wrong, resulting in missing data. Some statistical procedures will not work as well, or at all, with some data missing. The best recourse is always to repeat the experiment to generate the complete data set. Sometimes, however, this is not feasible, particularly where readings are taken at set times or the cost of repeating the experiment is prohibitive. Hence, alternative ways of addressing this problem are needed. Different procedures to handle missing data situations in a variety of ways like, using Expectation-Maximization (EM), Multiple Imputation (MI), last observation carried forward (LOCF), case-wise deletion, pairwise deletion, and so on are available. The simplest approach in dealing with missing data is known as complete-case analysis, in which any subject with one or more observations missing is excluded from the analysis. That is, only data from subjects who have no missing data are included in the analysis. Using this method with informative missing data will result in noticeably biased results. In other words, analyzing only complete observations alone may give rise to biased results, especially if the reason for dropout is related to outcome. Apart from EM, LOCF and MI most of these procedures are not preferred since they throw away information contained in the incomplete subjects.

The relation of the missingness mechanism to the underlying values of the variables in the data set is the primary issue in handling missing data, since the properties of statistical methods for missing data depend immensely on the nature of the dependencies in these mechanisms (Little and Rubin, 2002). Moreover, all available imputation methods make use of all observed data, but estimation of variance of imputed values is an additional issue that needs to be handled (Konietschke et al., 2012a). For brevity, in this study we are discussing missing responses. This doesn't include missing covariates. For instance, a response vector is available, but group membership is missing.

If missingness does not depend on the values of the data, missing, observed or the covariates, then it is called Missing Completely at Random (MCAR) and it is Missing at Random (MAR) if the missingness depends only on the components of data that are observed.

Methods that analyze all observed data without imputing the missing values or excluding subjects with missing data have been developed. One of the methods which uses all available information includes the use of Generalized Estimating Equations (GEE), known also as marginal models, which extends Generalized Linear Model (GLM) theory to correlated data (Diggle et al., 2013). These methods do not require any distributional assumptions be made about the outcome variable but rely on quasi-likelihood methods in order to estimate parameters and test hypotheses. Instead of making assumptions about the distribution of the outcome variable, a correlation structure must be specified.

Rank-based procedures such as the Mann-Whitney-Wilcoxon rank sum test, the signed rank test, and the Kruskal-Wallis test are very common methods to analyze complete datasets when there is a gross violation of assumptions of their parametric counterparts. A lot of research has been conducted for complete data analysis. For instance, Konietschke et al. (2010) developed a fully nonparametric repeated measures model, where non-metric data are covered by this model in a unified way. Furthermore, their study covers the so-called nonparametric Behrens-Fisher problem in any type, not limited to longitudinal data, of repeated measures designs. Akritas et al. (2002) extended the rank-based nonparametric procedure for matched-pairs by Conover and Iman (1981) to data with missing observations.

Paired samples t-test typically deal with a sample of matched pairs of similar subjects or one group of units that have been measured twice and thus considered as the simplest case of repeated measures. When measurements before and after treatment are of the same size, i.e., subjects are measured at both instances, the parametric paired t-test is the most appropriate procedure for analyzing the data. Different means of handling dataset with missing values have been developed. For example, Xu and Harrar (2012) developed a parametric procedure, which uses all available information, for comparison of means in partially paired data under the assumption of MCAR. A parametric statistical procedure based on permuting incomplete paired data by Amro and Pauly (2017) is shown to be asymptotically correct and also finitely exact when the distribution of the data is invariant with respect to the randomization group under consideration.

Repeated measures methods can be biased due to missing observations, especially if missing observation are related to treatment being studied. Modeling the dropout process requires assumptions and sophisticated modeling methods. The traditional repeated measures ANOVA encounter a problem when there are missing data on the response variable. If one measurement is missing, the entire case gets dropped as it uses listwise deletion. In the presence of missing observations in a dataset, it might be appealing to apply complete pairs analysis under the assumption of MCAR, but this approach fails to use all available information.

Although several test procedures are developed as solutions to the problem of missing data, most of them are designed under parametric setup. For instance, the test procedure by Xu and Harrar (2012) for mean comparison in partially matched pairs is valid in parametric setup only. Among several studies conducted on the development of nonparametric test procedure, those by Brunner and Puri (1996), Gao (2007) and Konietschke et al. (2012a) work under the MCAR missing mechanism. The test procedures developed by Akritas et al. (2002) work under MAR missing mechanism. In this study the missing mechanism MCAR in nonparametric setup is considered. For details of missing data analysis in general the reader can refer to Little and Rubin (2002). In the next section we discuss some of the nonparametric methods for handling missing data in repeated measures design.

1.4 Objective and Organization of the Study

In order to describe data, it is common to assume a specific probability model. Unfortunately, in many practical applications, it is not possible to identify a specific structure for the data. Nonparametric methods provide statistical tools for making inference under less stringent assumptions than the classical parametric procedures. Obviously, the parametric test procedures are powerful when the underlying distributional assumptions are met. This research was motivated by a repeated measures study in which measurements were collected on the same subject over a period of time. Measurements on the same subject are correlated and therefore any subject effect must be accounted for even though differences between subjects are not of interest. We are interested in developing rank-based procedures in determining if the outcome changed over time when some subjects have incomplete data.

Konietschke et al. (2012a) developed a fully nonparametric test procedure, under MCAR frame work, for testing significance of treatment effects when the matched pairs are only partially complete. This piece of work is very useful in a sense that it can be applied to any data type and also use all available information. However, the method is limited to one group. In the current study we have two main objectives. First, we extend the work of Konietschke et al. (2012a) to the two-group case and focus on developing a test procedure for testing the significance of group-time interaction effect and also develop a confidence interval for the interaction effect size.

A fully nonparametric rank-based test procedure is developed by Konietschke et al. (2010) to study a significance of change in effects over time. The study is also important in a sense that it handles any data type and is also robust to the so-called nonparametric Behrens-Fisher problem, in repeated measures designs. This method goes off the track when there is a missing observation. The second objective of this research work is extension of the result by Konietschke et al. (2010), to data with missing observations. We develop a fully nonparametric rank-based test procedure for testing the effect of time/treatment. The new test procedures work under the MCAR framework and are also robust to the Behrens-Fisher problem. The first method is limited to two-time points and two groups whereas the second method is limited to three-time points as covariance estimation in this method is a bit complex.

The dissertation is organized as follows. Literature review covering nonparametric methods for repeated measures and missing data is detailed in Chapter 2, specifically the rank-based methods; a general model for nonparametric repeated measures design is outlined in Section 2.1. For specific designs of this general model, namely one group two-time points, one group repeated measures and two-time points with two-groups; the model, effect size measure, interpretation of the effect size, estimation of the effect size vector and asymptotic properties are described in detail in Sections 2.1.1, 2.1.2, 2.1.3, respectively. In Section 2.2 we discuss nonparametric repeated measures designs with missing data. As an example, a discussion of a design with one group two-time points with missing data is presented in detail. To accomplish the aforementioned objectives the way forward is outlined in Section 2.3 of the chapter.

The discussion in Chapter 3 pertains to the analysis of partially matched pairs. This includes the statistical model for the nonparametric interaction effect size, rank-based expression of the estimator for the nonparametric interaction effect, the asymptotic behavior of the proposed estimator including hypothesis testing and confidence interval construction for the interaction effect size. An overview of the model and interaction-effect size measure are described in Sections 3.1 and 3.2. The asymptotic theory for the nonparametric method, which constitutes the main contribution of the chapter, is presented in Section 3.3. The application of the theory for carrying out significance test and constructing a confidence interval needs derivation of consistent estimator for the asymptotic variance. This is done in Section 3.4. The applications of the asymptotic theory for tests and confidence intervals on nonparametric effect are outlined in Section 3.5. More importantly, Section 3.5 proposes a small sample test and a confidence interval based on the nonparametric theory developed. Simulation studies are carried out in Section 3.6 to evaluate the numerical accuracy of the asymptotic results and finite sample approximations. Another aim of Section 3.6 is to numerically compare the performance of the new nonparametric test and its finite sample approximation with competing parametric methods in terms of size and power of tests. We close Section 3.6 by analyzing a real dataset from public health to illustrate the application of the methods.

The discussion in Chapter 4 pertains to the analysis of repeated measures designs with missing observations in the nonparametric setup. The setup of statistical model and effect size measure are described in Section 4.2. The definition of effect sizes, their rank-based estimators and asymptotic properties of the estimated vector of relative effects are detailed in Section 4.3. Estimation of covariance matrix based on the the idea of asymptotic rank transform is explained in Section 4.4. Test procedures based on WTS and ATS, MCTP statistics and confidence intervals for relative effects are discussed in Section 4.5. In order to study the numerical accuracy of the new test procedure, a size simulation is carried out and results detailed in Section 4.6. Power simulation was conducted to study the performance of the proposed test statistic and the results are presented in Subsection 4.6.2. Section 4.7 discusses the results of real data analysis in detail. Lastly, discussion, conclusions from both Chapters 3 and 4 and future directions are presented in Chapter 5.

Chapter 2

Nonparametric Methods for Repeated Measures and Missing Data

In this chapter we review recent nonparametric methods for the analysis of repeated measures data. Later in the chapter, we give an account of extensions to accommodate missing data. These methods form the basis for the research undertaken in this dissertation. In particular, this research seeks to develop methods for handling missing data in repeated measures studies when the missingness is MCAR. That is, the focus of this research will be on developing methods that preserve a preset type-I error rate while retaining all available observations and, thus, minimizing the loss of power due to missing data. Asymptotic theoretical results will be derived and the finite sample performance of these results will be investigated using simulated data.

2.1 General Nonparametric Analysis of Repeated Measures

In this section relevant literature on repeated measures in the nonparametric setup will be reviewed. Most of the materials presented are largely adapted from Konietschke et al. (2018) and Brunner et al. (2017).

First, we present a general nonparametric model in repeated measures factorial design and then consider specific designs (special cases of the general repeated measures design) in the following subsections. A general model that covers higher way layouts with repeated measures or longitudinal data is represented by the following model. Suppose that $\boldsymbol{X}_{gk}$ denotes independent random vectors containing l repeated measurements observed on k^{th} subject in g^{th} group. Then,

$$\boldsymbol{X}_{gk} = (X_{gtk})_{t=1}^{l} = (X_{g1k}, ..., X_{glk})', \tag{2.1}$$

where $g = 1, ..., d$, $t = 1, ..., l$, and $k = 1, ..., n_g$. All of the test procedures discussed in this and subsequent chapters are confined to those developed in terms of relative effects. Thus we first give a definition of relative effects.

Let F_{gt} denote the marginal distribution of X_{gtk}, i.e., $X_{gtk} \sim F_{gt}$. We define the marginal normalized distribution function, $F_{gt}(x)$, of X_{gtk} as defined in Ruymgaart (1980). For an observation in group g at occasion t, $F_{gt}(x)$ is defined as

$$F_{gt}(x) := \frac{1}{2}\left\{F_{gt}^{+}(x) + F_{gt}^{-}(x)\right\}, \tag{2.2}$$

where $F_{gt}^{-}(x) = P(X_{gtk} < x)$ and $F_{gt}^{+}(x) = P(X_{gtk} \leq x)$ are, respectively, the left and right continuous versions of the unknown distribution function.

The *relative summary effect*, p_{gt}, is defined as

$$p_{gt} = \int G dF_{gt} = P(X_{gtk} > Z), \tag{2.3}$$

where $Z \sim G$, G is unweighted pooled distribution function given by

$$G = \frac{1}{dl}\sum_{g=1}^{d}\sum_{t=1}^{l} F_{gt}. \tag{2.4}$$

This definition of G is independent of sample size allocation. The unweighted average distribution function, G, is preferred because the relative effects obtained via this definition of G depend only on the underlying distribution function (Brunner et al. (1999), Brunner et al. (2002), Gao and Alvo (2008), and Konietschke et al. (2018)).

Moreover, the *relative summary effect*, p_{gt}, can be rewritten as a function of *relative marginal effect (relative effect)*, w_{rsgt}, where

$$w_{rsgt} = \int F_{rs} dF_{gt}, \tag{2.5}$$

for $1 \leq g, r \leq d, 1 \leq s, t \leq l$. It is easy to see that the relative effect can be rewritten as

$$p_{gt} = \bar{w}_{..gt} = \frac{1}{dl}\sum_{r=1}^{d}\sum_{s=1}^{l} w_{rsgt},$$

where

$$w_{rsgt} = \int F_{rs} dF_{gt} = P(X_{rs} < X_{gt}) + \frac{1}{2}P(X_{rs} = X_{gt}). \tag{2.6}$$

Here w_{rsgt} is the probability that a randomly chosen observation from F_{rs} is less than or equal to an observation from F_{gt}. This way of defining the relative effect, w_{rsgt}, allows a unified treatment of ties which stemmed from the definition of distribution function, F_{gt}, in equation (2.2) (Fligner and Policello (1981), Brunner and Puri (1996), Munzel (1999), and Brunner and Munzel (2000)). The relative effect p_{gt} quantifies the tendency of the marginal distribution F_{gt} with respect to the mean distribution G. For instance, $p_{gt} < p_{g't'}$ means observations from F_{gt} have a tendency to be smaller than those from $F_{g't'}$ with respect to G. Similarly, a relative effect $p_{gt} < \frac{1}{2}$ means that randomly selected observation in the group *g* at time *t* tend to be smaller as compared to those of all groups at all times (Brunner et al., 2002).

Analogous to normalized distribution functions, the normalized empirical distribution function can be defined as

$$\hat{F}_{gt}(x) := \frac{1}{2}[\hat{F}^-_{gt}(x) + \hat{F}^+_{gt}(x)],$$

where

$$\hat{F}^-_{gt}(x) := \frac{1}{n_g}\sum_{k=1}^{n_g} c^-(x - X_{gtk}) \qquad \text{(left continuous)},$$

and

$$\hat{F}^+_{gt}(x) := \frac{1}{n_g}\sum_{k=1}^{n_g} c^+(x - X_{gtk}) \qquad \text{(right continuous)}.$$

The function $c(x)$ used in the above definitions is the normalized count function defined as

$$c(x) = \frac{1}{2}[c^-(x) + c^+(x)], \tag{2.7}$$

where $c^-(x)$ and $c^+(x)$ given by

$$c^-(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ 1, & \text{if x} > 0 \end{cases} \qquad \text{and} \qquad c^+(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}.$$

Let us assemble the relative effects in the *d* groups and *l* time points into a vector as

$$\boldsymbol{p} = (p_{11}, ..., p_{dl})'.$$

An estimator of this vector is given as

$$\hat{\boldsymbol{p}} = (\hat{p}_{11}, ..., \hat{p}_{dl})',$$

and the components are defined as follows

$$\hat{p}_{gt} = \int \hat{G} d\hat{F}_{gt},$$

where t = 1, ..., l and g = 1, ..., d, $\hat{F}_{gt}$ and $\hat{G}$ are empirical counterparts of distribution functions in equation 2.2.

Konietschke et al. (2018) show that the quantity $\sqrt{\boldsymbol{n}}(\hat{\boldsymbol{p}} - \boldsymbol{p})$ has an asymptotic multivariate normal distribution with mean **0** and an unknown covariance matrix $\boldsymbol{\Sigma}$. The expression of the covariance matrix $\boldsymbol{\Sigma}$ and its consistent estimator, $\hat{\boldsymbol{\Sigma}}$, are complicated and we refer the interested reader to Konietschke et al. (2018) for details.

The use of marginal means for the purpose of comparison may be questionable when the distribution of data is far from normality. To illustrate this phenomenon, consider four repeated measures $(l = 4)$ and one group $(d = 1)$. In Figure 2.1 below we illustrate the benefits of using *relative summary effects* instead of the marginal means when the assumption of normality is grossly violated. A sample of size n = 100 is used to generate data from multivariate distribution of dimension four from both Normal ($N_4((3,3,3,3)', \boldsymbol{\Sigma})$) and Non-normal distributions (Normal distribution 5% contaminated by Cauchy, $Cauchy((3,3,3,3)', \boldsymbol{\Sigma})$), where $\boldsymbol{\Sigma}$ contains unity on the diagonal and 0.5 on off-diagonal. Error bars for relative summary effect and the marginal means are computed for both distributions using 10,000 simulations. The error bars on the profile plot for the relative summary effects appear to be similar for both distributions whereas that of the marginal means are fluctuating. Therefore, consideration of summary relative effects for comparison effects seems to be appropriate.

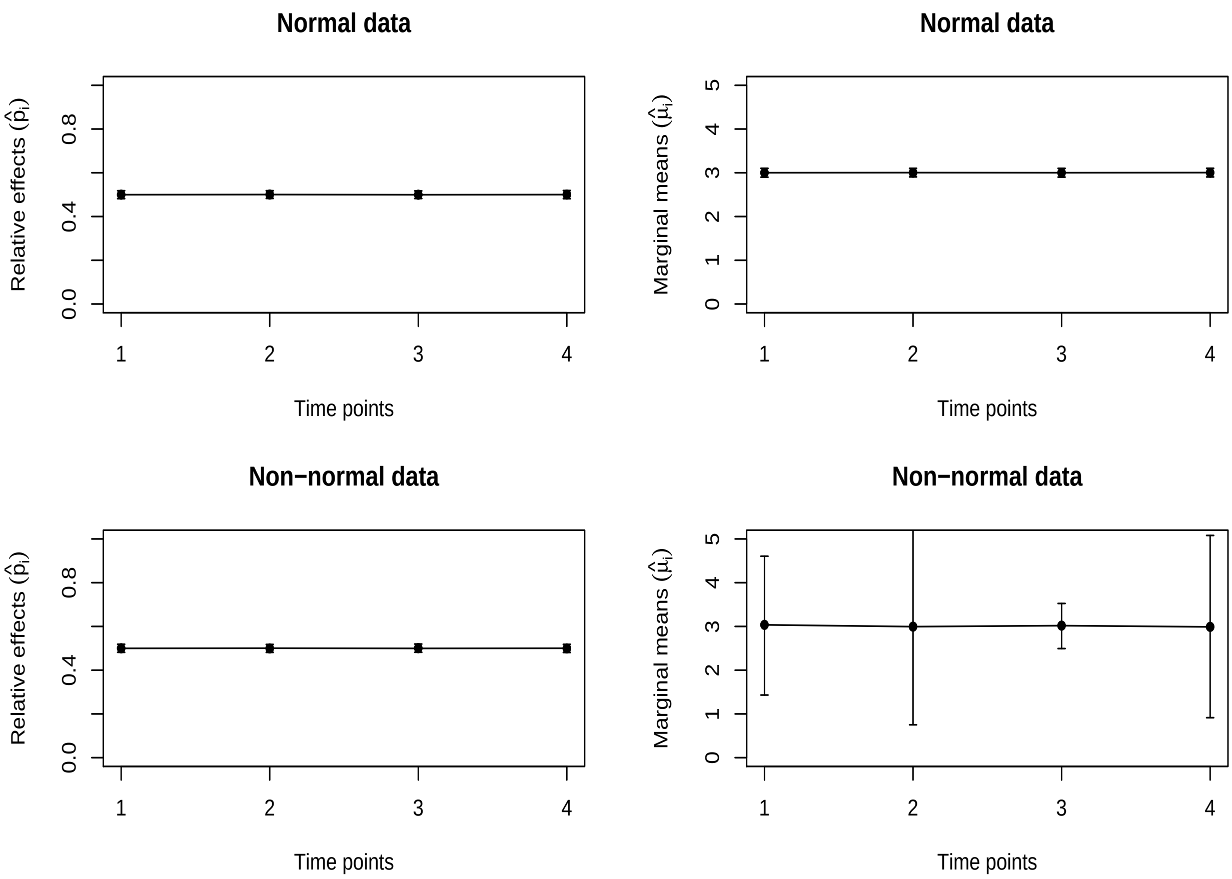


FIGURE 2.1: Profile plot for relative effects and marginal means from normal and non-normal (normal data 5% contaminated by Cauchy) multivariate data of dimension four. Sample sizes of 100 and simulation of size 10,000 is used for calculation of the error bars.

The null hypothesis of interest can be formulated in terms of either distribution functions or relative effects, by choosing the appropriate contrast matrix, **C**, as follows

$$H_o^F : \boldsymbol{CF} = \mathbf{0} \quad vs. \quad H_1^F : \boldsymbol{CF} \neq \mathbf{0}, \tag{2.8}$$

or

$$H_o^p : \boldsymbol{Cp} = \mathbf{0} \quad vs. \quad H_1^p : \boldsymbol{Cp} \neq \mathbf{0}, \tag{2.9}$$

where $\boldsymbol{p} = (p_{11}, ..., p_{dl})'$. Specification of null hypothesis through relative effects is less stricter than the one specified in terms of the distribution functions (Brunner et al., 2002). In the above hypotheses the null hypothesis (H_o^F) in the former (2.8) implies the the null hypothesis (H_o^p) in the later (2.9). In fact, the converse is not always true.

Test procedures for testing the hypothesis H_o^F are not consistent with respect to H_1^F. For instance, ranking procedures by Ryu and Agresti (2008), Ryu (2009), and Browne (2010) for testing H_o^F are not consistent with respect to alternatives of the form H_1^F (Konietschke et al., 2012a). They are not good to be used to construct confidence intervals for the effect size measure. Consequently, the test procedures which test the hypothesis, H_o^p, in terms of the treatment relative effect are more meaningful, and more importantly, easier to interpret (Brunner et al., 2002; Konietschke et al., 2012a; Brunner et al., 2017).

The Wald-Type Statistic (WTS) can be used to test the null hypothesis in equation (2.9) and is given by

$$WTS = \boldsymbol{N}\hat{\boldsymbol{p}}'\boldsymbol{C}'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{+}\boldsymbol{C}\hat{\boldsymbol{p}}, \tag{2.10}$$

where N is all available observations and $(A)^{+}$ denotes More-Penrose inverse of matrix A. The above test statistic follows $\chi^2_{(r(\boldsymbol{C}))}$ asymptotically, provided that $\boldsymbol{C}$ is of full row rank $r(\boldsymbol{C})$. Therefore, the above null hypothesis in equation (2.9) is rejected if the calculated WTS is greater than the $(1-\alpha)^{th}$ quantile of $\chi^2_{(r(\boldsymbol{C}))}$ distribution.

Alternatively, the same null hypothesis in equation (2.9) can also be tested by ANOVA-Type statistic given as

$$ATS = \frac{N}{tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}})}\hat{\boldsymbol{p}}'\boldsymbol{M}\hat{\boldsymbol{p}}, \tag{2.11}$$

where $\boldsymbol{M} = \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{+}\boldsymbol{C}$ denote the orthogonal projection onto the range of $\boldsymbol{C}'$. Under the null hypothesis, ATS has approximately χ^2_{ν}/ν distribution with approximated degrees of freedom $\hat{\nu} = \frac{[tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}})]^2}{tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}}\boldsymbol{M}\hat{\boldsymbol{\Sigma}})}$ (Brunner et al., 1997).

In order to accommodate higher way layouts with repeated measures or longitudinal data by the general model defined in equation (2.1), a factorial structure on the groups or repeated measures is easily obtained from the general setup by splitting the indices g or t into subindices g', g'', ..., or t', t'', ... respectively.

For example, consider a two-way layout with two crossed factors A and B, with levels g = *1, 2, ..., a* and t = *1, 2, ..., b*, respectively. Let F_{gt} denote the marginal distribution of X_{gtk}, i.e., $X_{gtk} \sim F_{gt}$ for $k = 1, ..., n_{gt}$. Then using the unweighted average distribution function in equation (2.4), we have the nonparametric relative effects, p_{gt}, collected in the vector $\boldsymbol{p} = (p_{11}, p_{12}, p_{21}, ..., p_{ab})'$. Note that if $p_{gt} \leq p_{rs}$ the observations under factor combination *(g, t)* tend to result in smaller values than the observations under factor combination *(r, s)*. The hypothesis of no effect of factor A, no effect of factor B and no interaction effect in terms of relative effects, can be formulated as

$$H_o^p : \boldsymbol{C}\boldsymbol{p} = \boldsymbol{0},$$

where the contrast matrices for the three hypotheses are $\boldsymbol{C} = \boldsymbol{P}_a \otimes \frac{1}{a}\boldsymbol{J}_a$, $\boldsymbol{C} = \frac{1}{2}\boldsymbol{J}_b \otimes \boldsymbol{P}_b$ and $\boldsymbol{C} = \boldsymbol{P}_a \otimes \boldsymbol{P}_b$, respectively. In the above notations $\boldsymbol{P}_a = \boldsymbol{I}_a - \frac{1}{a}\boldsymbol{J}_a$ is a centering matrix, I_a is a-dimensional identity matrix, and J_a is an a-dimensional square matrix where all entries are unity. The test statistics, WTS and ATS, mentioned above can be used to test the above three hypotheses. In the next subsections specific designs of the aforementioned general setup which are the basis for the subsequent chapters are presented. For the sake of brevity, we limit our discussion to the designs: two time points in one group $(d = 1, l = 2)$, two-time point in two-groups $(d = 2, l = 2)$ and one group with l time points $(d = 1)$. Our aim is to extend the latter two to missing data situation.

2.1.1 One group two-time points

When we consider one group with two time points the general model discussed in Section 2.3 reduces to the simplest design, which is commonly known as *matched pairs design*. The general model for repeated measures in equation (2.1) reduces to

$$\boldsymbol{X_{tk}} = (X_{1k}, X_{2k})',$$

where $k = 1, ..., n$. The random vectors $\boldsymbol{X_k}$ are assumed to be independent and the X_{tk} have the marginal distribution F_t, for $t = 1, 2$. Using this marginal distribution function the *relative summary effect* defined in equation (2.3) simplifies to

$$p_t = \int G dF_t,$$

where $G = \frac{1}{2}\sum_{t=1}^{2} F_t$.

The idea of hypothesis formulation in terms of relative summary effects for the specific design under consideration (the case of two-time points, $l = 2$) is explained as follows. Define p, as

$$p := \int F_1 dF_2.$$

Now,

$$p_1 = \int G dF_1 = \int \frac{1}{2}(F_1 + F_2)dF_1 = \frac{1}{4} + \frac{1}{2}\int F_2 dF_1 = \frac{1}{4} + \frac{1}{2}(1 - \int F_1 dF_2) = \frac{1}{4} + \frac{1}{2} - \frac{1}{2}p.$$

Similarly, for the relative summary effect at the second time point , p_2, we have

$$p_2 = \int G dF_2 = \frac{1}{2}\int F_1 dF_2 + \frac{1}{2}\int F_2 dF_2 = \frac{1}{4} + \frac{1}{2}p.$$

So, taking the difference of the above two quantities, p_2 and p_1, yields

$$p_2 - p_1 = p - \frac{1}{2}. \tag{2.12}$$

Notice that if $F_1 = F_2 = G$, then $p_2 = p_1$, implying that $p - \frac{1}{2} = 0$. The converse is not necessarily true.

In the subsequent discussion, we will utilize relative treatment effects to perform tests. For this purpose we introduce the notion of tendentiously smaller (larger) variables. Suppose the two random variables with respective marginal distribution function, say, $X_1 \sim F_1$ and $X_2 \sim F_2$. It can be seen that

$$p = \int F_1 dF_2 = P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2), \tag{2.13}$$

that is, the quantity p is the probability that X_1 is less than X_2.

If $p < \frac{1}{2}$, then we say that X_2 is *tendentiously smaller* than X_1, and if $p > \frac{1}{2}$, then we say that X_2 is *tendentiously larger* than X_1. If $p = \frac{1}{2}$, no tendency exists for the values of X_2 to be either larger or smaller than those of X_1.

For the choice of an appropriate contrast matrix, inthis case say **C** = (1, -1)′, the general hypothesis of no treatment effect in terms of relative effects in equation (2.9) simplifies to

$$H_o^p : p_1 - p_2 = 0$$

and is equivalent to

$$H_o^p : p = \frac{1}{2}.$$

Estimators for Relative Effects

The relative effect p_t is estimated by replacing the distribution functions $G(x)$ and $F(x)$ with their respective empirical distribution functions. Once the empirical distribution functions are estimated, $G(x)$ can be estimated as

$$\hat{G}(x) = \frac{1}{2}\sum_{t=1}^{2}\hat{F}_t(x) = \frac{1}{N}\sum_{t=1}^{2}\sum_{k=1}^{n}c(x - X_{tk}),$$

where $N = 2n$.

Consequently, the relative effect, p_t is estimated by

$$\hat{p}_t = \int \hat{G}d\hat{F}_t(x) = \frac{1}{n}\sum_{k=1}^{n}\hat{G}(X_{tk}).$$

The relationship between $\hat{G}(X_{tk})$ and the rank of X_{tk}, which is called Rank Transform (RT), makes computation of estimated relative effects very easy (Conover and Iman, 1981). In particular, the two-time point special case allows expression of relative effects and their estimated variances in terms of ranks. Based on the three versions of empirical distribution function it is possible to establish the following relationship between $\hat{G}(X_{tk})$ and the rank of X_{tk} among all $N = 2n$ observations (Brunner et al., 2002). In the following $\hat{G}^{-}$ and $\hat{G}^{+}$ are averages of empirical versions of left and right continuous distribution functions respectively.

$$R^{-}_{tk} = 1 + N\hat{G}^{-}(X_{tk}), \qquad \text{(minimum rank)},$$

$$R^{+}_{tk} = N\hat{G}^{+}(X_{tk}), \qquad \text{(maximum rank)},$$

and

$$R_{tk} = \frac{1}{2} + N\hat{G}(X_{tk}) \qquad \text{(mid-rank)}.$$

This relationship may not hold in general for $d > 1$. From the relationship between $\hat{G}(X_{tk})$ and the rank of X_{tk}, one can obtain the following result

$$\hat{p}_t = \int \hat{G}d\hat{F}_t = \frac{1}{n}\sum_{k=1}^{n}\hat{G}(X_{tk}) = \frac{1}{N}(\bar{R}_{t.} - \frac{1}{2}).$$

So, the empirical estimator of p_t in terms of ranks is given by

$$\hat{p}_t = \frac{1}{N}(\bar{R}_{t.} - \frac{1}{2}).$$

Brunner et al. (2002) show that $\hat{p}_t$ is a consistent and asymptotically unbiased estimator of p_t. Since the relative effects are expressed in terms of ranks, it is computationally easy to obtain estimates for p_t and construct a confidence interval for the same. For two-time points the vector $\boldsymbol{p}$ is estimated by $\hat{\boldsymbol{p}} = (\hat{p}_1, \hat{p}_2)$. Choosing the contrast matrix $\boldsymbol{C} = (-1, 1)$, the quantity $\sqrt{n}\boldsymbol{C}\hat{\boldsymbol{p}}$ is simplified as

$$\sqrt{n}\boldsymbol{C}\hat{\boldsymbol{p}} = \sqrt{n}\boldsymbol{C}(\hat{p}_1, \hat{p}_2) = \sqrt{n}\boldsymbol{C}\frac{1}{2n}(\bar{R}_{1.} - \frac{1}{2}, \bar{R}_{2.} - \frac{1}{2}) = \frac{1}{2\sqrt{n}}(\bar{R}_{2.} - \bar{R}_{1.}).$$

Under the null hypothesis, H_o^p, the WTS statistic in equation (2.10) for large sample size reduces to

$$T = \sqrt{n}\frac{(\bar{R}_{2.} - \bar{R}_{1.})}{2S_n}.$$

Under H_o^p, T converges in distribution to N(0,1). In the above case $\bar{R}_{t.}$ is average of overall rank of X_{tk} and

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left([R_{2k} - R_{2k}^{(2)}] - [R_{1k} - R_{1k}^{(1)}] - [\bar{R}_{2.} - \bar{R}_{1.}]\right)^2.$$

For small sample sizes $(n \leq 15)$, T can be approximated by the *t(n-1)*-distribution (Brunner et al., 2002). Since $\boldsymbol{C}$ is of full row rank, the distribution of T^2 can also be approximated by $\chi^2_{(1)}$.

2.1.2 One group repeated measures

Again we consider the general model for repeated measures in equation (2.1), which for one group and multiple time points, reduces to the following independent and identically distributed random vectors

$$\boldsymbol{X_k} = (X_{1k}, ..., X_{lk})', \tag{2.14}$$

where $k = 1, ..., n$. Note that X_{tk} has the marginal distribution function F_t. The model above in equation (2.14) has been studied in detail by Konietschke et al. (2010). Therefore, here we highlight the results from that paper.

Consider the relative effect given in equation (2.1.1) which can be rewritten as

$$p_t = \int G dF_t = \frac{1}{l}\sum_{j=1}^{l}\int F_j dF_t,$$

where $G = \frac{1}{l}\sum\limits_{t=1}^{l} F_t$.

Now collecting p_t s in a vector $\boldsymbol{p} = (p_1, ..., p_l)'$, its estimator is $\hat{\boldsymbol{p}} = (\hat{p}_1, ..., \hat{p}_l)'$, where $\hat{p}_t$ are estimated by

$$\hat{p}_t = \frac{1}{l} \sum_{j=1}^{l} \hat{p}_{jt},$$

with

$$\hat{p}_{jt} = \int \hat{F}_j d\hat{F}_t = \frac{1}{N} \left(\bar{R}_t^{(jt)} - \frac{(n+1)}{2} \right).$$

and $\bar{R}_t^{(jt)}$ is the average of overall rank of X_{tk} for time points $\{j, t\}$.

Konietschke et al. (2010) show that $\hat{\boldsymbol{p}}$ is asymptotically unbiased and is a consistent estimator of $\boldsymbol{p}$. They also show that $\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p})$ has asymptotically a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, and also obtained a consistent rank-based estimator for this covariance matrix.

The nonparametric null hypotheses of no treatment effect, in terms of relative effects, p_t, can then be stated as

$$H_o^p : \boldsymbol{C}\boldsymbol{p} = 0,$$

where $\boldsymbol{p} = (p_1, ..., p_l)'$. For example, for $l = 3$, the following contrast matrix, $\boldsymbol{C}$, can be considered

$$\boldsymbol{C} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

to test the null hypothesis $H_o : p_1 = p_2 = p_3$.

The two common test statistics WTS and ATS can be used to test the above overall null hypothesis. Moreover, Konietschke et al. (2010) introduced the MCTP approach to test pairwise hypotheses simultaneously. The inferential procedures discussed in that paper are developed for complete data. They are not appropriate to use when the data set contains missing values. Consequently, the need for a test procedure which accommodates missing value situation arises. It is our objective to develop this type of test procedure and is the subject of Chapter 4 of the dissertation.

2.1.3 Two-time points with two-groups

The general model for repeated measures in equation (2.1), for two-groups and two-time points, reduces to a 2×2 experimental design type for longitudinal data, with the following random vectors

$$\boldsymbol{X_{gk}} = (X_{g1k}, X_{g2k})', \tag{2.15}$$

where the observations X_{gtk} have the marginal distributions F_{gt}, for $g = 1, 2$, $t = 1, 2$. Here also the relative summary effects can be described using the relative marginal effects as

$$p_{gt} = \int G dF_{gt},$$

where $t, g = 1, 2$ and $G = \frac{1}{4} \sum_{t=1}^{2} \sum_{g=1}^{2} F_{gt}$. The estimator of p_{gt} can easily be obtained by replacing F_{gt} by their empirical counterparts. Consequently,

$$\hat{p}_{gt} = \int \hat{G} d\hat{F}_{gt}.$$

The asymptotically unbiased and a consistent estimator for the relative summary effect, p_{gt}, in terms of rank is give by

$$\hat{p}_{gt} = \frac{1}{N}(\bar{R}_{gt.} - \frac{1}{2}),$$

where $\bar{R}_{gt.}$ is the average of overall ranks of X_{gtk} among N observations.

The null hypothesis for testing group-time interaction effect in terms of marginal distribution function is formulated from the general hypotheses in equation (2.9) as $H_o^p : \boldsymbol{Cp} = 0$, for the choice of contrast matrix **C** = (-1, 1, 1, -1)′, where $\boldsymbol{p} = (p_{11}, p_{12}, p_{21}, p_{22})'$ and this null hypothesis can be tested by test statistic, expressed in terms of rank as

$$T = \frac{(\bar{R}_{12.} - \bar{R}_{11.} + \bar{R}_{21.} - \bar{R}_{22.})}{\sqrt{\sum_{k=1}^{2} \hat{S}_g{}^2 / n_g}},$$

where

$$\hat{S}_g^2 = \frac{1}{n_g - 1} \sum_{k=1}^{n_g} (R_{g2k} - R_{g1k} + \bar{R}_{g1.} - \bar{R}_{g2.})^2,$$

for g = 1, 2.

Under the null hypothesis, H_o^p, the statistic T has, asymptotically, a standard normal distribution. For small sample sizes $t(\hat{\nu})$-distribution is used for the same testing purpose, where $\hat{\nu}$ is approximated by Satterthwaite approach (Brunner et al., 2002). The above inferential procedure discussed in detail in (Brunner et al., 2002) is developed for complete data. Complete case analyses are common for analyzing data with missingness, especially under MCAR mechanism. They are appropriate in the sense that they maintain the level of the test. However, they may be less powerful. Hence, it is not appropriate to use it when the data set contains missing values. Consequently, the need for a test procedure which accommodates missing values in this setup arises. We are interested to develop this type of test procedure and that is the discussion point of Chapter 3 of the dissertation.

2.2 Nonparametric Repeated Measures with Missing Data

When the compulsory assumption for the t-test is not met, nonparametric counterparts, like the sign-test and Wilcoxon signed-rank test, are used provided that the non-stringent assumptions are satisfied (Hollander et al., 2015). However, the aforementioned procedures are limited to complete cases. When the subjects are partially paired, i.e., when both paired observations and independent observations are present in the two sample design, assuming data are missing completely at random (MCAR), the paired observations or independent observations may be discarded in order to proceed with the standard tests above (Derrick et al., 2017). But discarding information is not a good idea; rather making use of all available data is recommended.

In fact, there are some methodological developments in the direction of accommodating subjects with missing observations in data analysis. For instance, a fully nonparametric method for analysis of factorial designs developed by Akritas et al. (2006) works under MAR. Brunner et al. (1999) developed procedures which work in the absence of the continuity assumption, under scenarios of MCAR and singular covariance matrices. In their paper, Brunner et al., the null hypothesis is formulated in terms of distribution functions. The methods developed by Xu and Harrar (2012) are a viable tool for handling partially paired data but requires the existence of the eighth moment of the sampling population. Akritas et al. (2002) propose a ranking approach for $H_o^F : F_1 = F_2$ which is an extension of the matched pairs test procedure by Conover and Iman (1981) to data with missing observations, under the assumption of missing at random (MAR). A nonparametric approach for testing H_o^p under MCAR data, applicable to arbitrary distributions was derived by Brunner and Puri (1996). But the procedure does not maintain the pre-assigned Type I Error level for small sample sizes or in unbalanced sample size allocations.

Konietschke et al. (2012a) extended the idea of weighted rank estimators for relative effects in factorial diagnostic trials with clustered data by Konietschke and Brunner (2009) to MCAR data. They also derived asymptotic test procedures for H_0^p and constructed asymptotic confidence interval for p. The asymptotic framework considered is either for the complete cases or when both the incomplete and complete cases are large. Their approach is purely nonparametric and allows one to formulate hypotheses in terms of the effect size measure p. Moreover, the marginal distribution functions are not required to have the same shape. Hence, this enables to cover the so-called nonparametric Behrens–Fisher problem for matched pairs with missing values. Below we provide some details from that paper.

One group two-time points with missing data

Consider paired observations from n subjects; $n = n_c + n_1 + n_2$, where n_c, n_1 and n_2 are the number of complete cases, incomplete data in time 1, and incomplete data in time 2, respectively. The complete data vector is

$$\boldsymbol{X}_k^{(c)} = (X_{1k}^{(c)}, X_{2k}^{(c)}), k = 1, \ldots, n_c,$$

with marginal distributions

$$X_{tk}^{(c)} \sim F_t,$$

t = 1, 2 and $k = 1, \ldots, n_c$. Here F_t is the normalized distribution function defined in equation (1.2).

Let X_k be a random vector defined as

$$X_k = (\Delta_{1k} X_{1k}, \Delta_{2k} X_{2k}), \tag{2.16}$$

where k=1,. . . , n – paired observations. Define Δ_{ik}, a missing indicator, as

$$\Delta_{ik} = \begin{cases} 1, & \text{if } X_{ik} \text{ is observed;} \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$n_c = \sum_{k=1}^{n} \Delta_{1k} \Delta_{2k}$$

and

$$n_t = \sum_{k=1}^{n} \Delta_{tk}(1 - \Delta_{jk}),$$

for $j \neq t$, are the number of complete and incomplete cases in sample t; j, t = 1, 2, respectively.

Under the MCAR assumption, the incomplete cases,

$$X_{tk}^{(i)} \sim F_t,$$

k = 1, . . . , n_t , where n_t denotes the number of incomplete data in sample *t; t = 1, 2*.

Most of the test procedures, for partially paired design, mentioned before assume one or more combination of the following sufficient conditions.

1. $n_c \to \infty, n_1, n_2 \leq M < \infty$, or
2. $n_c \to \infty, n_1 \to \infty, n_2 \leq M < \infty$, or
3. $n_c \to \infty, n_1 \to \infty, n_2 \to \infty$, or
4. $n_c \leq N_c, n_1 \to \infty, n_2 \to \infty$
5. $min\{n_c, n_t\} \to \infty$
6. $min\{n_c + n_1, n_c + n_2\} \to \infty$.

It is not realistic to assume that the number of incomplete cases or missing values goes to infinity. The procedure developed by Konietschke et al. (2012a) works under a more realistic assumption in a sense that it assumes that the sum of number of complete and incomplete cases at both time points is large, i.e. $n_c + n_t \to \infty$, such that $\frac{n}{n_c+n_t} \leq N_o < \infty, t = 1, 2$. Under this assumption, they show that an asymptotically unbiased and consistent estimator of the relative effect size, *p*, defined in equation (2.13) is given by

$$\hat{p}_\theta = \int \hat{F}_{1,\theta_1} d\hat{F}_{2,\theta_2} = \frac{1}{N}\left(\theta_2 \bar{R}_{2.}^{(c)} - \theta_1 \bar{R}_{1.}^{(c)} + (1-\theta_2)\bar{R}_{2.}^{(i)} - (1-\theta_1)\bar{R}_{1.}^{(i)}\right) + \frac{1}{2}, \tag{2.17}$$

where N = $2n_c+n_1+n_2$ all non-missing data, R_{tk} are the rank of X_{tk} among N observations, $\bar{R}_{t.}^{(c)}$ are the average of overall ranks of complete cases, $\bar{R}_{t.}^{(i)}$ are the average of overall ranks of incomplete cases, and $\theta_t = \frac{n_c}{n_c+n_t}$, for $t = 1, 2$. In the current study we will make use of this estimator for the method discussed in Chapter 4.

Konietschke et al. (2012a) show that the quantity $\sqrt{n}(\hat{p}_\theta - p)$ is has asymptotically normal distribution with mean 0 and variance σ^2. When σ_c^2 and σ_g^2 are the variances of rank transforms in the complete and incomplete subjects, respectively

$$\sigma^2 = n\left(\frac{\sigma_c^2}{n_c} + (1-\theta_2)^2\frac{\sigma_2^2}{n_2} - (1-\theta_1)^2\frac{\sigma_1^2}{n_1}\right). \tag{2.18}$$

To test the null hypothesis

$$H_o^p : p = \frac{1}{2},$$

Konietschke et al. (2012a) propose the tests statistic T_θ^p which follows standard normal distribution asymptotically and is given by

$$T_\theta = \sqrt{n}\frac{(\hat{p}_\theta - 1/2)}{\hat{\sigma}_\theta},$$

where $\hat{\sigma}_\theta^2$ is a consistent estimator of σ^2 in equation (2.18). Extension of this test procedure to two-groups case is developed in Chapter 3 of this research work.

For small sample sizes, the distribution of the test statistic, T_θ can be approximated by $t(\hat{\nu})$-distribution. Konietschke et al. (2012a) give asymptotic $(1-\alpha)$ confidence intervals for the relative effect.

Konietschke et al. (2018) developed a test procedure called Multiplication-Combination Test (MCT) for the same setup (one group and two time points). This procedure is applicable in semi-parametric and nonparametric setup to test hypotheses formulated in terms of various effect sizes and it works under the MCAR framework. Moreover, they show that their method is more accurate in maintaining the nominal level of significance than the existing competing procedures like Konietschke et al. (2012a).

2.3 The Way Forward

The nonparametric test procedure for data with missing values by Konietschke et al. (2012a) discussed in the previous section details the way for handling partially matched pairs but its use is limited to the case where there are repeated measures in one group only. Consequently, the need for a test procedure which can handle repeated measures from two or more groups arises. It is one of the objective of this research work to extend this test procedure to two-group and two-time points setup and derive a test procedure for testing significance of group-time interaction effect. In response to the first objective, the model in equation (2.14) is modified as

$$\boldsymbol{X_k} = (\Delta_{1k}X_{1k}, \Delta_{2k}X_{2k})', \tag{2.19}$$

where $g = 1, 2$, and $k = 1, ..., n_g$. In this model measurements at both time points can contain missing values, where Δ_{1k} is a missing value indicator which assumes the value 1 when a datum is observed and 0 otherwise.

Under the model in equation (2.19) above, we define a measure of interaction (group-time) effect size, propose an estimator for it and then study the asymptotic behavior of the proposed estimator.

We also develop procedures for hypothesis testing and build a confidence interval for the interaction effect size. Next, we investigate the performance of the procedures via simulation. Lastly, we apply the test procedure to a real dataset example. All of that is detailed in Chapter 3 of the dissertation.

Regarding the second objective we extend the test procedure developed by Konietschke et al. (2010) to missing data situation. We specify the model measurements as follows

$$\boldsymbol{X_k} = (\Delta_{1k}X_{1k}, \Delta_{2k}X_{2k}, \Delta_{3k}X_{3k})', \tag{2.20}$$

where $k = 1, ..., n_l$. In this model measurements are missing at least at one time point. The Δ_{lk}s defined above are not going to be directly used in the derivation of theoretical results. We will make use of them when the simulation is carried out. It comes in handy when defining the various counts of observations involved in the theoretical derivations.

The consideration in 2.20 refers to a situation where we have repeated measures in one group. Here the test procedure developed is limited to three time points as the derivation of the theoretical results becomes laborious, specifically that of the covariance matrix. We propose an estimator for a vector of relative effects (l = 3) and derive the asymptotic theoretical results. We then setup a test procedure for testing the significance of time effect. The properties of the test procedures are studied via size and power simulation. Finally, all the steps followed in the derivation of this test procedure and the application to real data will be presented in Chapter 4.

Chapter 3

Partially Matched-Pairs Design

3.1 Introduction

In typical two-arm randomized controlled trials or observational studies, the efficacy of the active treatment is validated by comparing it with a control. In such trials, outcomes are typically assessed at baseline, before treatments are applied, and at follow-up, after the treatments are applied. Due to the baseline measurement, each subject serves as its own control, thereby allowing a precise comparison of the two treatments. The primary question of interest is whether the changes in the outcome in the active treatment group from baseline to followup is different from the change in the control group. The two assessments of the outcome can be viewed as levels of a within-subject factor and the two groups as levels of a between-subject factor. Therefore, in the terminology of of repeated measures analysis, the primary question of interest becomes investigation of time-by-treatment interaction (Davis, 2002).

Under the assumption of normality, the data analysis for the above inferential problem can be carried out by a t-test with equal or unequal variances as appropriate. In the absence of normality, Generalized Estimating Equations (GEE) or Generalized Linear Mixed Model (GLMM) can be applied if the data fit certain parametric models (Diggle et al., 2013). When data are not measured in a metric scale (such as ordered categorical data) or when data have heavy tails or are skewed for any parametric model to be appropriate, nonparametric methods are the preferred analytic methods.

The seemingly elementary data analysis problem discussed above could quickly turn into a challenge if the data are subject to missing values and the assumption of normality is grossly violated. In today's data collection methods and procedures, unavailability of data from study units due to, for example, failure of devices or progression of disease is the norm rather than exception. Data could be *missing completely at random* (MCAR) for a reason unrelated to the study variables or *missing at random* (MAR) for a reason unrelated to the actual value missed but may be related to other study variables. In this chapter we focus on the former type. Albeit simple, sometimes this missing type can be motivated from a design and survey perspectives, for example, see the discussion at the end of Section 3.2. See also Fong et al. (2017), Fuchs et al. (2017), Samawi and Vogel (2014) and Samawi and Vogel (2015) and Xu and Harrar (2012) for examples in medicine and public health.

Under the assumptions of parametric models, Expectation-Maximization (EM) or Multiple Imputation (MI) algorithms (Dempster et al., 1977) can be applied in conjunction with the assumed models for the complete data. Recent semi-parametric approaches that use all available information, but at least require existence of the first few moments, include Xu and Harrar (2012), Samawi and Vogel (2014) and Samawi and Vogel (2015), Amro and Pauly (2017) and the references therein. The methods of Samawi and Vogel (2014) and Samawi and Vogel (2015) are in the context of partially paired data and work by combining paired and independent sample tests for the paired and unpaired, respectively, portion of the data. Apart from using permutations to determine the null distribution, Amro and Pauly (2017) also combine t statistics calculated from the complete and incomplete data separately.

Nonparametric methods for incomplete paired data received the attention of researchers fairly recently (Akritas et al., 2002; Akritas et al., 2006; Konietschke et al., 2012a; Samawi and Vogel, 2014; Samawi and Vogel, 2015; Fong et al., 2017). The papers by Akritas et al. (2002) and Akritas et al. (2006) introduce nonparametric tests for one and multiple-group, respectively, paired data under MAR-type missing values. However, the hypotheses of interest are formulated in terms of marginal distributions and the asymptotic variances are estimated under the null hypotheses. Therefore, the tests cannot be used to construct confidence intervals. The methods of Fong et al. (2017) and Samawi and Vogel (2014) and Samawi and Vogel (2015), which are constructed by combining the Wilcoxon-Signed-Rank and Wilcoxon-Mann-Whitney tests, also suffer from this problem. Further, they are intended for partially paired data in one group. Konietschke et al. (2012a) consider a nonparametric method for partially paired data. Its strengths are that the hypothesis are formulated in terms of nonparametric relative effect and the asymptotic variance of the test statistic is derived under general conditions. Therefore, the asymptotic results can be used to construct confidence intervals. However, the method mentioned above is designed for one group situation and is not applicable for treatment comparisons in a two-arms and two-assessments studies which is the focus of this study. Therefore, it is an aim of this research work to derive nonparametric tests for two-arm trials where the outcome, assessed at two time points, is subject to missing values.

The nonparametric method will accommodate binary, ordered categorical, discrete and continuous data in a unified manner. The within-pair dependence can be different in the two groups. The approach can also be used to construct a confidence interval for the primary effect of interest, which is the interaction effect. Unlike some of the other methods (e.g. Samawi and Vogel (2014), Samawi and Vogel (2015), Xu and Harrar (2012), and Amro and Pauly (2017)), the new nonparametric method explicitly compares the complete and incomplete pieces of the data to perform a test and construct a confidence interval. Table 3.1 shows schematic layout of data of this chapter.

Treatment 1			Treatment 2		
Subject	Time = 1	Time = 2	Subject	Time = 1	Time = 2
1	x	x	1	x	x
⋮	⋮	⋮	⋮	⋮	⋮
n_{c1}	x	x	n_{c2}	x	x
1	x	?	1	x	?
⋮	⋮	⋮	⋮	⋮	⋮
n_{11}	x	?	n_{21}	x	?
1	?	x	1	?	x
⋮	⋮	⋮	⋮	⋮	⋮
n_{12}	?	x	n_{22}	?	x

TABLE 3.1: Schematic display of the dataset in paired design in two groups. In the table “x" indicates available observation and “?" indicates missing observation.

3.2 Model and Interaction-Effect Size Measure

Suppose there are n subjects in two groups measured at two time points, containing n_{cg} complete cases and $n_{g1} + n_{g2}$ incomplete cases in group g, where n_{g1} is the number of incomplete cases at the first time point and n_{g2} is the number of incomplete cases at the second time point in group g for $g = 1, 2$. Denote the paired observations from the complete cases by $(X^{(c)}_{g1k}, X^{(c)}_{g2k})$ for $g = 1, 2$, and $k = 1, 2, \ldots, n_{cg}$, with marginal distributions given as

$$X^{(c)}_{gtk} \sim F_{gt},$$

for $g, t = 1, 2$. Under the assumption of MCAR the incomplete cases, denoted by $X^{(i)}_{gtk}$, follow the same marginal distribution as the complete cases, i.e.

$$X^{(i)}_{gtk} \sim F_{gt},$$

for $k = 1, \ldots, n_{gt}$. There are in total N observations from n subjects. For the sake of later use, we also define m_{gt} as the total number of available observations in group g at occasion t for $g, t = 1, 2$. Here,

$$N = \sum_{g=1}^{2}(2n_{cg} + n_{g1} + n_{g2}) = \sum_{g,t=1}^{2} m_{gt} \quad \text{and} \quad n = \sum_{g=1}^{2}(n_{cg} + n_{g1} + n_{g2}),$$

where $m_{gt} = n_{cg} + n_{gt}$.

When defining non-parametric treatment effects and their estimators, the normalized version of the distribution function of a random variable in equation (3.1) below will be used in place of the usual (right continuous) distribution function (Ruymgaart (1980) and Brunner et al. (2002)).

The normalized distribution function for an observation in group g at occasion t is defined as

$$F_{gt}(x) := \frac{1}{2}\left\{F_{gt}^{+}(x) + F_{gt}^{-}(x)\right\}, \tag{3.1}$$

where $F_{gt}^{-}(x) = P(X_{gtk} < x)$ and $F_{gt}^{+}(x) = P(X_{gtk} \leq x)$ are, respectively, the left and right continuous versions of the unknown distribution function. The use of normalized distribution function allows the treatment of binary, ordinal, discrete and continuous data in a unified manner. Nonparametric treatment effects are defined by comparing each marginal distribution function with the average distribution function. Let G denote the average of the distribution functions in the two groups at the two time points. That is,

$$G := \frac{1}{4}(F_{11} + F_{12} + F_{21} + F_{22}).$$

Using this average distribution function, define

$$p_{gt} := \int G dF_{gt}$$

for $g, t = 1, 2$ which is known as the *relative summary effect* at time point t in group g, with respect to the average of the marginal distributions, G. The use of unweighted rather than weighted average of the distribution function removes the dependence of test results on sample size allocation (see for example, Gao and Alvo, 2005). The magnitude of p_{gt} has interpretation in terms of the corresponding marginal distribution having a tendency to generate larger or smaller values compared to the overall sample. For example, $p_{12} < p_{22}$ means observations from F_{12} have a tendency to be larger than those from F_{22} (Brunner et al., 2002). Using the nonparametric relative summary effects, the interaction effect is defined by

$$p_I := (p_{12} - p_{11}) - (p_{22} - p_{21}). \tag{3.2}$$

Our aims are to find an estimator for summary interaction effects size and to develop methods for confidence intervals and significance test. The null hypothesis of interest is that there is no interaction effect, i.e. $H_0 : p_I = 0$, versus the alternative that there is an interaction effect. Figure 3.1 below is the graphical display of null and alternative hypothesis in terms of relative effects.

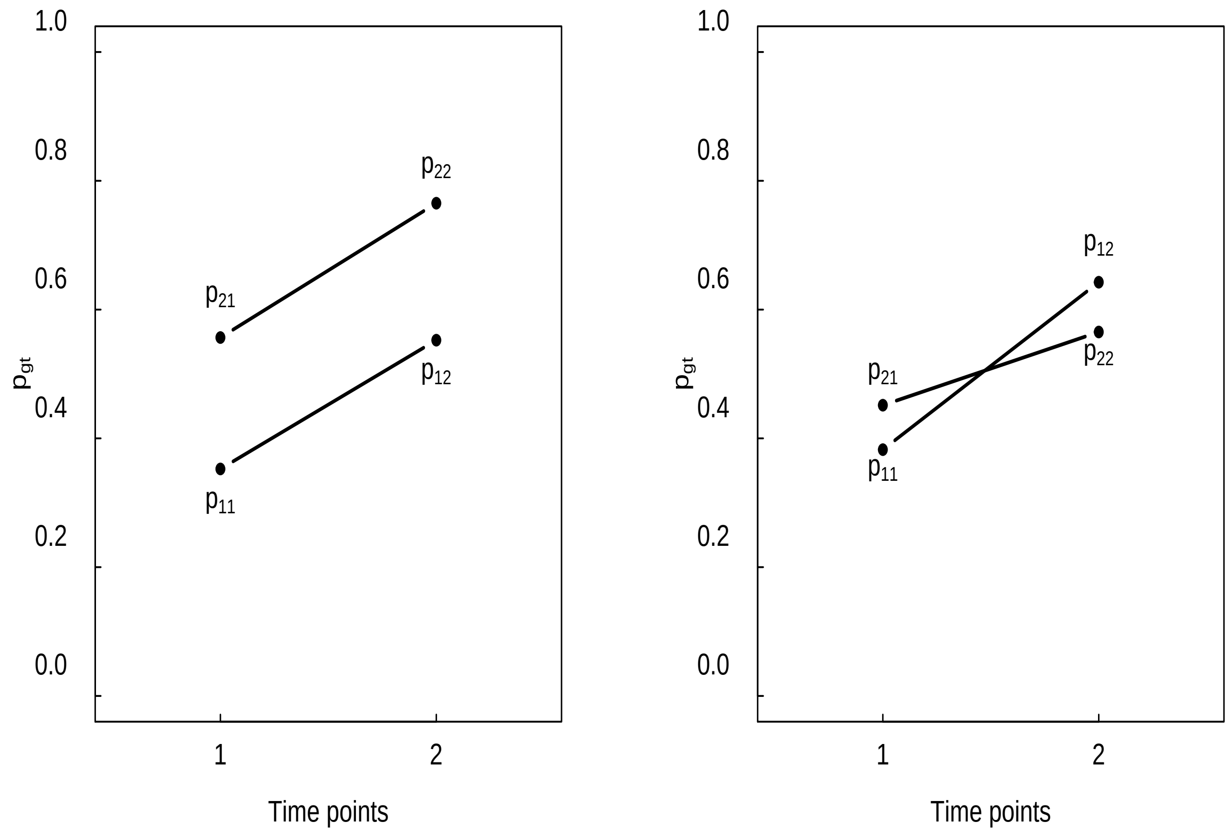


FIGURE 3.1: Null hypothesis (Left) and Alternative hypothesis (Right).

To facilitate the development of these inferential procedures, Proposition 1 below gives a more elaborate form of this interaction effect. To prove Proposition 1 we need the following Lemma.

Lemma 1. *For given normalized distribution functions F_1 and F_2,*
$\int_{-\infty}^{\infty} F_1 dF_2 = 1 - \int_{-\infty}^{\infty} F_2 dF_1.$

Proof. The result follows immediately by applying integration by parts formula on the Riemann–Stieltjes integrals (Hewitt, 1960).

$$\begin{aligned}\int_{-\infty}^{\infty} F_1 dF_2 + \int_{-\infty}^{\infty} F_2 dF_1 &= F_1(\infty)F_2(\infty) - F_1(-\infty)F_2(-\infty) \\ &= (1)(1) - (0)(0) \\ &= 1 \\ \Rightarrow \int_{-\infty}^{\infty} F_1 dF_2 = 1 - \int_{-\infty}^{\infty} F_2 dF_1 &\end{aligned}$$

□

Next Proposition 1 is stated as follows.

Proposition 1. *Consider the interaction effect defined in equation (3.2), then*

$$p_{\mathrm{I}} = \int \frac{1}{2}(F_{11} + F_{22})d(F_{12} + F_{21}) - 1. \tag{3.3}$$

Proof. By definition

$$\begin{aligned}p_{\mathrm{I}} &= \int G d(F_{12} + F_{21}) - \int G d(F_{11} + F_{22}) \\ &= \int \frac{1}{4}(F_{11} + F_{22})d(F_{12} + F_{21}) + \int \frac{1}{4}(F_{12} + F_{21})d(F_{12} + F_{21}) - \int \frac{1}{4}(F_{11} + F_{22})d(F_{11} + F_{22}) \\ &\quad - \int \frac{1}{4}(F_{12} + F_{21})d(F_{11} + F_{22}) \\ &= \int \frac{1}{4}(F_{11} + F_{22})d(F_{12} + F_{21}) + \frac{1}{2} - \frac{1}{2} - (1 - \frac{1}{4}\int (F_{11} + F_{22})d(F_{12} + F_{21})) \\ &= \frac{1}{2}\int (F_{11} + F_{22})d(F_{12} + F_{21}) - 1.\end{aligned}$$

□

It is easy to show (Brunner et al., 2002) that $1/8 \leq p_{gt} \leq 7/8$ which occurs as an artifact of the definition of p_{gt}. As can clearly be seen from (3.3), that the interaction effect p_{I} does not suffer from this constraint on p_{gt}. Moreover, the integral expression for p_{I} illuminates its interpretation as an interaction effect. Notice that

$$\int \frac{1}{2}(F_{11} + F_{22})d\{\frac{1}{2}(F_{12} + F_{21})\} = P(X < Y) + \frac{1}{2}P(X = Y)$$

where X and Y are independently distributed as $(1/2)(F_{11}+F_{22})$ and $(1/2)(F_{12}+F_{21})$, respectively. Therefore, $p_{\text{I}} < 0$ means the combined observations in group 1 at times 1 and group 2 at time 2 have a tendency to be larger than the combined observations in group 2 at times 1 and group 1 at time 2, which is an indication of time-by-group interaction effect.

Apart from constants, the interaction effect can equivalently and, yet, more conveniently be expressed as

$$q_I := \frac{1}{2}\int (F_{11} + F_{22})d(F_{12} + F_{21}). \tag{3.4}$$

More specifically, it should be noted that $q_{\text{I}} \in [0, 2]$ and $p_{\text{I}} = 0$ if and only if $q_{\text{I}} = 1$.

The nonparametric interaction effect size measure, q_{I}, can be estimated by replacing the distribution functions F_{gt} with their corresponding empirical counterparts. However, the contribution of the complete and incomplete cases to the empirical distribution must be weighted based on their sizes. To that end, we define the weights

$$\theta_{gt} := \frac{n_{cg}}{n_{cg} + n_{gt}},$$

which are the relative sample sizes of the complete and incomplete observations in group g, and at occasion t, for g, t = 1, 2. The weighted empirical distribution functions are defined by

$$\hat{F}_{gt,\theta_{gt}}(x) := \theta_{gt}\hat{F}_{gt}^{(c)}(x) + (1 - \theta_{gt})\hat{F}_{gt}^{(i)}(x); \tag{3.5}$$

for $g, t = 1, 2$, where

$$\hat{F}_{gt}^{(c)}(x) = \frac{1}{n_{cg}}\sum_{k=1}^{n_{cg}} c(x - X_{gtk}^{(c)}) \quad \text{and} \quad \hat{F}_{gt}^{(i)}(x) = \frac{1}{n_{gt}}\sum_{k=1}^{n_{gt}} c(x - X_{gtk}^{(i)}),$$

and $c(\cdot)$ is the (normalized) counting function defined by

$$c(x) = \frac{1}{2}[c^-(x) + c^+(x)], \quad c^-(x) = \begin{cases} 0, & \text{if } x \le 0 \\ 1, & \text{if x} > 0 \end{cases}, \quad \text{and} \quad c^+(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \ge 0 \end{cases}.$$

According to the definition of $c(x)$, it is clear that $c(x) = 0, \frac{1}{2}, 1$ according as $x < 0$, $x = 0$ and $x > 0$. The use of normalized count function provides a normalized version of empirical distribution function (analogous to the normalized distribution function) that permits seamless handling of ties in the data. In the above expressions, *c* and *i* are written in the superscripts of distributions functions to indicate reference to the complete and incomplete cases, respectively. A natural estimator for q_I arises by replacing the distribution functions in (3.4) with their empirical counterparts given in (3.5),

$$\hat{q}_{I,\theta} = \frac{1}{2}\int (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}). \tag{3.6}$$

An alternative but more intuitive form of this estimator is the expression give in terms of ranks.

Proposition 2. *A rank-based expression for the point estimator* $\hat{q}_{\mathrm{I},\theta}$ *is*

$$\begin{aligned}
\hat{q}_{\mathrm{I},\theta} = \frac{1}{2}\left(\frac{\theta_{12}}{m_{11}}\left(\bar{R}^{(c)}_{11.12} - \bar{R}^{(c,12)}_{12}\right) + \frac{(1-\theta_{12})}{m_{11}}\left(\bar{R}^{(i)}_{11.12} - \bar{R}^{(i,12)}_{12}\right)\right)\\
+ \frac{1}{2}\left(\frac{\theta_{21}}{m_{11}}\left(\bar{R}^{(c)}_{11.21} - \bar{R}^{(c,21)}_{21}\right) + \frac{(1-\theta_{21})}{m_{11}}\left(\bar{R}^{(i)}_{11.21} - \bar{R}^{(i,21)}_{21}\right)\right)\\
+ \frac{1}{2}\left(\frac{\theta_{12}}{m_{22}}\left(\bar{R}^{(c)}_{22.12} - \bar{R}^{(c,12)}_{12}\right) + \frac{(1-\theta_{12})}{m_{22}}\left(\bar{R}^{(i)}_{22.12} - \bar{R}^{(i,12)}_{12}\right)\right)\\
+ \frac{1}{2}\left(\frac{\theta_{21}}{m_{22}}\left(\bar{R}^{(c)}_{22.21} - \bar{R}^{(c,21)}_{21}\right) + \frac{(1-\theta_{21})}{m_{11}}\left(\bar{R}^{(i)}_{22.21} - \bar{R}^{(i,21)}_{21}\right)\right)
\end{aligned}$$

where for $l \neq g$ or $s \neq t$, $R^{(w)}_{ls.gtk}$ is the mid-rank of $X^{(w)}_{gtk}$ among all $n_{cl} + n_{ls} + n_{cg} + n_{gt}$ observations available in group l at time s and group g at time t; and $R^{(w,gt)}_{gtk}$ is the mid-rank of $X^{(w)}_{gtk}$ among all $n_{cg} + n_{gt}$ observations available in group g at time t for $w \in \{c, i\}$.

Proof. Observe that,

$$\hat{q}_{I,\theta} = \frac{1}{2}\left(\int \hat{F}_{11,\theta_{11}} d\hat{F}_{12,\theta_{12}} + \int \hat{F}_{11,\theta_{11}} d\hat{F}_{21,\theta_{21}} + \int \hat{F}_{22,\theta_{22}} d\hat{F}_{12,\theta_{12}} + \int \hat{F}_{22,\theta_{22}} d\hat{F}_{21,\theta_{21}}\right). \tag{3.7}$$

Now any of the terms inside the parenthesis of equation (3.7) can be written as

$$\int \hat{F}_{ls,\theta_{ls}} d\hat{F}_{gt,\theta_{gt}} = \theta_{gt}\frac{1}{n_{cg}}\sum_{k=1}^{n_{cg}} \hat{F}_{ls,\theta_{ls}}(X^{(c)}_{gtk}) + (1-\theta_{gt})\frac{1}{n_{gt}}\sum_{k=1}^{n_{gt}} \hat{F}_{ls,\theta_{ls}}(X^{(i)}_{gtk}),$$

where $l \neq g$ or $s \neq t$. It can be seen that

$$\hat{F}_{ls,\theta_{ls}}(X^{(w)}_{gtk}) = \frac{1}{m_{ls}}\left(R^{(w)}_{ls.gtk} - R^{(w,gt)}_{gtk}\right) \quad for \quad w \in \{c, i\}.$$

Therefore,

$$\int \hat{F}_{ls,\theta_{ls}} d\hat{F}_{gt,\theta_{gt}} = \frac{\theta_{gt}}{m_{ls}}\left(\bar{R}^{(c)}_{ls.gt} - \bar{R}^{(c,gt)}_{gt}\right) + \frac{(1-\theta_{gt})}{m_{ls}}\left(\bar{R}^{(i)}_{ls.gt} - \bar{R}^{(i,gt)}_{gt}\right).$$

So, plugging the above into equation (3.7) gives the desired result.

In view of the definition of p_I (see equation (1)) and using results from Proposition 1, the interaction effect p_I, is estimated by

$$\hat{p}_{\text{I},\theta} = \hat{q}_{\text{I},\theta} - 1. \tag{3.8}$$

Due to the simple relationship between $\hat{p}_{\text{I},\theta}$ and $\hat{q}_{\text{I},\theta}$, the theory derived for $\hat{q}_{\text{I},\theta}$ extends to that of $\hat{p}_{\text{I},\theta}$ in a straightforward manner. For the derivation of theoretical results for $\hat{q}_{\text{I},\theta}$, we will need the following assumptions. Besides, the existence of the required moments is assumed in all forthcoming sections.

Assumption 1. *$n_{cg} + n_{gt} \to \infty$ such that $\frac{n}{n_{cg}+n_{gt}} \leq N_o < \infty$ for $g, t = 1, 2$.*

Assumption 2. *Missing is completely at random.*

Assumption 1 is satisfied if the number of available observation at each time point is large in both groups. This can happen, for example, if either the complete cases or both of the incomplete cases are large in each group. More precisely, either $n_{cg} \to \infty$ or $\min(n_{g1}, n_{g2}) \to \infty$ for each $g = 1, 2$. For instance, $n_{c1} \to \infty$ and $\min(n_{21}, n_{22}) \to \infty$ but $n_{11}, n_{12} \leq M_0 < \infty$ is a scenario covered by the asymptotic framework. Assumption 2, appears somewhat restrictive but the theory developed in this study uses all available data from the complete as well as incomplete cases to make valid inference. This type of missingness can be motivated from partially-paired data points of view. For example, in surveys complete data are collected only from some of the study participants that are randomly drawn from the collection of all participants (Rubin, 1987). This leads to paired data where some of the observations are correlated and others are independent. In a design of experiment context, incomplete block design also leads to partially correlated data. For example, consider a design where the two treatments are randomized within twins (matched-pairs) when available in the pool of experimental units. Otherwise, the two treatments are assigned to independent units. For additional examples of partially-paired data see Fong et al. (2017), Fuchs et al. (2017), Samawi and Vogel (2014) and Samawi and Vogel (2015) and Xu and Harrar (2012).

□

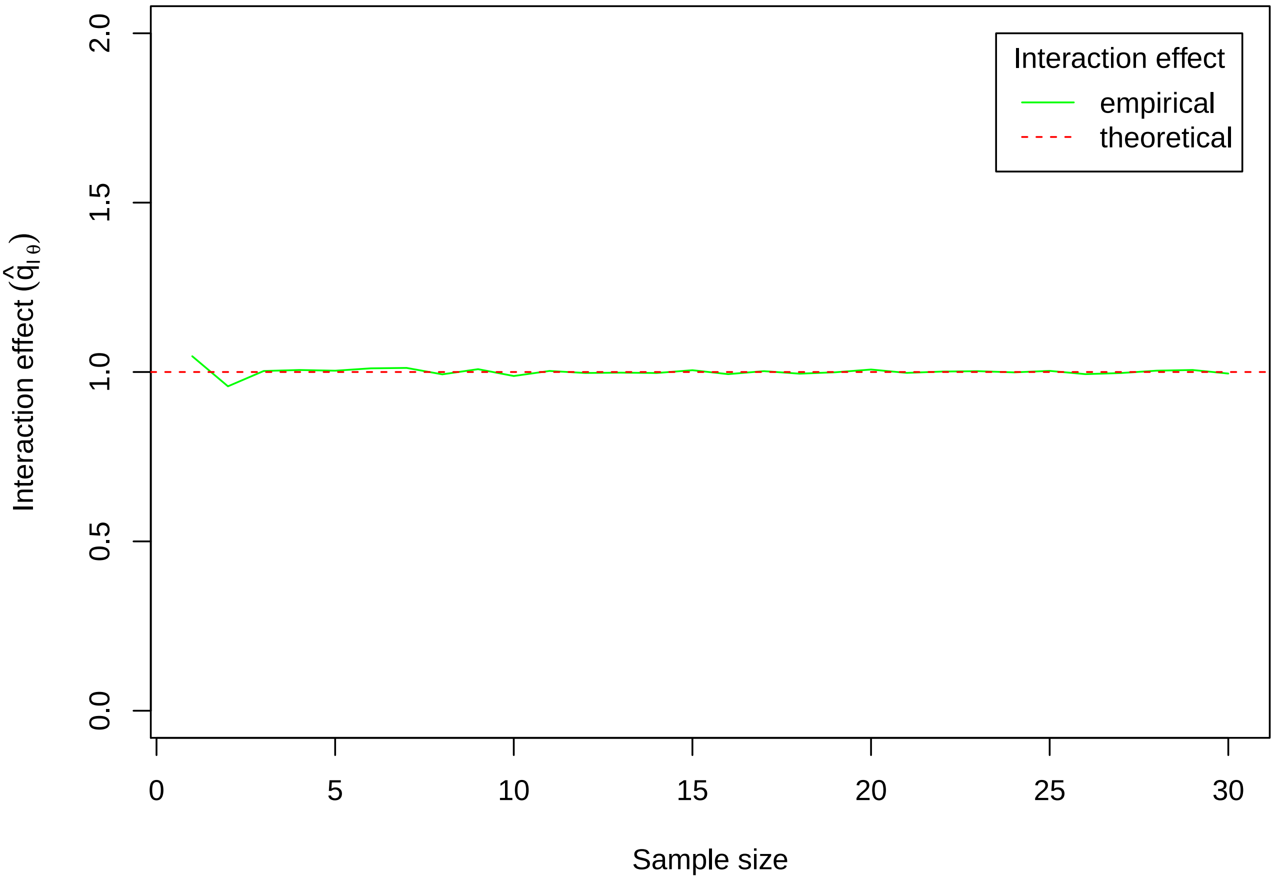


FIGURE 3.2: Convergence of interaction effect estimator ($\hat{q}_{I,\theta}$) based on the simulation size of 1,000. Sample of size 30 is generated from Cauchy distribution and 5% missing percentage is considered in estimating the interaction effect.

Figure 3.2 shows the property of the interaction effect estimator as the sample size gets large. The theoretical value of the interaction effect size equals 1, and the estimator converges quickly as the sample size increases.

3.3 Asymptotic Theory

Under the asymptotic framework of Assumption 1 and 2, the estimator $\hat{q}_{\mathrm{I},\theta}$ is asymptotically unbiased and strongly consistent. These facts are stated and proved in Proposition 3.

Proposition 3. *Under assumption 1 and 2,*

(i) $\hat{q}_{I,\theta}$ *is an asymptotically unbiased estimator of* q_I*.*

(ii) $\hat{q}_{I,\theta}$ *is a strongly consistent estimator of* q_I*, i.e.* $|\hat{q}_{I,\theta} - q_I| \xrightarrow{a.s} 0$.

Proof. (i) Unbiasedness: To show that $\hat{q}_{I,\theta}$ is an asymptotically unbiased estimator of q_I, we first apply the linearity property of expectation to equation (3.7) to get

$$E\left(\hat{q}_{\mathrm{I},\theta}\right) = \frac{1}{2}\left[E\left(\int \hat{F}_{11,\theta_{11}} d\hat{F}_{12,\theta_{12}}\right) + E\left(\int \hat{F}_{11,\theta_{11}} d\hat{F}_{21,\theta_{21}}\right)\right] + \frac{1}{2}\left[E\left(\int \hat{F}_{22,\theta_{22}} d\hat{F}_{12,\theta_{12}}\right) + E\left(\int \hat{F}_{22,\theta_{22}} d\hat{F}_{21,\theta_{21}}\right)\right]. \tag{3.9}$$

Taking the first term inside the square brackets of equation (3.9); i.e $E\left(\int \hat{F}_{11,\theta_{11}} d\hat{F}_{12,\theta_{12}}\right)$, we will show that $\int \hat{F}_{11,\theta_{11}} d\hat{F}_{12,\theta_{12}}$ is asymptotically an unbiased estimator of $\int F_{11} dF_{12}$. Define $p := \int F_{11} dF_{12}$ and its estimator by $\hat{p}_\theta := \int \hat{F}_{11,\theta_{11}} d\hat{F}_{12,\theta_{12}}$,

$$\hat{p}^{(cc)} := \int \hat{F}_{11}^{(c)} d\hat{F}_{12}^{(c)}, \quad \hat{p}^{(ci)} := \int \hat{F}_{11}^{(c)} d\hat{F}_{12}^{(i)}, \quad \hat{p}^{(ic)} := \int \hat{F}_{11}^{(i)} d\hat{F}_{12}^{(c)} \quad \text{and} \quad \hat{p}^{(ii)} := \int \hat{F}_{11}^{(i)} d\hat{F}_{12}^{(i)}.$$

Since $X_{11k}^{(c)}$ and $X_{12k}^{(i)}$ are independent, it follows that

$$\begin{aligned} E(\theta_{11}\theta_{12}\hat{p}^{(cc)}) &= \frac{1}{m_{11}m_{12}} \sum_{s=1}^{n_{c1}} \sum_{k=1}^{n_{c1}} E(c(X_{12s}^{(c)} - X_{11k}^{(c)})) \\ &= \frac{1}{m_{11}m_{12}} \sum_{s\neq k}^{n_{c1}} E(c(X_{12s}^{(c)} - X_{11k}^{(c)})) + \frac{1}{m_{11}m_{12}} \sum_{k=1}^{n_{c1}} E(c(X_{12k}^{(c)} - X_{11k}^{(c)})) \\ &= \frac{n_{c1}(n_{c1}-1)}{m_{11}m_{12}} p + \frac{n_{c1}}{m_{11}m_{12}} \Delta, \end{aligned}$$

where $\Delta = E(c(X_{12k}^{(c)} - X_{11k}^{(c)}))$.

Similarly,

$$\begin{aligned} E(\theta_{11}(1-\theta_{12})\hat{p}^{(ci)}) &= \frac{1}{m_{11}m_{12}} \sum_{s=1}^{n_{12}} \sum_{k=1}^{n_{c1}} E(c(X_{12s}^{(i)} - X_{11k}^{(c)})) \\ &= \frac{n_{12}n_{c1}}{m_{11}m_{12}} p, \\ E((1-\theta_{11})\theta_{12}\hat{p}^{(ic)}) &= \frac{1}{m_{11}m_{12}} \sum_{s=1}^{n_{c1}} \sum_{k=1}^{n_{11}} E(c(X_{12s}^{(c)} - X_{11k}^{(i)})) \\ &= \frac{n_{11}n_{c1}}{m_{11}m_{12}} p \quad \text{and} \\ E((1-\theta_{11})(1-\theta_{12})\hat{p}^{(ii)}) &= \frac{1}{m_{11}m_{12}} \sum_{s=1}^{n_{12}} \sum_{k=1}^{n_{11}} E(c(X_{12s}^{(i)} - X_{11k}^{(i)})) \\ &= \frac{n_{11}n_{12}}{m_{11}m_{12}} p. \end{aligned}$$

Pulling together the expectations, we have

$$
\begin{aligned}
E(\hat{p}_\theta) &= E(\theta_{11}\theta_{12}\hat{p}^{(cc)} + \theta_{11}(1-\theta_{12})\hat{p}^{(ci)} + (1-\theta_{11})\theta_{12}\hat{p}^{(ic)} + (1-\theta_{11})(1-\theta_{12})\hat{p}^{(ii)}) \\
&= \frac{n_{c1}(n_{c1}-1)}{m_{11}m_{12}}p + \frac{n_{c1}}{m_{11}m_{12}}\Delta + \frac{n_{11}n_{c1}}{m_{11}m_{12}}p + \frac{n_{12}n_{c1}}{m_{11}m_{12}}p + \frac{n_{11}n_{12}}{m_{11}m_{12}}p \\
&= p + \frac{n_{c1}}{m_{11}m_{12}}(\Delta - p) \\
&= \int F_{11}dF_{12} + O(\frac{1}{n}),
\end{aligned}
$$

where the order $O(1/n)$ follows by Assumption 1 and 2. Applying similar arguments for the remaining three components in equation (3.9),

$$
E\left(\int \hat{F}_{ls,\theta_{ls}} d\hat{F}_{gt,\theta_{gt}}\right) = \int F_{ls}dF_{gt} + O(\frac{1}{n}),
$$

where $l \neq g$ or $s \neq t$. Collecting these expectations together,

$$
\begin{aligned}
E\left(\hat{q}_{I,\theta}\right) &= \frac{1}{2}\left[\int F_{11}dF_{12} + \int F_{11}dF_{21} + \int F_{22}dF_{12} + \int F_{22}dF_{21}\right] + O(\frac{1}{n}) \\
&= \frac{1}{2}\int (F_{11}+F_{22})d(F_{12}+F_{21}) + O(\frac{1}{n}) \\
&= \mathrm{q_I} + \mathrm{O}(\frac{1}{\mathrm{n}}).
\end{aligned}
$$

(ii) Strong consistency:

Recall that $||f||_\infty := \sup_x |f(x)|$, then,

$$
\begin{aligned}
|\hat{q}_{\mathrm{I},\theta} - q_{\mathrm{I}}| &= |\frac{1}{2}\int (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - \frac{1}{2}\int (F_{11}+F_{22})d(F_{12}+F_{21})| \\
&= |\frac{1}{2}\int (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) + \frac{1}{2}\int (F_{11}+F_{22})d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) \\
&- \frac{1}{2}\int (F_{11}+F_{22})d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - \frac{1}{2}\int (F_{11}+F_{22})d(F_{12}+F_{21})| \\
&= |\frac{1}{2}\int \left[(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11}+F_{22})\right] d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) \\
&+ \frac{1}{2}\int (F_{11}+F_{22})d\left[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - (F_{12}+F_{21})\right]| \\
&= |\frac{1}{2}\int \left[(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11}+F_{22})\right] d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) \\
&+ \frac{1}{2}\int \left[(F_{12}+F_{21}) - (\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}})\right] d(F_{11}+F_{22})| \\
&\leq |\frac{1}{2}\int \left[(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11}+F_{22})\right] d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}})| \\
&+ |\frac{1}{2}\int \left[(F_{12}+F_{21}) - (\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}})\right] d(F_{11}+F_{22})|, by \quad triangular \quad inequality
\end{aligned}
$$

$$\leq \frac{1}{2}\int |(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11} + F_{22})| d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}})$$
$$+ \frac{1}{2}\int |(F_{12} + F_{21}) - (\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}})| d(F_{11} + F_{22})$$
$$= \| (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11} + F_{22}) \|_\infty + \| (F_{12} + F_{21}) - (\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) \|_\infty .$$

Now,

$$\| (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11} + F_{22}) \|_\infty = \| (\hat{F}_{11,\theta_{11}} - F_{11}) + (\hat{F}_{22,\theta_{22}} - F_{22}) \|_\infty$$
$$\leq \| (\hat{F}_{11,\theta_{11}} - F_{11}) \|_\infty + \| \hat{F}_{22,\theta_{22}} - F_{22} \|_\infty \xrightarrow{a.s} 0$$

since by the Glivenko-Cantelli theorem, eg. (van der Vaart, 1998);

$$\| (\hat{F}_{11,\theta_{11}} - F_{11}) \|_\infty \quad \textit{and} \quad \| \hat{F}_{22,\theta_{22}} - F_{22} \|_\infty \xrightarrow{a.s} 0,$$

it follows that

$$\| (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11} + F_{22}) \|_\infty \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \| (F_{12} + F_{21}) - (\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) \|_\infty \xrightarrow{a.s} 0.$$

Hence, $|\hat{q}_{I,\theta} - q_I| \xrightarrow{a.s} 0$; as $n_{cg} + n_{gt} \to \infty$.

□

Next, we study the asymptotic distribution of $\hat{q}_{\mathrm{I},\theta}$. Our method of derivation involves obtaining an asymptotic equivalent version of $\sqrt{n}(\hat{q}_{\mathrm{I},\theta} - q_{\mathrm{I}})$ which can be expressed as the sum of independent random variables. The asymptotic distribution of the latter, under the asymptotic framework of Assumption 1, can be established by applying the standard central limit theorem for independent random variables. To that end, define

$$M_\theta = \frac{1}{2}\int (F_{11} + F_{22}) d(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - \frac{1}{2}\int (F_{12} + F_{21}) d(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) + 2 - 2q_{\mathrm{I}}. \tag{3.10}$$

Proposition 4. *Under Assumption 2, the quantity, M_θ, given in equation 3.10 can be expressed as sum of independent random variables as*

a) $M_\theta = \frac{1}{n_{c1}}\sum_{k=1}^{n_{c1}} Z_{1k}^{(c)} + \frac{1}{n_{c2}}\sum_{k=1}^{n_{c2}} Z_{2k}^{(c)} + \frac{1}{n_{12}}\sum_{k=1}^{n_{12}} Y_{12k}^{(i)} + \frac{1}{n_{21}}\sum_{k=1}^{n_{21}} Y_{21k}^{(i)} - \frac{1}{n_{11}}\sum_{k=1}^{n_{11}} Y_{11k}^{(i)} - \frac{1}{n_{22}}\sum_{k=1}^{n_{22}} Y_{22k}^{(i)} + 2 - 2q_{\mathrm{I}}.$
and

b) $E(M_\theta) = 0,$

where

$$Z_{1k}^{(c)} := \frac{1}{2}\left(\theta_{12}\left[F_{11}(X_{12k}^{(c)}) + F_{22}(X_{12k}^{(c)})\right] - \theta_{11}\left[F_{12}(X_{11k}^{(c)}) + F_{21}(X_{11k}^{(c)})\right]\right), \quad for \quad k = 1, ..., n_{c1}$$
$$Z_{2k}^{(c)} := \frac{1}{2}\left(\theta_{21}\left[F_{11}(X_{21k}^{(c)}) + F_{22}(X_{21k}^{(c)})\right] - \theta_{22}\left[F_{12}(X_{22k}^{(c)}) + F_{21}(X_{22k}^{(c)})\right]\right), \quad for \quad k = 1, ..., n_{c2},$$

and, for $k = 1, \dots, n_{gt}$;

$$Y_{11k}^{(i)} := \frac{(1-\theta_{11})}{2}\left(F_{12}(X_{11k}^{(i)}) + F_{21}(X_{11k}^{(i)})\right), \quad Y_{12k}^{(i)} := \frac{(1-\theta_{12})}{2}\left(F_{11}(X_{12k}^{(i)}) + F_{22}(X_{12k}^{(i)})\right),$$
$$Y_{21k}^{(i)} := \frac{(1-\theta_{21})}{2}\left(F_{11}(X_{21k}^{(i)}) + F_{22}(X_{21k}^{(i)})\right) \quad \text{and} \quad Y_{22k}^{(i)} := \frac{(1-\theta_{22})}{2}\left(F_{12}(X_{22k}^{(i)}) + F_{21}(X_{22k}^{(i)})\right).$$

Proof. Part a) Consider

$$\begin{aligned}
&\int (F_{11}+F_{22}) d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - \int (F_{12}+F_{21}) d(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})] \\
&= \int (F_{11}+F_{22}) d\hat{F}_{12,\theta_{12}} + \int (F_{11}+F_{22}) d\hat{F}_{21,\theta_{21}} - \int (F_{12}+F_{21}) d\hat{F}_{11,\theta_{11}} + \int (F_{12}+F_{21}) \hat{F}_{22,\theta_{22}} \\
&= \frac{1}{n_{c1}} \sum_{k=1}^{n_{c1}} \left(\theta_{12}\left[F_{11}(X_{12k}^{(c)}) + F_{22}(X_{12k}^{(c)})\right] - \theta_{11}\left[F_{12}(X_{11k}^{(c)}) + F_{21}(X_{11k}^{(c)})\right]\right) \\
&+ \frac{1}{n_{c2}} \sum_{k=1}^{n_{c2}} \left(\theta_{21}\left[F_{11}(X_{21k}^{(c)}) + F_{22}(X_{21k}^{(c)})\right] - \theta_{22}\left[F_{12}(X_{22k}^{(c)}) + F_{21}(X_{22k}^{(c)})\right]\right) \\
&+ \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} (1-\theta_{12})\left(F_{11}(X_{12k}^{(i)}) + F_{22}(X_{12k}^{(i)})\right) + \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} (1-\theta_{21})\left(F_{11}(X_{21k}^{(i)}) + F_{22}(X_{21k}^{(i)})\right) \\
&- \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} (1-\theta_{11})\left(F_{12}(X_{11k}^{(i)}) + F_{21}(X_{11k}^{(i)})\right) - \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} (1-\theta_{22})\left(F_{12}(X_{22k}^{(i)}) + F_{21}(X_{22k}^{(i)})\right)
\end{aligned}$$

The desired result follows immediately.

Part b)

$$\begin{aligned}
E(M_\theta) &= \frac{1}{n_{c1}} \sum_{k=1}^{n_{c1}} E(Z_{1k}^{(c)}) + \frac{1}{n_{c2}} \sum_{k=1}^{n_{c2}} E(Z_{2k}^{(c)}) + \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} E(Y_{12k}^{(i)}) + \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} E(Y_{21k}^{(i)}) - \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} E(Y_{11k}^{(i)}) \\
&- \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} E(Y_{22k}^{(i)}) + E(2 - 2q_I) \\
&= \frac{1}{n_{c1}} \sum_{k=1}^{n_{c1}} E(Z_{11}^{(c)}) + \frac{1}{n_{c2}} \sum_{k=1}^{n_{c2}} E(Z_{21}^{(c)}) + \frac{1}{n_{12}} \sum_{k=1}^{n_{12}} E(Y_{121}^{(i)}) + \frac{1}{n_{21}} \sum_{k=1}^{n_{21}} E(Y_{211}^{(i)}) - \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} E(Y_{111}^{(i)}) \\
&- \frac{1}{n_{22}} \sum_{k=1}^{n_{22}} E(Y_{221}^{(i)}) + 2 - 2q_I \\
&= E(Z_{11}^{(c)}) + E(Z_{21}^{(c)}) + E(Y_{121}^{(i)}) + E(Y_{211}^{(i)}) - E(Y_{111}^{(i)}) - E(Y_{221}^{(i)}) + 2 - 2q_I.
\end{aligned}$$

Replacing quantities in the bracket by their respective expression

$$
\begin{aligned}
2E(M_\theta) = \theta_{12}E\left[F_{11}(X_{121}^{(c)}) + F_{22}(X_{121}^{(c)})\right] - \theta_{11}E\left[F_{12}(X_{111}^{(c)}) + F_{21}(X_{111}^{(c)})\right] \\
+ \theta_{21}E\left[F_{11}(X_{211}^{(c)}) + F_{22}(X_{211}^{(c)})\right] - \theta_{22}E\left[F_{12}(X_{221}^{(c)}) + F_{21}(X_{221}^{(c)})\right] \\
+ (1-\theta_{12})E\left(F_{11}(X_{121}^{(i)}) + F_{22}(X_{121}^{(i)})\right) + (1-\theta_{21})E\left(F_{11}(X_{211}^{(i)}) + F_{22}(X_{211}^{(i)})\right) \\
-(1-\theta_{11})E\left(F_{12}(X_{111}^{(i)}) + F_{21}(X_{111}^{(i)})\right) - (1-\theta_{22})E\left(F_{12}(X_{221}^{(i)}) + F_{21}(X_{221}^{(i)})\right) + 2(2-2q_I)
\end{aligned}
$$

$$
\begin{aligned}
2E(M_\theta) = \theta_{12}\left[\int F_{11}dF_{12} + \int F_{22}dF_{12}\right] - \theta_{11}\left[\int F_{12}dF_{11} + \int F_{21}dF_{11}\right] \\
+ \theta_{21}\left[\int F_{11}dF_{21} + \int F_{22}dF_{21}\right] - \theta_{22}\left[\int F_{12}dF_{22} + \int F_{21}dF_{22}\right] \\
+ (1-\theta_{12})\left(\int F_{11}dF_{12} + \int F_{22}dF_{12}\right) + (1-\theta_{21})\left(\int F_{11}dF_{21} + \int F_{22}dF_{21}\right) \\
-(1-\theta_{11})\left(\int F_{12}dF_{11} + \int F_{21}dF_{11}\right) - (1-\theta_{22})\left(\int F_{12}dF_{22} + \int F_{21}dF_{22}\right) + 2(2-2q_I) \\
= \int (F_{11}+F_{22})d(F_{12}+F_{12}) - \left(4 - \int (F_{11}+F_{22})d(F_{12}+F_{12})\right) + 2(2-2q_I)
\end{aligned}
$$

So,

$$
E(M_\theta) = \frac{1}{2}\int (F_{11} + F_{22})d(F_{12} + F_{12}) - 2 + \frac{1}{2}\int (F_{11} + F_{22})d(F_{12} + F_{12}) + 2 - 2q_I = 0.
$$

□

Now, we are ready to establish the asymptotic equivalence of $\sqrt{n}(\hat{q}_{I,\theta} - q_I)$ and $\sqrt{n}M_\theta$.

Theorem 1. *Let M_θ be as given in* (3.10)*. Under Assumption 1 and 2,*

$$
E\left[\sqrt{n}(\hat{q}_{I,\theta} - q_I) - \sqrt{n}M_\theta\right]^2 = O(\frac{1}{n}).
$$

Proof. Consider the quantity

$$
\begin{aligned}
&(\hat{q}_{\mathrm{I},\theta} - q_{\mathrm{I}}) - M_\theta \\
&= (\hat{q}_{I,\theta} - q_I) - \left[2 - 2q_I - \frac{1}{2}\int (F_{12}+F_{21})d(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})] + \frac{1}{2}\int (F_{11}+F_{22})d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}})\right] \\
&= \hat{q}_{I,\theta} - [2 - \frac{1}{2}\int (F_{12}+F_{21})d(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})] - \frac{1}{2}\int (F_{11}+F_{22})d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) + q_I
\end{aligned}
$$

$$= \frac{1}{2}\int (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - \frac{1}{2}\int (\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}})d(F_{12} + F_{21})$$
$$- \frac{1}{2}\int (F_{11} + F_{22})d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) + \frac{1}{2}\int (F_{11} + F_{22})d(F_{12} + F_{21})$$
$$= \frac{1}{2}\int \Big[(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11} + F_{22})\Big] d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - (F_{12} + F_{21})]. \tag{3.11}$$

Now,

$$\int \Big[(\hat{F}_{11,\theta_{11}} + \hat{F}_{22,\theta_{22}}) - (F_{11} + F_{22})\Big] d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - (F_{12} + F_{21})]$$
$$= \int \Big[(\hat{F}_{11,\theta_{11}} - F_{11}) + (\hat{F}_{22,\theta_{22}} - F_{22})\Big] d[(\hat{F}_{12,\theta_{12}} + \hat{F}_{21,\theta_{21}}) - (F_{12} + F_{21})]$$
$$= \underbrace{\int (\hat{F}_{11,\theta_{11}} - F_{11})d\hat{F}_{12,\theta_{12}} - \int (\hat{F}_{11,\theta_{11}} - F_{11})dF_{12}}_{A_1} + \underbrace{\int (\hat{F}_{11,\theta_{11}} - F_{11})d\hat{F}_{21,\theta_{21}} - \int (\hat{F}_{11,\theta_{11}} - F_{11})dF_{21}}_{A_2}$$
$$+ \underbrace{\int (\hat{F}_{22,\theta_{22}} - F_{22})d\hat{F}_{12,\theta_{12}} - \int (\hat{F}_{22,\theta_{22}} - F_{22})dF_{12}}_{A_3} + \underbrace{\int (\hat{F}_{22,\theta_{22}} - F_{22})d\hat{F}_{21,\theta_{21}} - \int (\hat{F}_{22,\theta_{22}} - F_{22})dF_{21}}_{A_4}.$$

To arrive at the desired result, it suffices to show that

$$nE(A_i)^2 = O(\frac{1}{n}), \quad for \quad i = 1, 2, 3, 4.$$

Now,

$$nE(A_1)^2 = nE\left[\int (\hat{F}_{11,\theta_{11}} - F_{11})d\hat{F}_{12,\theta_{12}} - \int (\hat{F}_{11,\theta_{11}} - F_{11})dF_{12}\right]^2.$$

Notice that,

$$\hat{F}_{11,\theta_{11}}(x) - F_{11}(x) = \frac{\sum_{s=1}^{n_{c1}} c(x - X_{11s}^{(c)}) + \sum_{s'=1}^{n_{11}} c(x - X_{11s'}^{(i)})}{n_{c1} + n_{11}} - F_{11}(x)$$
$$= \frac{\sum_{s=1}^{n_{c1}} [c(x - X_{11s}^{(c)}) - F_{11}(x)] + \sum_{s'=1}^{n_{11}} [c(x - X_{11s'}^{(i)}) - F_{11}(x)]}{n_{c1} + n_{11}}.$$

Therefore,

$$\int(\hat{F}_{11,\theta_{11}}-F_{11})d\hat{F}_{12,\theta_{12}}=\int(\hat{F}_{11,\theta_{11}}-F_{11})d\left(\frac{\sum_{k=1}^{n_{c1}}c(x-X_{12k}^{(c)})+\sum_{k'=1}^{n_{12}}c(x-X_{12k'}^{(i)})}{n_{c1}+n_{12}}\right)$$
$$=\frac{\sum_{k=1}^{n_{c1}}[\hat{F}_{11,\theta_{11}}(X_{12k}^{(c)})-F_{11}(X_{12k}^{(c)})]}{n_{c1}+n_{12}}+\frac{\sum_{k'=1}^{n_{12}}[\hat{F}_{11,\theta_{11}}(X_{12k'}^{(i)})-F_{11}(X_{12k'}^{(i)})]}{n_{c1}+n_{12}}$$
$$=\frac{\sum_{k=1}^{n_{c1}}\sum_{s=1}^{n_{c1}}[c(X_{12k}^{(c)}-X_{11s}^{(c)})-F_{11}(X_{12k}^{(c)})]}{(n_{c1}+n_{11})(n_{c1}+n_{12})}+\frac{\sum_{k=1}^{n_{c1}}\sum_{s'=1}^{n_{11}}[c(X_{12k}^{(c)}-X_{11s'}^{(i)})-F_{11}(X_{12k}^{(c)})]}{(n_{c1}+n_{11})(n_{c1}+n_{12})}$$
$$+\frac{\sum_{k'=1}^{n_{12}}\sum_{s=1}^{n_{c1}}[c(X_{12k'}^{(i)}-X_{11s}^{(c)})-F_{11}(X_{12k'}^{(i)})]}{(n_{c1}+n_{11})(n_{c1}+n_{12})}+\frac{\sum_{k'=1}^{n_{12}}\sum_{s'=1}^{n_{11}}[c(X_{12k'}^{(i)}-X_{11s'}^{(i)})-F_{11}(X_{12k'}^{(i)})]}{(n_{c1}+n_{11})(n_{c1}+n_{12})}$$

and

$$\int(\hat{F}_{11,\theta_{11}}-F_{11})dF_{12}(x)=\frac{1}{n_{c1}+n_{11}}\int\sum_{k=1}^{n_{c1}}\left(\sum_{s=1}^{n_{c1}}c(x-X_{11s}^{(c)})-F_{11}(x)+\sum_{s'=1}^{n_{11}}c(x-X_{12s'}^{(i)})-F_{11}(x)\right)$$
$$=\frac{\sum_{k=1}^{n_{c1}}\sum_{s=1}^{n_{c1}}\int[c(x-X_{11s}^{(c)})-F_{11}(x)]dF_{12}(x)}{(n_{c1}+n_{11})(n_{c1}+n_{12})}+\frac{\sum_{k=1}^{n_{c1}}\sum_{s'=1}^{n_{11}}\int[c(x-X_{11s'}^{(i)})-F_{11}(x)]dF_{12}(x)}{(n_{c1}+n_{11})(n_{c1}+n_{12})}$$
$$+\frac{\sum_{k'=1}^{n_{12}}\sum_{s=1}^{n_{c1}}\int[c(x-X_{11s}^{(c)})-F_{11}(x)]dF_{12}(x)}{(n_{c1}+n_{11})(n_{c1}+n_{12})}+\frac{\sum_{k'=1}^{n_{12}}\sum_{s'=1}^{n_{11}}\int[c(x-X_{11s'}^{(i)})-F_{11}(x)]dF_{12}(x)}{(n_{c1}+n_{11})(n_{c1}+n_{12})}.$$

Now A_1 can be rewritten as:

$$A_1=\underbrace{\frac{1}{(n_{c1}+n_{11})(n_{c1}+n_{12})}\sum_{k=1}^{n_{c1}}\sum_{s=1}^{n_{c1}}\left(c(X_{12k}^{(c)}-X_{11s}^{(c)})-F_{11}(X_{12k}^{(c)})-\int[c(x-X_{11s}^{(c)})-F_{11}(x)]dF_{12}(x)\right)}_{A_{11}}$$
$$+\underbrace{\frac{1}{(n_{c1}+n_{11})(n_{c1}+n_{12})}\sum_{k=1}^{n_{c1}}\sum_{s'=1}^{n_{11}}\left(c(X_{12k}^{(c)}-X_{11s'}^{(i)})-F_{11}(X_{12k}^{(c)})-\int[c(x-X_{11s'}^{(i)})-F_{11}(x)]dF_{12}(x)\right)}_{A_{12}}$$
$$+\underbrace{\frac{1}{(n_{c1}+n_{11})(n_{c1}+n_{12})}\sum_{k'=1}^{n_{12}}\sum_{s=1}^{n_{c1}}\left(c(X_{12k'}^{(i)}-X_{11s}^{(c)})-F_{11}(X_{12k'}^{(i)})-\int[c(x-X_{11s}^{(c)})-F_{11}(x)]dF_{12}(x)\right)}_{A_{13}}$$
$$+\underbrace{\frac{1}{(n_{c1}+n_{11})(n_{c1}+n_{12})}\sum_{k'=1}^{n_{12}}\sum_{s'=1}^{n_{11}}\left(c(X_{12k'}^{(i)}-X_{11s'}^{(i)})-F_{11}(X_{12k'}^{(i)})-\int[c(x-X_{11s'}^{(i)})-F_{11}(x)]dF_{12}(x)\right)}_{A_{14}}.$$

Again by C_r-inequality,

$$nE(A_1^2) \leq n4[E(A_{11})^2 + E(A_{12})^2 + E(A_{13})^2 + E(A_{14})^2].$$

Now,

$$\begin{aligned} nE(A_{11}^2) &= \frac{n}{(n_{c1}+n_{11})^2(n_{c1}+n_{12})^2}\sum_{k=1}^{n_{c1}}\sum_{s=1}^{n_{c1}}\sum_{l=1}^{n_{c1}}\sum_{t=1}^{n_{c1}} E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})\right) \\ &= \frac{n}{(n_{c1}+n_{11})^2(n_{c1}+n_{12})^2}[\sum_{\substack{\text{two indices are equal} \\ \text{and the other two equal}}}\sum\sum\sum E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})\right) \\ &+ \frac{n}{(n_{c1}+n_{11})^2(n_{c1}+n_{12})^2}\sum_{\substack{\text{one index is} \\ \text{different from the other three}}}\sum\sum\sum E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})\right)], \end{aligned}$$

where

$$\phi(X,Y) = c(X-Y) - F_{11}(X) - \int (c(x-Y) - F_{11}(x))dF_{12}(x).$$

Since the summand in the first sum is the expectation of a uniformly bounded term, its order is $O(n_{c1}^2)$. For the second sum, if any one of the indices is different from the other three, for example $k \neq (l,s,t)$, then

$$E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})\right) = E\left[E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})|X_{11s}^{(c)}, X_{12l}^{(c)}, X_{11t}^{(c)}\right)\right].$$

Consider the inner expectation,

$$E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})|X_{11s}^{(c)}, X_{12l}^{(c)}, X_{11t}^{(c)}\right) = \phi(X_{12l}^{(c)}, X_{11t}^{(c)})E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})|X_{11s}^{(c)}\right).$$

It then follows that

$$\begin{aligned} E(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})|X_{11s}^{(c)}) &= E(c(X_{12k}^{(c)} - X_{11s}^{(c)}) - F_{11}(X_{12k}^{(c)})|X_{11s}^{(c)}) - E(\int [c(x - X_{11s}^{(c)}) - F_{11}(x)]dF_{12}(x)|X_{11s}^{(c)}) \\ &= \int [c(x - X_{11s}^{(c)}) - F_{11}(x)]dF_{12}(x) - \int [c(x - X_{11s}^{(c)}) - F_{11}(x)]dF_{12}(x) = 0. \end{aligned}$$

Similarly, for $s \neq (k,l,t)$,

$$E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})\right) = E\left[E\left(\phi(X_{12k}^{(c)}, X_{11s}^{(c)})\cdot\phi(X_{12l}^{(c)}, X_{11t}^{(c)})|X_{121k}^{(c)}, X_{12l}^{(c)}, X_{11t}^{(c)}\right)\right].$$

But,

$$E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})\cdot\phi(X^{(c)}_{12l}, X^{(c)}_{11t})|X^{(c)}_{12k}, X^{(c)}_{12l}, X^{(c)}_{11t}\right) = \phi(X^{(c)}_{12l}, X^{(c)}_{11t})E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})|X^{(c)}_{12k}\right)$$

Now,

$$E(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})|X^{(c)}_{12k}) = E\left\{c(X^{(c)}_{12k} - X^{(c)}_{11s}) - F_{11}(X^{(c)}_{12k})|X^{(c)}_{12k}\right\} - E(\int[c(x - X^{(c)}_{11s}) - F_{11}(x)]dF_{12}(x)|X^{(c)}_{12k}). \tag{3.12}$$

Here the term inside the braces is equal to 0 because

$$E(c(X^{(c)}_{12k} - X^{(c)}_{11s})|X^{(c)}_{12k}) = P(X^{(c)}_{12k} > X^{(c)}_{11s}|X^{(c)}_{12k}) + \frac{1}{2}P(X^{(c)}_{12k} = X^{(c)}_{11s}|X^{(c)}_{12k}) = F_{11}(X^{(c)}_{12k}).$$

By applying Fubini's theorem to the second term in equation (3.12)

$$\begin{aligned} E\left(\int[c(x - X^{(c)}_{11s}) - F_{11}(x)]dF_{12}(x)\right) &= \int\left(\int[c(x - X^{(c)}_{11s}) - F_{11}(x)]dF_{12}(x)\right)dF_{11}(x) \\ &= \int\left(\int[c(x - X^{(c)}_{11s}) - F_{11}(x)]dF_{11}(x)\right)dF_{12}(x) \\ &= \int\left(F_{11}(x) - F_{11}(x)\right)dF_{12}(x) = 0. \end{aligned}$$

For another example, if $k = l \neq (s, t)$,

$$\begin{aligned} E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})\cdot\phi(X^{(c)}_{12l}, X^{(c)}_{11t})\right) &= E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})\cdot\phi(X^{(c)}_{12k}, X^{(c)}_{11t})\right) \\ &= E[E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})\cdot\phi(X^{(c)}_{12k}X^{(c)}_{11t})|X^{(c)}_{12k}, X^{(c)}_{11t}\right)] \end{aligned}$$

Analogous to the previous case,

$$E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})\cdot\phi(X^{(c)}_{12k}, X^{(c)}_{11t})|X^{(c)}_{12k}, X^{(c)}_{11t}\right) = \phi(X^{(c)}_{12k}, X^{(c)}_{11t})E\left(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})|X^{(c)}_{12k}\right).$$

Now,

$$E(\phi(X^{(c)}_{12k}, X^{(c)}_{11s})|X^{(c)}_{12k}) = E(c(X^{(c)}_{12k} - X^{(c)}_{11s}) - F_{11}(X^{(c)}_{12k})|X^{(c)}_{12k}) - E\left(\int[c(x - X^{(c)}_{11s}) - F_{11}(x)]dF_{12}(x)|X^{(c)}_{12k}\right).$$

Here also the first term is 0 since $E(c(X_{12k}^{(c)} - X_{11s}^{(c)})) = F_{11}(X_{12k}^{(c)})$ and, by applying Fubini's theorem, the second term reduces to

$$\begin{aligned} E\left[\int [c(x - X_{11s}^{(c)}) - F_{11}(x)]dF_{12}\right] &= \int \left[\int (c(x - X_{11s}^{(c)}) - F_{11}(x)dF_{12})\right] dF_{11} \\ &= \int \left[\int (c(x - X_{11s}^{(c)}) - F_{11}(x)dF_{11})\right] dF_{12} \\ &= \int (F_{11}(x) - F_{11}(x))dF_{12} = 0. \end{aligned}$$

The other cases of the second term in the equation can similarly be shown to be zero. Thus, putting these pieces together and using Assumption 1 and 2,

$$nE(A_{11}^2) = \frac{n}{(n_{c1} + n_{11})^2(n_{c1} + n_{12})^2} O(n_{1c}^2) = O(\frac{1}{n}).$$

That $nE(A_2)^2$, $nE(A_3)^2$ and $nE(A_4)^2$ are of order $O(n^{-1})$ can be shown along the same lines, and the proof is complete. □

Together with Chebychev's inequality, Theorem 1 says that

$$\sqrt{n}(\widehat{q}_{\mathrm{I},\theta} - q_{\mathrm{I}}) - \sqrt{n}M_\theta \xrightarrow{p} 0, \tag{3.13}$$

which essentially means the asymptotic distribution of $\sqrt{n}(\widehat{q}_{\mathrm{I},\theta} - q_{\mathrm{I}})$ and $\sqrt{n}M_\theta$ is the same.

Theorem 2. *Assume that $\sigma_{cg}^2 = Var(Z_{g1}^{(c)}) > 0$ and $\sigma_{gt}^2 = Var(Y_{gt1}^{(i)}) > 0$ for $g, t = 1, 2$. Under Assumption 1 and 2, the statistic $\sqrt{n}(\hat{q}_{I,\theta} - q_I)$ has, asymptotically, a normal distribution with expectation 0 and variance:*

$$\sigma_\theta^2 = n\left(\frac{\sigma_{c1}^2}{n_{c1}} + \frac{\sigma_{c2}^2}{n_{c2}} + \frac{\sigma_{12}^2}{n_{12}} + \frac{\sigma_{21}^2}{n_{21}} + \frac{\sigma_{11}^2}{n_{11}} + \frac{\sigma_{22}^2}{n_{22}}\right).$$

Proof. In view of (3.13), the desired result in Theorem 2 follows from Slutsky's theorem if we establish that $\sqrt{n}M_\theta$ has asymptotic distribution with mean 0 and variance σ_θ^2. But M_θ is the sum of independent and, by Assumptions 1 and 2, uniformly bounded random variables. Therefore a direct application of the Liapunov's Central Limit Theorem (Billingsley, 1995) leads to the desired result. □

Attention needs to be called to the fact that the asymptotic variance is composed of terms corresponding to the three pieces (one complete and two incomplete) of data in each of the two groups. It is pointed out at the end of Section 3.2 that the asymptotic framework Assumptions 1 and 2 covers includes scenarios where some sample sizes diverge while others remain bounded. In such scenarios, the terms in the asymptotic variance corresponding to the bounded sample sizes falls off and the asymptotic variance estimator in Section 3.4 needs to be adjusted accordingly.

3.4 Estimation of the Asymptotic Variance

In order to be able to use the asymptotic distribution result for statistical inference, we need a consistent estimator of σ_θ^2. From the Weak Law of Large Numbers, we know that

$$\tilde{\sigma}_{cg}^2 \xrightarrow{p} \sigma_{cg}^2 \quad \text{and} \quad \tilde{\sigma}_{gt}^2 \xrightarrow{p} \sigma_{gt}^2$$

for $g, t = 1, 2$ where

$$\tilde{\sigma}_{cg}^2 := \frac{1}{n_{cg}-1}\sum_{k=1}^{n_{cg}}(Z_{gk}^{(c)} - \bar{Z}_{g.}^{(c)})^2 \quad \text{and} \quad \tilde{\sigma}_{gt}^2 := \frac{1}{n_{gt}-1}\sum_{k=1}^{n_{gt}}(Y_{gtk}^{(i)} - \bar{Y}_{gt.}^{(i)})^2.$$

However, $\tilde{\sigma}_{cg}^2$ and $\tilde{\sigma}_{gt}^2$ cannot directly be used to construct a consistent estimator of σ_θ^2 because they are defined in terms of unobservable random variables $Z_{gk}^{(c)}$ and $Y_{gtk}^{(i)}$. In the following, we will show that it is possible to obtain a consistent estimator by replacing these random variables with their analogs defined in terms of the empirical distribution functions. To that end, first define

$$\tilde{\sigma}_\theta^2 := n\left(\frac{\tilde{\sigma}_{c1}^2}{n_{c1}} + \frac{\tilde{\sigma}_{c2}^2}{n_{c2}} + \frac{\tilde{\sigma}_{12}^2}{n_{12}} + \frac{\tilde{\sigma}_{21}^2}{n_{21}} + \frac{\tilde{\sigma}_{11}^2}{n_{11}} + \frac{\tilde{\sigma}_{22}^2}{n_{22}}\right), \tag{3.14}$$

where for $g, t = 1, 2$. Obviously, under Assumptions 1 and 2, $\tilde{\sigma}_\theta^2 \xrightarrow{p} \sigma_\theta^2$. Let $\hat{Z}_{1k}^{(c)}, \hat{Z}_{2k}^{(c)}, \hat{Y}_{11k}^{(i)}, \hat{Y}_{12k}^{(i)}, \hat{Y}_{21k}^{(i)}$ and $\hat{Y}_{22k}^{(i)}$ be defined analogous to $Z_{1k}^{(c)}, Z_{2k}^{(c)}, Y_{11k}^{(i)}, Y_{12k}^{(i)}, Y_{21k}^{(i)}$ and $Y_{22k}^{(i)}$, respectively, by replacing F_{gt} by $\hat{F}_{gt,\theta}$ for all $g, t = 1, 2$. Further, define

$$\hat{\sigma}_{cg}^2 := \frac{1}{n_{cg}-1}\sum_{k=1}^{n_{cg}}(\hat{Z}_{gk}^{(c)} - \bar{\hat{Z}}_{g.}^{(c)})^2 \quad \text{and} \quad \hat{\sigma}_{gt}^2 := \frac{1}{n_{gt}-1}\sum_{k=1}^{n_{gt}}(\hat{Y}_{gtk}^{(i)} - \bar{\hat{Y}}_{gt.}^{(i)})^2. \tag{3.15}$$

It is clear that the proof of $\hat{\sigma}_\theta^2$ being a consistent estimator for σ_θ^2 will be complete if we show $\hat{\sigma}_\theta^2 - \tilde{\sigma}_\theta^2 \xrightarrow{p} 0$. This is claimed and proved in Theorem 3. Essentially, Theorem 3 shows that $\hat{\sigma}_\theta^2$ is a consistent estimator of σ_θ^2 under the null as well as alternative hypotheses.

Theorem 3. *Define* $\hat{\sigma}_\theta^2 := n\left(\frac{\hat{\sigma}_{c1}^2}{n_{c1}} + \frac{\hat{\sigma}_{c2}^2}{n_{c2}} + \frac{\hat{\sigma}_{12}^2}{n_{12}} + \frac{\hat{\sigma}_{21}^2}{n_{21}} + \frac{\hat{\sigma}_{11}^2}{n_{11}} + \frac{\hat{\sigma}_{22}^2}{n_{22}}\right)$. *Then under Assumptions 1 and 2 and* $\sigma_\theta^2 > 0$,
$\frac{\hat{\sigma}_\theta^2}{\sigma_\theta^2} \xrightarrow{L_2} 1.$

Proof. In order to show that contribution of the variance from the complete cases are asymptotically unbiased, it suffices to show that $E[\tilde{\sigma}^2_{gt} - \hat{\sigma}^2_{gt}]^2$ goes to 0 as $n \to \infty$. Using the definition of variances in equation (3.15)

$$
\begin{aligned}
E[\tilde{\sigma}^2_{gt} - \hat{\sigma}^2_{gt}]^2 &= E\left[\frac{1}{n_{gt}-1}\sum_{k=1}^{n_{gt}}(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.})^2 - \frac{1}{n_{gt}-1}\sum_{k=1}^{n_{gt}}(\hat{Y}^{(i)}_{gtk} - \bar{\hat{Y}}^{(i)}_{gt.})^2\right]^2 \\
&= \frac{1}{(n_{gt}-1)^2} E\left[\sum_{k=1}^{n_{gt}}(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.})^2 - (\hat{Y}^{(i)}_{gtk} - \bar{\hat{Y}}^{(i)}_{gt.})^2\right]^2 \\
&= \frac{1}{(n_{gt}-1)^2} E\left[\sum_{k=1}^{n_{gt}}(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.} - \hat{Y}^{(i)}_{gtk} + \bar{\hat{Y}}^{(i)}_{gt.})(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.} + \hat{Y}^{(i)}_{gtk} - \bar{\hat{Y}}^{(i)}_{gt.})\right]^2 \\
&\le \frac{1}{(n_{gt}-1)^2} E\left[\sum_{k=1}^{n_{gt}}(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.} - \hat{Y}^{(i)}_{gtk} + \bar{\hat{Y}}^{(i)}_{gt.})^2 \sum_{k=1}^{n_{gt}} \underbrace{(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.} + \hat{Y}^{(i)}_{gtk} - \bar{\hat{Y}}^{(i)}_{gt.})^2}_{\le 4}\right] \\
&\le \frac{4n_{gt}}{(n_{gt}-1)^2} E\left[\sum_{k=1}^{n_{gt}}(Y^{(i)}_{gtk} - \bar{Y}^{(i)}_{gt.} - \hat{Y}^{(i)}_{gtk} + \bar{\hat{Y}}^{(i)}_{gt.})^2\right] \\
&\le \frac{4n_{gt}}{(n_{gt}-1)^2} E\left[\sum_{k=1}^{n_{gt}}[(Y^{(i)}_{gtk} - \hat{Y}^{(i)}_{gtk}) - (\bar{Y}^{(i)}_{gt.} - \bar{\hat{Y}}^{(i)}_{gt.})]^2\right] \\
&= \frac{4n_{gt}}{(n_{gt}-1)^2} E\left[\sum_{k=1}^{n_{gt}}(Y^{(i)}_{gtk} - \hat{Y}^{(i)}_{gtk})^2 - \underbrace{n_{gt}(\bar{Y}^{(i)}_{gt.} - \bar{\hat{Y}}^{(i)}_{gt.})^2}_{\ge 0}\right] \\
&\le \frac{4n_{gt}}{(n_{gt}-1)^2} \sum_{k=1}^{n_{gt}} E(Y^{(i)}_{gtk} - \hat{Y}^{(i)}_{gtk})^2 \\
&= 4\left(\frac{n_{gt}}{n_{gt}-1}\right)^2 E(Y^{(i)}_{gt1} - \hat{Y}^{(i)}_{gt1})^2 \\
&= 4\left(\frac{n_{gt}}{n_{gt}-1}\frac{n_{gt}}{2m_{gt}}\right)^2 E\left[\{F_{ls}(Y^{(i)}_{gt1}) - \hat{F}_{ls,\theta_{ls}}(Y^{(i)}_{gt1})\} + \{F_{sl}(Y^{(i)}_{gt1}) - \hat{F}_{sl,\theta_{sl}}(Y^{(i)}_{gt1})\}\right]^2; \\
& l \neq g \quad or \quad s \neq t \\
&\le 2\left(\frac{n_{gt}}{n_{gt}-1}\frac{n_{gt}}{m_{gt}}\right)^2 \{E\left[F_{ls}(Y^{(i)}_{gt1}) - \hat{F}_{ls,\theta_{ls}}(Y^{(i)}_{gt1})\right]^2 + E\left[F_{sl}(Y^{(i)}_{gt1}) - \hat{F}_{sl,\theta_{sl}}(Y^{(i)}_{gt1})\right]^2\} \\
&= O(\frac{1}{m_{ls}} + \frac{1}{m_{sl}}).
\end{aligned}
$$

□

The estimator $\hat{\sigma}^2_\theta$ can be calculated in terms of ranks. To show that, it can easily be verified that

$$
\hat{F}_{ls,\theta_{ls}}(X^{(w)}_{gtk}) = \frac{1}{m_{ls}}\left(R^{(w)}_{ls.gtk} - R^{(w,gt)}_{gtk}\right),
$$

where one of (g, s, t, l) is different from the other three and $w \in \{c, i\}$. Here, $R_{gtk}^{(w)}$ and $R_{gtk}^{(w,gt)}$ are as defined in Proposition 2. Using these in (3.15), the rank expression for $\hat{\sigma}_\theta^2$ follows immediately.

3.5 Test Procedures and Confidence Intervals

Recall that in equation (3.2), the quantity $p_I := (p_{12} - p_{11}) - (p_{22} - p_{21})$ is defined as interaction effect. So, $p_I = 0$ means that there is no interaction effect, which in turn is equivalent to saying $q_I = 1$. Therefore, testing the hypothesis of no interaction effect can be formulated as

$$H_0 : q_I = 1 \quad \text{versus} \quad H_1 : q_I \neq 1.$$

The statistic $T_{M,\theta}$ (given below) and its limit distribution, under Assumptions 1 and 2,

$$T_{M,\theta} = \sqrt{n}\frac{(\hat{q}_{I,\theta} - q_I)}{\hat{\sigma}_\theta} \xrightarrow{D} N(0, 1) \tag{3.16}$$

can be used to test this hypothesis. The asymptotic $(1 - \alpha)100\%$ confidence interval for q_{I} is obtained from

$$P\left(\hat{q}_{I,\theta} - \frac{z_{\alpha/2}\hat{\sigma}_\theta}{\sqrt{n}} \leq q_I \leq \hat{q}_{I,\theta} + \frac{z_{\alpha/2}\hat{\sigma}_\theta}{\sqrt{n}}\right) \to 1 - \alpha, \tag{3.17}$$

which also holds under Assumptions 1 and 2. In practical applications, small sample sizes may arise for which use of the asymptotic results may not control the sizes of the tests and coverage probability of the confidence intervals at the desired level. We apply a Box-Type approximation of degrees of freedom as was done in Brunner et al. (1997) for nonparametric factorial designs. These approximations are obtained by matching the first two moments of a chi-square distribution with that of the estimated asymptotic variance (see also Konietschke et al., 2012a). Consequently, the distribution of our test statistic, $T_{M,\theta}$, for small sample sizes can be approximated by a central $t(\nu)$ – distribution, where the degree of freedom ν is given by

$$\nu = \frac{\left(\frac{\hat{\sigma}_{c1}^2}{n_{c1}} + \frac{\hat{\sigma}_{11}^2}{n_{11}} + \frac{\hat{\sigma}_{21}^2}{n_{21}} + \frac{\hat{\sigma}_{c2}^2}{n_{c2}} + \frac{\hat{\sigma}_{12}^2}{n_{12}} + \frac{\hat{\sigma}_{22}^2}{n_{22}}\right)^2}{\left(\frac{[\hat{\sigma}_{c1}^2/n_{c1}]^2}{n_{c1}-1} + \frac{[\hat{\sigma}_{11}^2/n_{11}]^2}{n_{11}-1} + \frac{[\hat{\sigma}_{21}^2/n_{21}]^2}{n_{21}-1} + \frac{[\hat{\sigma}_{c2}^2/n_{c2}]^2}{n_{c2}-1} + \frac{[\hat{\sigma}_{12}^2/n_{12}]^2}{n_{12}-1} + \frac{[\hat{\sigma}_{22}^2/n_{22}]^2}{n_{22}-1}\right)}. \tag{3.18}$$

A small sample version of the confidence interval in (3.17) can be obtained from

$$P\left(\hat{q}_{I,\theta} - \frac{t_{\alpha/2}^{(\nu)}\hat{\sigma}_\theta}{\sqrt{n}} \leq q_I \leq \hat{q}_{I,\theta} + \frac{t_{\alpha/2}^{(\nu)}\hat{\sigma}_\theta}{\sqrt{n}}\right) \approx 1 - \alpha. \tag{3.19}$$

Although both intervals in (3.17) and (3.19) might not preserve the range of the interval, a suitable transformation discussed by Brunner et al. (2002) could be applied.

The density plot in Figure 3.3 depicts the empirical distribution of our test statistic in equation 3.16. As it can be seen from the figure the density plot of $T_{M,\theta}$ approximately overlaps with that of standard normal curve, and this is in agreement with the theoretical derivation of the test statistic.

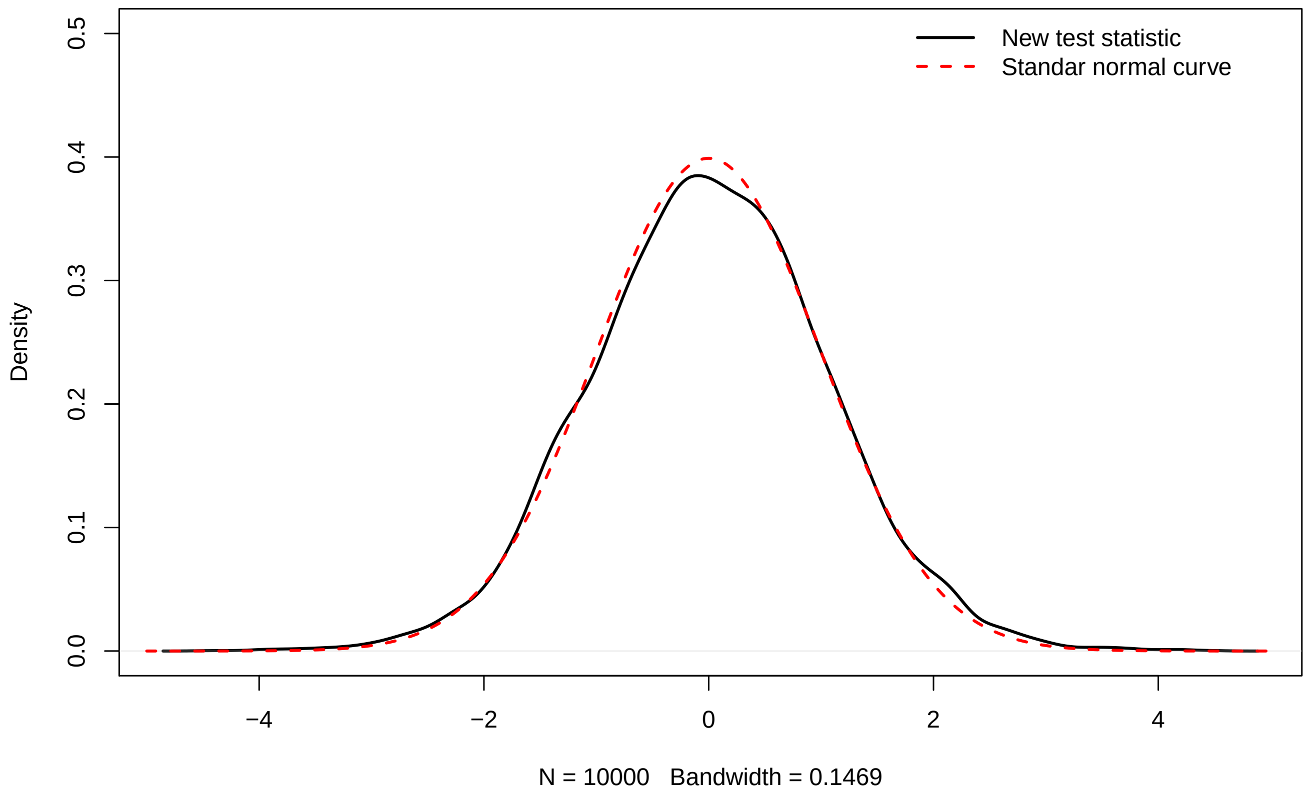


FIGURE 3.3: Empirical distribution of the test statistic ($T_{M,\theta}$) based on the simulation of size 10,000. The theoretical value of interaction effect, q_I=1 is used.

3.6 Numerical Examples

3.6.1 Simulation Study

In this subsection we investigate the numerical accuracy of the asymptotic results in Sections 3.3 and 3.4 via a simulation study. We also use simulation studies to perform numerical comparisons of the nonparametric test proposed in this study with parametric tests for the same problem developed elsewhere. We carry out comparisons in terms of size and power of the tests. In all the simulation studies, the run size is $10,000$. The methods compared are:

$T_{\mathrm{M},\theta}$: The new test statistic $T_{M,\theta}$ with the asymptotic distribution as shown in (3.16).

$T^s_{\mathrm{M},\theta}$: The new test statistic $T_{M,\theta}$ where the null distribution is approximated by a t-distribution with degrees of freedom given by (3.18).

T_{XH}: The parametric test in Xu and Harrar (2012) where the null distribution is approximated by second order approximation derived in their paper.

T^s_{XH}: The parametric test in Xu and Harrar (2012) where the null distribution is approximated by a t-distribution with degrees of freedom obtained in a way similar to (3.18).

T: The parametric paired sample t-test applied to the complete cases only.

These methods will be compared with respect to their actual sizes and powers by setting the nominal size to $\alpha = 0.05$.

Size Simulation

We numerically compare achieved sizes of the tests under various settings for missing percentage (MP), the distribution of data, sample size allocations (n_1 and n_2) and within-pair dependence. The effect of missing percentage is evaluated by varying MP in $\{0\%, 5\%, 10\%, 25\%\}$. Data will be generated from Normal, Cauchy and Lognormal distributions which represent light-tailed, heavy-tailed and skewed, distributions, respectively. For the paired data, we use a scale matrix with unity on the diagonals and ρ_g on the off diagonals for group $g = 1, 2$. For the unpaired cases, we set the scales to 1. For the lognormal distribution, the the scale matrix is the covariance of the variables in logarithmic scale. Therefore, for both normal and lognormal, the off diagonal elements of the scale matrix are essentially correlations and a value of zero implies independence. In order to investigate the effect of sample size allocation as the degrees of within-pair dependence varies, we allow equal and unequal values for the dependence parameter with $\rho_1, \rho_2 \in \{-0.5, 0, 0.85\}$ and pairing them positively as well as negatively with the sample size allocations $(n_1, n_2) \in \{(10, 10), (5, 15), (20, 20), (15, 25), (30, 30), (25, 35)\}$.

	(n_1,n_2)	%	$T_{M,\theta}$			$T^s_{M,\theta}$			T_{XH}			T^s_{XH}			T		
ρ_1			-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85
ρ_2			-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85
Normal	(10,10)	0	7.7	6.7	3.8	6.1	5.3	2.7	6.8	6.8	6.7	5.0	5.2	4.9	4.9	5.1	4.9
		5	8.2	7.7	4.8	6.3	5.7	3.0	7.4	7.5	8.6	5.3	5.4	6.4	5.1	5.2	5.3
		10	8.6	8.0	5.8	6.8	6.1	3.6	7.6	7.9	9.8	5.8	5.9	7.3	5.4	5.4	5.5
		25	8.9	8.1	7.1	7.1	5.9	4.5	7.9	7.9	11	5.9	5.8	8.3	5.0	4.9	5.0
	(5,15)	0	11.3	9.5	5.6	7.9	4.4	4.5	8.8	8.7	8.9	8.0	7.8	7.9	5.4	5.6	5.6
		5	11.8	10.1	6.5	7.9	5.4	3.7	9.2	9.1	10.5	8.5	8.4	9.6	5.7	5.5	5.1
		10	12.2	10.6	7.7	8.4	5.9	4.3	9.6	9.5	11.8	8.9	8.6	10.9	5.9	5.6	5.3
		25	13.7	12.0	9.8	10.0	7.1	6.3	10.6	10.5	14.0	9.7	9.6	13.0	5.9	5.9	5.8
	(20,20)	0	6.5	5.9	4.0	5.6	5.2	3.3	5.9	5.6	5.7	5.0	5.0	5.0	5.0	5.1	5.0
		5	6.5	6.3	5.9	5.5	5.4	4.	5.9	5.1	6.8	4.9	4.9	6.0	4.8	4.9	4.9
		10	6.5	6.5	5.6	5.5	5.4	4.7	6.0	6.0	7.4	5.0	5.0	6.3	5.0	4.9	4.5
		25	6.6	6.3	5.5	5.7	5.3	4.7	6.1	6.0	6.5	5.2	4.9	5.4	5.1	4.9	4.8
	(15,25)	0	6.6	5.9	4.2	5.7	5.1	3.6	6.2	6.1	6.1	5.6	5.8	5.6	5.2	5.3	5.1
		5	6.7	6.1	4.9	5.9	5.3	4.2	6.2	6.0	7.1	5.7	5.5	6.6	5.0	5.0	5.1
		10	6.8	6.3	5.3	5.9	5.4	4.0	6.2	6.1	7.3	5.6	5.4	5.9	5.1	5.1	5.0
		25	6.7	6.3	5.9	6.0	5.4	4.2	6.2	6.3	7.5	5.5	5.6	5.3	4.8	5.0	4.8
	(30,30)	0	5.8	5.6	4.4	5.4	5.0	3.9	5.6	5.4	5.5	5.2	5.1	5.0	5.2	5.1	5.0
		5	6.0	5.7	5.2	5.5	5.2	4.8	5.5	5.6	6.4	5.0	5.1	6.0	5.0	5.1	5.0
		10	6.1	5.8	5.5	5.5	5.2	4.3	5.7	5.6	6.7	5.2	5.2	4.9	5.0	5.2	4.9
		25	6.2	6.0	5.4	5.6	5.4	4.6	5.8	5.5	6.0	5.2	5.1	4.6	5.0	4.9	4.9
	(25,35)	0	5.6	5.6	4.2	5.2	4.9	3.8	5.2	5.5	5.4	5.0	5.3	5.1	4.8	5.1	4.9
		5	6.1	5.6	5.1	5.5	5.0	4.6	5.7	5.6	6.6	5.4	5.2	6.0	5.1	5.1	4.5
		10	6.1	5.6	5.5	5.5	5.0	4.8	5.8	5.7	6.8	5.4	5.3	6.2	5.1	5.0	4.8
		25	6.0	5.7	5.8	5.3	5.1	4.9	5.7	5.6	6.4	5.0	4.9	5.3	5.2	4.9	5.0

TABLE 3.2: Achieved size I at $\alpha = 0.05$ by the new method ($T_{M,\theta}$), the new method with t approximation ($T^s_{M,\theta}$), the method of Xu and Harrar (2012) (T_{XH}), the method of Xu and Harrar (2012) with t approximation (T^s_{XH}) and complete-case t test (T). Data in the two groups are generated from normal distribution with equal correlation. The simulation size is $10,000$.

In Tables 3.2 – 3.7, the simulation results for the size of the test are shown. From Tables 3.2 and 3.3, under normality, we see complete-case analysis based on paired t-test controls the size of the test most accurately. The other parametric methods also perform reasonably well for medium and larger sample sizes whence the new nonparametric method catches up as the sample sizes increase. The effect of increasing missing percentage (MP) appears to be pronounced when the amount of the dependence in the two groups is different and the sample sizes are smaller.

	(n_1, n_2)	%	$T_{M,\theta}$			$T^s_{M,\theta}$			T_{XH}			T^s_{XH}			T		
ρ_1			-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0
ρ_2			0	0.85	0.85	0	0.85	0.85	0	0.85	0.85	0	0.85	0.85	0	0.85	0.85
Normal	(5,15)	0	11.9	13.6	10.8	7.3	7.8	5.5	9.2	11.7	10.6	8.8	10.9	10.0	5.7	5.4	5.7
		5	12.7	14.7	12.1	8.4	10.6	7.3	10.1	12.5	12.0	9.1	11.6	11.0	5.8	5.7	5.7
		10	13.1	15.2	13.0	8.8	11.1	8.0	10.4	13.0	12.6	9.5	12.1	11.5	6.1	5.8	5.9
		25	14.7	17.3	14.6	10.2	13.4	9.8	11.6	14.4	13.7	10.7	13.3	12.6	6.2	6.2	6.3
	(15,5)	0	9.2	5.7	5.1	4.7	4.0	3.1	8.4	7.3	7.1	3.3	0.6	1.2	5.6	5.5	5.2
		5	9.6	6.1	5.8	4.6	4.8	4.5	8.8	7.3	7.7	3.7	0.1	0.3	5.7	5.0	5.1
		10	10.1	6.6	6.5	5.1	5.1	5.0	9.3	7.8	8.3	4.0	0.1	0.3	5.6	5.1	5.1
		25	11.9	8.1	8.2	6.2	6.5	6.6	10.2	9.2	9.6	4.8	0.2	0.7	5.8	4.7	4.6
	(15,25)	0	6.5	7.5	6.6	5.1	5.7	4.9	6.2	6.9	6.7	5.5	6.2	6.1	5.1	5.1	5.1
		5	6.7	7.8	6.9	5.8	6.2	5.3	6.4	7.2	6.2	5.9	6.5	5.6	5.0	5.2	5.0
		10	7.0	8.0	6.8	6.0	6.5	5.6	6.3	7.3	6.9	5.7	6.3	6.0	5.1	5.2	5.2
		25	6.9	7.9	6.8	5.9	6.3	5.5	6.5	7.1	6.8	5.8	6.2	5.6	4.9	5.1	5.1
	(25,15)	0	6.0	5.5	5.2	5.3	4.6	4.2	5.7	5.8	5.8	4.3	3.4	3.5	4.9	4.8	4.9
		5	6.1	5.9	5.5	5.4	5.0	4.6	5.7	6.0	6.2	4.6	3.8	3.8	4.8	4.7	4.6
		10	6.2	6.0	5.8	5.5	5.1	4.8	5.9	6.1	6.2	4.7	4.0	4.2	4.7	4.8	4.6
		25	6.4	6.5	6.0	5.4	5.6	5.1	6.0	6.2	6.1	4.0	2.9	3.3	4.8	4.9	4.8
	(25,35)	0	5.8	6.3	5.6	5.1	5.3	4.7	5.2	5.9	5.6	4.9	5.4	5.1	4.7	4.9	4.7
		5	6.2	6.7	6.2	5.4	5.8	5.3	5.7	6.5	6.2	5.2	5.8	5.6	5.0	5.0	5.0
		10	6.1	6.7	6.2	5.4	5.4	5.3	5.6	6.3	6.2	5.3	5.7	5.8	4.9	4.9	4.8
		25	6.2	6.5	6.1	5.4	5.4	5.2	5.7	6.1	6.1	5.2	5.5	5.4	5.0	5.0	5.0
	(35,25)	0	5.4	5.4	5.1	5.0	4.8	4.6	5.3	5.4	5.5	4.6	4.5	4.6	4.8	5.0	4.9
		5	5.9	6.0	5.7	5.3	5.5	5.1	5.6	6.0	6.0	4.8	4.5	4.6	5.2	5.0	5.1
		10	5.9	5.8	5.8	5.4	5.2	5.1	5.6	5.8	5.8	4.9	4.6	4.6	5.3	5.1	5.2
		25	6.0	6.1	5.8	5.4	5.5	5.4	5.8	6.0	6.0	4.9	4.8	4.9	5.1	5.5	5.4

TABLE 3.3: Achieved size I at $\alpha = 0.05$ by the new method ($T_{M,\theta}$), the new method with t approximation ($T^s_{M,\theta}$), the method of Xu and Harrar (2012) (T_{XH}), the method of Xu and Harrar (2012) with t approximation (T^s_{XH}) and complete-case t test (T). Data in the two groups are generated from normal distribution with unequal correlation. The simulation size is $10,000$.

When data are generated from a Cauchy (heavy-tailed) distribution (Tables 3.4 and 3.5), the parametric methods have unacceptably conservative behavior, more so when the amount of dependence in the two groups is different and when larger samples are drawn from the group in which the dependence is negative.

	(n_1,n_2)	%	$T_{M,\theta}$			$T^s_{M,\theta}$			T_{XH}			T^s_{XH}			T		
ρ_1			-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85
ρ_2			-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85
Cauchy	(10,10)	0	7.4	7.2	3.5	5.9	5.4	2.1	3.0	3.3	3.4	1.8	1.8	1.7	1.7	1.9	2.0
		5	8.1	7.4	4.9	6.2	6.1	3.3	4.6	4.6	6.7	3.0	2.6	4.0	1.9	1.7	1.8
		10	8.3	7.9	5.9	6.5	6.2	4.2	5.5	5.5	8.4	3.8	3.5	5.7	1.9	1.8	1.7
		25	8.6	8.2	8.5	6.1	6.2	4.9	6.1	6.3	9.4	3.9	3.7	6.9	1.8	1.8	1.6
	(5,15)	0	10.2	8.5	4.2	6.7	4.0	1.4	3.2	3.2	3.1	2.7	2.5	2.6	1.6	1.5	1.6
		5	10.9	9.7	6.2	6.2	2.9	2.4	4.3	5.0	7.2	2.3	4.3	5.2	1.6	1.6	1.8
		10	11.1	7.9	5.9	6.4	6.2	4.2	5.2	5.5	8.4	3.1	3.5	5.7	1.6	1.8	1.7
		25	12.9	8.2	7.5	7.7	6.2	4.9	6.8	6.3	9.4	4.5	3.7	6.9	1.6	1.8	1.6
	(20,20	0	6.0	5.8	4.1	5.3	5.1	3.3	2.5	2.5	2.7	2.0	1.9	2.1	2.0	1.9	2.1
		5	6.7	6.4	5.3	5.8	5.5	4.4	3.5	3.7	5.2	2.8	3.0	4.6	1.9	2.1	2.0
		10	6.6	6.4	5.8	5.7	5.4	4.7	3.7	4.0	5.7	1.2	0.9	0.7	1.9	2.0	2.0
		25	6.8	6.3	6.3	5.8	5.3	4.3	3.3	3.4	4.5	1.3	1.0	1.0	1.8	1.8	1.8
	(15,25)	0	6.4	6.2	3.9	5.5	5.1	2.9	2.7	2.8	2.6	2.3	2.4	2.1	2.0	2.2	1.8
		5	6.3	6.2	5.4	5.5	5.4	4.5	3.4	3.5	5.0	2.6	2.7	4.3	1.8	1.7	1.9
		10	6.6	6.4	5.8	5.7	5.4	4.7	3.7	4.0	5.7	1.2	0.9	0.7	1.9	2.0	2.0
		25	6.8	6.3	6.3	5.8	5.3	4.3	3.3	3.4	4.5	4.5	1.3	1.0	1.8	1.8	1.8
	(30,30)	0	5.4	5.3	4.3	4.9	4.8	3.7	2.3	2.3	2.3	2.0	1.8	2.0	2.0	1.8	1.9
		5	5.8	5.7	5.3	5.2	5.3	4.7	2.9	3.0	4.0	2.5	2.6	3.5	1.7	1.8	1.9
		10	5.8	5.7	5.6	5.1	5.2	5.0	3.1	3.0	4.2	2.6	2.6	3.6	1.8	1.9	2.0
		25	5.8	5.8	5.1	5.1	5.1	4.4	2.5	2.4	2.9	0.3	0.3	0.3	1.7	1.7	1.9
	(25,35)	0	5.6	5.7	4.7	5.1	5.2	4.1	2.4	2.3	2.5	2.1	2.0	2.0	1.9	1.9	2.0
		5	5.9	5.7	5.2	5.3	5.3	3.3	3.1	3.2	4.5	2.0	2.5	1.0	1.8	2.1	2.1
		10	5.8	5.7	5.6	5.1	5.2	5.0	3.1	3.0	4.2	2.6	2.6	3.6	1.9	1.8	2.0
		25	5.8	5.8	5.1	5.1	5.1	4.4	2.5	2.4	2.9	0.3	0.3	0.3	1.7	1.7	1.9

TABLE 3.4: Achieved size I at $\alpha = 0.05$ by the new method ($T_{M,\theta}$), the new method with t approximation ($T^s_{M,\theta}$), the method of Xu and Harrar (2012) (T_{XH}), the method of Xu and Harrar (2012) with t approximation (T^s_{XH}) and complete-case t test (T). Data in the two groups are generated from Cauchy distribution with equal correlation. The simulation size is $10,000$.

The new method shows a clear edge in performance for the heavy-tailed distribution where the quality of approximation progressively improves as sample sizes increase. Even for smaller samples, this method attain the desired level ($\alpha = 0.05$) when larger sample are from the group that has negatively correlated pairs.

	(n_1, n_2)	%	$T_{M,\theta}$			$T^s_{M,\theta}$			T_{XH}			T^s_{XH}			T		
ρ_1			-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0
ρ_2			0	0.85	0.85	0	0.85	0.85	0	0.85	0.85	0	0.85	0.85	0	0.85	0.85
Cauchy	(5,15)	0	11.1	13.0	10.0	6.4	6.6	5.1	3.5	4.6	4.3	2.9	4.0	1.8	1.8	2.0	1.8
		5	11.8	14.4	11.2	6.4	7.3	4.8	4.7	6.6	6.4	2.2	2.3	2.4	1.5	2.0	1.9
		10	12.3	14.8	11.8	6.8	7.9	5.4	5.5	7.5	7.8	3.1	3.1	3.2	1.7	2.0	1.9
		25	14.0	16.8	13.7	8.6	9.9	7.2	7.0	9.2	9.5	4.4	4.5	4.7	1.7	2.4	2.3
	(15,5)	0	8.0	5.0	4.5	6.7	3.4	2.0	3.4	2.8	3.0	1.9	1.8	1.6	1.6	1.8	1.6
		5	8.7	5.8	5.4	5.6	4.1	3.9	4.8	4.5	4.8	3.3	3.1	3.4	1.7	1.7	1.8
		10	9.2	6.3	6.1	4.3	4.0	3.8	5.5	5.5	5.8	4.0	3.9	4.2	1.7	1.7	1.7
		25	10.6	8.0	8.1	3.7	2.8	2.9	6.9	7.3	7.6	5.3	5.6	5.9	1.6	1.5	1.5
	(15,25)	0	6.7	7.4	6.8	5.5	5.8	5.3	2.8	3.0	2.6	2.3	2.4	2.4	2.1	2.1	2.0
		5	6.2	7.4	6.7	5.5	5.8	5.4	3.4	4.0	4.2	2.5	3.4	3.6	1.6	1.8	1.8
		10	6.4	7.4	7.0	5.6	6.0	5.5	3.6	4.5	4.5	2.8	3.8	3.8	1.7	1.8	1.8
		25	6.3	7.4	6.8	5.4	6.0	5.4	3.5	4.1	4.2	1.5	1.5	1.5	1.6	2.1	1.9
	(25,15)	0	6.0	5.5	5.5	5.3	4.6	4.8	5.7	5.8	2.9	4.3	3.4	1.7	4.9	4.8	2.2
		5	6.1	5.9	5.9	5.4	5.0	5.1	5.7	6.0	4.0	4.6	3.8	2.2	4.8	4.7	1.8
		10	6.2	6.0	6.1	5.5	5.1	5.1	5.9	6.1	4.6	4.7	4.0	2.9	4.7	4.8	1.8
		25	6.4	6.5	6.0	5.4	5.6	4.6	6.0	6.2	3.7	4.0	2.9	2.5	4.8	4.9	1.8
	(25,35)	0	6.0	6.4	6.3	5.3	5.4	5.3	2.4	2.5	2.6	2.2	2.2	2.4	1.9	2.0	2.1
		5	6.0	6.4	6.0	5.3	5.3	5.1	3.1	3.5	3.7	2.0	2.0	2.3	2.0	2.0	2.0
		10	6.3	6.3	6.0	5.6	5.3	5.2	3.1	3.4	3.6	2.3	2.4	2.7	2.1	1.9	1.9
		25	6.4	6.4	6.0	5.6	5.2	4.9	2.9	3.0	2.8	1.9	1.5	1.7	2.0	2.1	2.1
	(35,25)	0	5.4	5.4	5.1	5.0	4.8	4.5	5.3	5.4	2.5	4.6	4.5	1.8	4.8	5.0	2.0
		5	5.9	6.0	5.5	5.3	5.5	4.9	5.6	6.0	3.7	4.8	4.5	2.9	5.2	5.0	2.1
		10	5.9	5.8	5.7	5.4	5.2	5.1	5.6	5.8	3.6	4.9	4.6	3.0	5.3	5.1	1.9
		25	6.0	6.1	5.7	5.4	5.5	5.0	5.8	6.0	3.0	4.9	4.8	2.5	5.1	5.5	1.9

TABLE 3.5: Achieved size I at $\alpha = 0.05$ by the new method ($T_{M,\theta}$), the new method with t approximation ($T^s_{M,\theta}$), the method of Xu and Harrar (2012) (T_{XH}), the method of Xu and Harrar (2012) with t approximation (T^s_{XH}) and complete-case t test (T). Data in the two groups are generated from Cauchy distribution with unequal correlation. The simulation size is $10,000$.

Looking at Tables 3.6 and 3.7, although the performance of the complete-case paired t-test is not as bad as it is for the heavy-tailed distribution, it is still far too low to be considered acceptable.

On the contrary, the second-order asymptotic method of Xu and Harrar (2012) performs well . This is expected because the second-order asymptotic explicitly corrects for the skewness in the data. On the other hand, the new nonparametric method has very good performance under skewness as well. We also note that the performance of the new method under skewness (Tables 3.6 and 3.7) and heavy tail (Tables 3.4 and 3.5) is somewhat similar.

	(n_1, n_2)	%	$T_{M,\theta}$			$T^s_{M,\theta}$			T_{XH}			T^s_{XH}			T		
ρ_1			-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85
ρ_2			-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85	-0.5	0	0.85
Lognormal	(10,10)	0	7.4	6.7	3.9	5.9	5.1	2.6	5.3	5.1	5.2	3.6	3.0	3.3	3.3	3.3	3.1
		5	7.5	6.9	5.0	5.7	5.3	3.6	5.7	5.6	7.1	3.6	3.9	4.3	3.6	3.1	3.0
		10	7.7	6.9	5.8	5.8	4.7	4.1	6.1	6.0	8.8	4.5	2.4	6.6	3.5	3.3	3.3
		25	13.7	12.0	9.8	10.0	7.1	6.3	9.0	8.5	10.9	8.0	7.7	10.	3.5	3.2	2.8
	(5,15)	0	10.8	9.6	5.3	6.3	5.7	3.5	6.3	6.0	5.5	5.1	5.3	3.5	3.5	3.3	2.7
		5	11.8	10.1	6.5	7.9	5.4	3.7	7.6	7.2	7.8	6.7	6.3	6.9	3.7	3.4	2.8
		10	12.2	10.6	7.7	8.4	5.9	4.3	8.2	7.7	9.2	7.3	6.8	8.2	3.6	3.2	2.9
		25	13.7	12.0	9.8	10.0	7.1	6.3	9.0	8.5	10.9	8.0	7.7	10.0	3.5	3.2	2.8
	(20,20	0	6.2	5.9	3.8	5.5	5.2	3.2	5.4	4.9	4.6	4.6	4.0	3.6	4.5	3.9	3.6
		5	6.5	6.0	5.4	5.6	5.1	4.6	5.4	5.0	6.1	4.3	3.9	5.1	4.2	3.8	3.7
		10	6.0	5.8	5.8	5.2	4.9	5.0	5.4	5.3	6.1	3.8	4.5	5.2	4.2	4.1	3.8
		25	6.6	6.3	5.2	5.8	5.5	3.9	5.2	5.2	5.7	4.4	4.0	4.3	3.9	3.8	3.4
	(15,25)	0	6.6	5.9	4.0	5.8	5.0	5.3	5.5	4.6	4.5	5.1	4.2	4.0	4.5	3.7	3.6
		5	6.5	6.3	5.0	5.5	5.4	4.3	5.1	5.4	6.3	4.6	4.9	5.7	4.1	4.2	3.7
		10	6.5	7.0	5.9	5.1	5.3	4.8	5.7	6.1	6.8	5.0	5.5	5.9	4.2	4.3	3.6
		25	6.7	6.3	5.9	6.0	5.4	4.2	5.5	5.3	6.0	4.6	4.6	1.2	4.0	3.8	3.3
	(30,30)	0	5.7	5.7	4.5	5.2	5.2	4.0	5.1	4.8	4.3	4.5	4.2	3.9	4.6	4.2	3.8
		5	5.8	5.6	5.3	5.3	5.2	4.8	5.0	4.8	5.6	4.1	4.2	4.9	4.2	4.0	3.8
		10	6.2	5.8	5.4	5.7	5.3	4.8	5.3	5.2	5.4	4.5	4.5	4.7	4.6	4.4	4.0
		25	6.0	5.4	5.6	5.4	4.9	4.5	5.4	5.1	4.9	4.3	4.4	4.0	4.6	4.5	4.1
	(25,35)	0	6.0	5.5	4.0	5.4	4.9	3.7	5.1	5.0	5.0	4.6	4.7	4.6	4.5	4.4	4.4
		5	5.5	6.0	5.2	5.0	5.5	4.7	4.9	5.0	5.9	4.5	4.6	4.9	4.2	4.2	3.8
		10	6.0	5.8	5.6	5.4	5.2	2.7	5.2	5.0	4.9	4.7	4.6	1.6	4.4	4.2	4.1
		25	6.0	5.7	5.8	5.3	5.1	4.9	5.3	4.9	5.1	4.8	4.3	4.3	4.3	4.3	4.0

TABLE 3.6: Achieved size I at $\alpha = 0.05$ by the new method ($T_{M,\theta}$), the new method with t approximation ($T^s_{M,\theta}$), the method of Xu and Harrar (2012) (T_{XH}), the method of Xu and Harrar (2012) with t approximation (T^s_{XH}) and complete-case t test (T). Data in the two groups are generated from Log-Normal distribution with equal correlation. The simulation size is $10,000$.

Despite lack of rigorous theory to justify its favorable performance in small samples, the t-approximation with degrees of freedom in (3.18) continues to perform well in our simulation study as well. Therefore, we recommend its use for smaller sample sizes, in particular, when the dependence in the two groups are about the same and when it is not too strong.

	(n_1, n_2)	%	$T_{M,\theta}$			$T^s_{M,\theta}$			T_{XH}			T^s_{XH}			T		
ρ_1			-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5	0
ρ_2			0	0.85	0.85	0	0.85	0.85	0	0.85	0.85	0	0.85	0.85	0	0.85	0.85
Lognormal	(5,15)	0	11.9	13.6	10.8	7.3	7.8	5.5	7.0	8.4	7.2	6.3	7.4	6.4	3.9	3.5	3.1
		5	12.7	14.7	12.1	8.4	10.6	7.3	7.8	9.2	8.7	7.0	8.2	7.7	3.6	3.4	3.0
		10	13.1	15.2	13.0	8.8	11.1	8.0	8.3	9.9	9.3	7.4	8.9	8.4	3.6	3.5	3.1
		25	14.7	17.3	14.6	10.2	13.4	9.8	9.0	11.3	10.8	8.1	10.3	9.7	3.7	3.5	3.4
	(15,5)	0	9.2	5.7	5.1	4.7	4.0	3.1	6.3	5.3	5.2	1.8	0.0	0.0	3.6	3.6	3.5
		5	9.6	6.1	5.8	4.6	4.8	4.5	6.8	6.4	6.5	1.6	0.7	1.0	3.5	3.5	3.4
		10	10.1	6.6	6.5	5.1	5.1	5.0	7.2	7.0	7.4	1.8	1.0	1.5	3.4	3.6	3.5
		25	11.9	8.1	8.2	6.2	6.5	6.6	8.5	8.5	8.5	2.7	1.8	2.4	3.4	3.2	3.1
	(15,25)	0	6.5	7.5	6.6	5.1	5.7	4.9	5.0	5.4	5.0	4.4	4.7	4.4	4.0	4.0	3.6
		5	6.7	7.8	6.9	5.8	6.2	5.5	5.6	6.0	5.7	4.9	5.4	5.1	4.2	4.2	3.8
		10	7.0	8.0	6.8	6.0	6.5	5.6	5.5	6.1	5.9	4.8	4.8	3.8	4.2	4.1	4.0
		25	6.9	7.9	6.8	5.9	6.3	5.5	5.6	6.2	5.7	4.8	3.9	2.5	4.2	4.1	3.8
	(25,15)	0	6.0	5.5	5.2	5.3	4.6	4.2	5.1	5.1	4.8	3.4	2.5	2.1	4.1	4.0	3.8
		5	6.1	5.9	5.5	5.4	5.0	4.6	5.2	5.5	5.5	4.1	4.0	4.2	4.0	3.8	3.6
		10	6.2	6.0	5.8	5.5	5.1	4.8	5.4	5.8	5.6	4.0	4.2	4.3	3.9	3.7	3.5
		25	6.4	6.5	6.0	5.4	5.6	5.1	5.2	5.7	5.1	2.7	1.9	1.2	3.8	3.7	3.5
	(25,35)	0	5.8	6.3	5.6	5.1	5.3	4.7	5.1	4.9	4.7	4.7	4.5	4.3	4.4	4.0	3.9
		5	6.2	6.7	6.2	5.4	5.8	5.3	5.2	5.4	5.2	4.9	4.8	4.8	4.5	4.2	4.0
		10	6.1	6.7	6.2	5.4	5.4	5.3	5.4	5.4	5.2	5.0	4.8	4.8	4.5	4.3	3.9
		25	6.2	6.5	6.1	5.4	5.4	5.2	5.2	5.2	5.1	4.7	4.7	4.7	4.2	4.0	3.9
	(35,25)	0	5.4	5.4	5.1	5.0	4.8	4.6	4.7	4.7	4.8	3.8	3.6	3.8	4.0	4.0	4.2
		5	5.9	6.0	5.7	5.3	5.5	5.1	5.2	5.3	5.4	4.3	3.7	3.6	4.5	4.2	4.4
		10	5.9	5.8	5.8	5.4	5.2	5.1	5.2	5.5	5.2	4.2	3.9	3.8	4.4	4.4	4.2
		25	6.0	6.1	5.8	5.4	5.1	5.4	5.0	5.2	5.2	4.2	3.8	3.6	4.3	4.5	4.4

TABLE 3.7: Achieved size I at $\alpha = 0.05$ by the new method ($T_{M,\theta}$), the new method with t approximation ($T^s_{M,\theta}$), the method of Xu and Harrar (2012) (T_{XH}), the method of Xu and Harrar (2012) with t approximation (T^s_{XH}) and complete-case t test (T). Data in the two groups are generated from Log-Normal distribution with unequal correlation. The simulation size is $10,000$.

Power Simulation

To show an example of the power advantage of the new method, we set the sample size allocation to $n_1 = 15$ and $n_2 = 25$, and the missing percentage to $\mathrm{MP} = 20\%$.

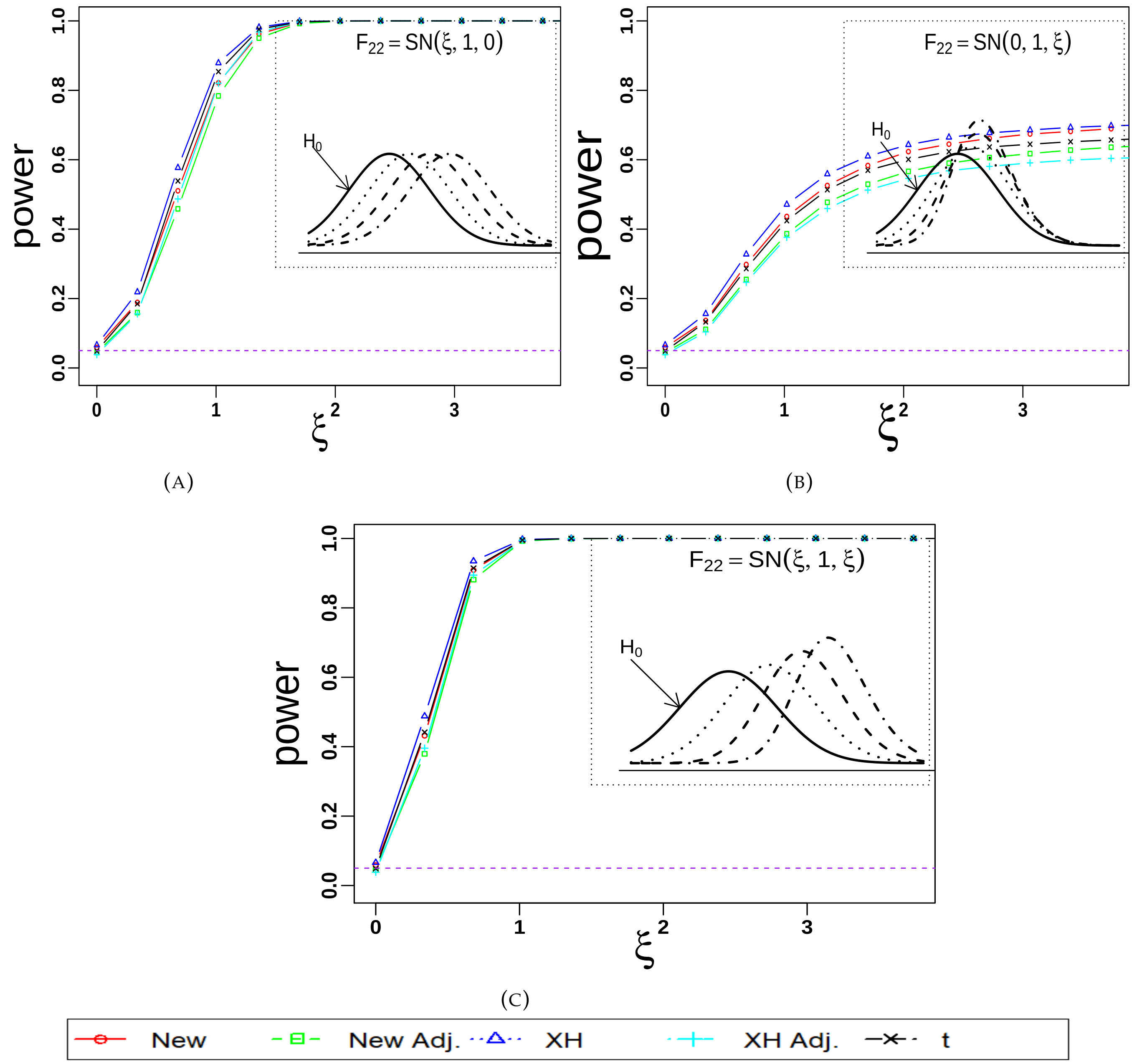


FIGURE 3.4: Power curves for the new method (New), the new method with t approximation (New Adj), the method of Xu and Harrar (2012) (XH), the method of Xu and Harrar (2012) with t approximation (XH Adj.) and complete-case t test (t). The inset plots show the pdf of the normal distribution from which the data are generated. The alternatives considered are $F_{11} = F_{12} = F_{21}$ the pdf marked as H_0 ($\xi = 0$) in the inset plot whereas F_{22} varies with respect to location and/or shape. The (A), (B) and (C) panels are for location, shape and location-shape, respectively, alternatives. For all plots the sample sizes n_1 and n_2 are 15 and 25, respectively, the missing percentage is 20% and the simulation of size is $10,000$.

Here also the size of the test is set at $\alpha = 0.05$. For the alternative hypothesis, we take three of the distributions to be the same but add location, shape or both to the fourth one. We choose the fourth distribution in such a way that when the location and/or shape parameters are zeros, it will be the same as the other three distributions.

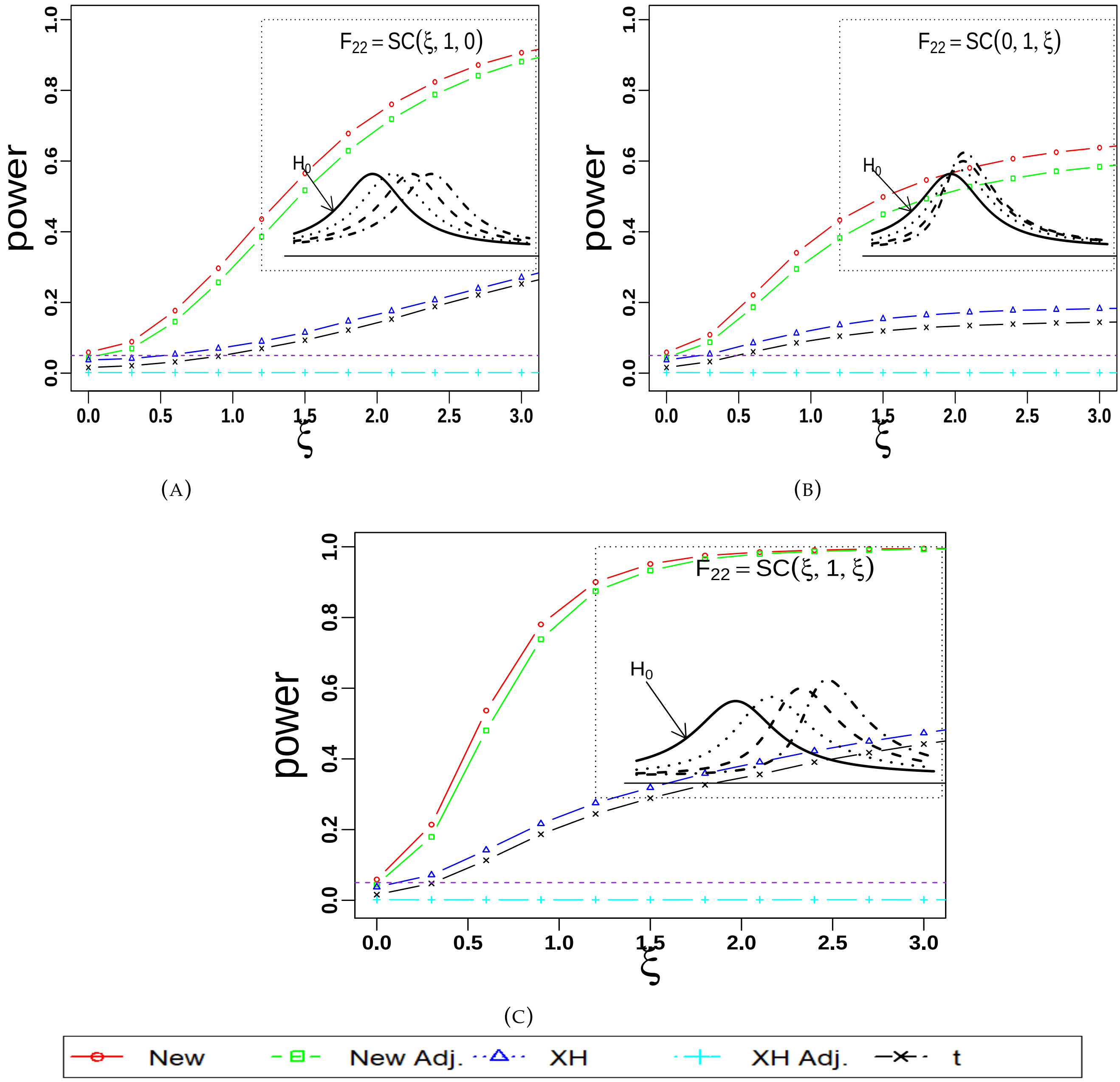


FIGURE 3.5: Power curves for the new method (New), the new method with t approximation (New Adj), the method of Xu and Harrar (2012) (XH), the method of Xu and Harrar (2012) with t approximation (XH Adj.) and complete-case t test (t). The inset plots show the pdf of the cauchy distribution from which the data are generated. The alternatives considered are $F_{11} = F_{12} = F_{21}$ the pdf marked as H_0 ($\xi = 0$) in the inset plot whereas F_{22} varies with respect to location and/or shape. The (A), (B) and (C) panels are for location, shape and location-shape, respectively, alternatives. For all plots the sample sizes n_1 and n_2 are are 15 and 25, respectively, the missing percentage is 20% and the simulation of size is $10,000$.

This scenario constitutes the null case and a desirable value of power would be 0.05 or less. The values of the scale parameter in all the four distributions is set to unity. In order to induce a measured amount of dependence within the paired observations, we generate data from the respective marginal distributions and glue them together with Clayton's copula (Nelsen, 2006) setting the parameter value at 1.5.

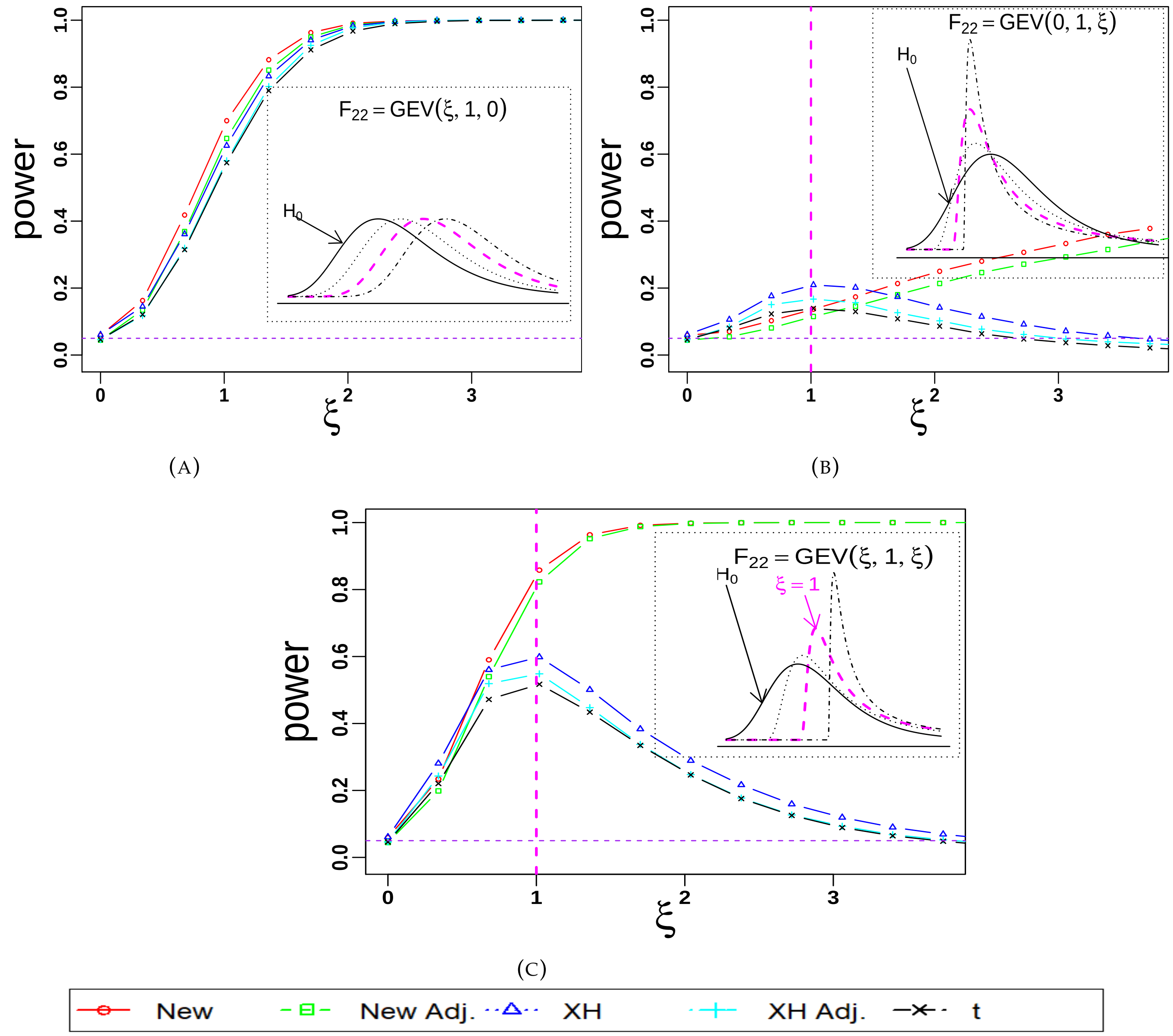


FIGURE 3.6: Power curves for the new method (New), the new method with t approximation (New Adj), the method of Xu and Harrar (2012) (XH), the method of Xu and Harrar (2012) with t approximation (XH Adj.) and complete-case t test (t). The inset plots show the pdf of the generalized extereme value distribution from which the data are generated. The alternatives considered are $F_{11} = F_{12} = F_{21}$ the pdf marked as H_0 ($\xi = 0$) in the inset plot whereas F_{22} varies with respect to location and/or shape. The (A), (B) and (C) panels are for location, shape and location-shape, respectively, alternatives. For all plots the sample sizes n_1 and n_2 are are 15 and 25, respectively, the missing percentage is 20% and the simulation of size is $10,000$.

This value of the parameter induces a dependence of Kendal's $\tau = 0.4286$, and Spearman's $\rho = 0.5987$. The marginal distributions we consider are skew-normal (SN) of Azzalini (1985), skew-t (ST) of Azzalini and Capitanio (2003) and generalized extreme value (GEV) of McFadden (1978). For the purpose of power simulation we consider three distributions namely, Normal, Cauchy as an example of heavy tailed distribution and generalized extreme value distribution as an example of skewed distribution. For these distributions, the power curves are shown in the following Figures 3.4, 3.5 and 3.6, respectively.

The results from the power simulations shown in Figures 3.4, 3.5 and 3.6 are rather unequivocal. For location alternatives (panel (A) in Figures 3.4, 3.5 and 3.6), the new nonparametric method has a clear advantage when data come from heavy tailed distributions. Otherwise, all the five methods perform comparably well. On the other hand, the performance of the new nonparametric method excels that of the other three parametric methods for shape-alone alternatives(panel (B) in Figures 3.4, 3.5 and 3.6), in particular, for heavy-tailed case. In Figure 3.6 we should point out that that GEV does not have any of its moments when the shape parameter exceeds one (indicated by the vertical lines in panels (B) and (C)) which means the tails of this distribution get heavier as the shape parameter increases. Another point that needs to be stressed is that for the shape-only alternatives (panel (B)), and location and shape alternative (panel (C))as the shape parameter increases the distributions approach a limit and, therefore, the power levels off at a value below 1. Finally, for the location and shape alternatives the parametric methods exhibit a surprisingly-poor performance that deteriorates with increasing shape and location in the cases of Cauchy and GEV distributions. The method of Xu and Harrar (2012) gains accuracy by correcting skewness and kurtosis effects on the asymptotic distribution. When these moments do not exist ($= \infty$) the corrections cannot be effective. This explains the unpleasant behavior especially under GEV because existence of moments depends on the value the shape parameter ξ. The undesirable behavior of the other two parametric methods can also be explained similarly.

3.6.2 Real Data Examples

In this subsection we illustrate the application of the new method using two real data examples. The first one involves a dataset from the Asthma Randomized Trial of Indoor Wood Smoke (ARTIS) Noonan and Ward (2012) which evaluated the changes in household air pollution within wood-burning homes, and the second dataset comes from Greek Health Project (GHP), a clinical trial to assess the efficacy of a particular form of counseling known as motivational interviewing (MI) for motivating college students to change their smoking behavior (Harris et al., 2010). Albeit both datasets are from public health, both feature different levels of missing values and outcome types.

ARTIS Dataset

Briefly, ARTIS was a three-arm randomized placebo-controlled intervention trial with two intervention strategies for reducing in-home woodsmoke particulate matter (PM). Eligible participants included children with asthma, age 6–18 years, residing in a non-tobacco-smoking household that used an older-model woodstove as their primary source of heating. Each household participated in two consecutive winter periods with household interventions occurring between the two winter periods. Homes were randomly assigned to one of three treatments: the woodstove-intervention group receiving improved-technology wood-burning appliances (i.e., EPA-certified woodstoves), the air-filter group receiving functioning air-filtration devices, and the placebo group receiving sham air-filtration devices. The aim was to test the hypothesis that household-level interventions, specifically improved-technology wood-burning appliances or air-filtration devices, would improve health outcomes measure. The primary health outcome of interest was Pediatric Asthma Quality of Life Questionnaire (PAQLQ). Prior to enrollment of the final cohort of homes, woodstove changeouts intervention arm was discontinued as an interim analysis indicated that the new stoves were not efficacious in reducing indoor PM2.5. Therefore, we disregard this arm in our analysis. More details on the ARTIS trial and the dataset can be found in Noonan and Ward (2012) and Noonan et al. (2017). In order to avoid the complications of the home clustering effect, only the primary child (the child with more severe asthma based on screening questions) is included if a household had more than one eligible child with asthma.

A total of 39 and 42 homes were recruited and visited in the pre-treatment winters in the sham filter and air filter groups, respectively. These numbers reduced, respectively, to 36 and 35 in the post-treatment winters. In the way the study was designed, there was no possibility for missing values to occur in the pre-treatment winters. The plots of the mean and relative summary effects profiles are displayed in Figure 3.7. The outcome PAQLQ is a measure with an arbitrary scale and, thus, a nonparametric analysis would be an appropriate choice.

The results in Figure 3.7 and Table 3.8 indicate the difference in the change associated with the two treatment, if any, is insignificant. The conclusions from all the five methods are the same. The complete case analysis gives a wider confidence interval than the other two semi-parametric

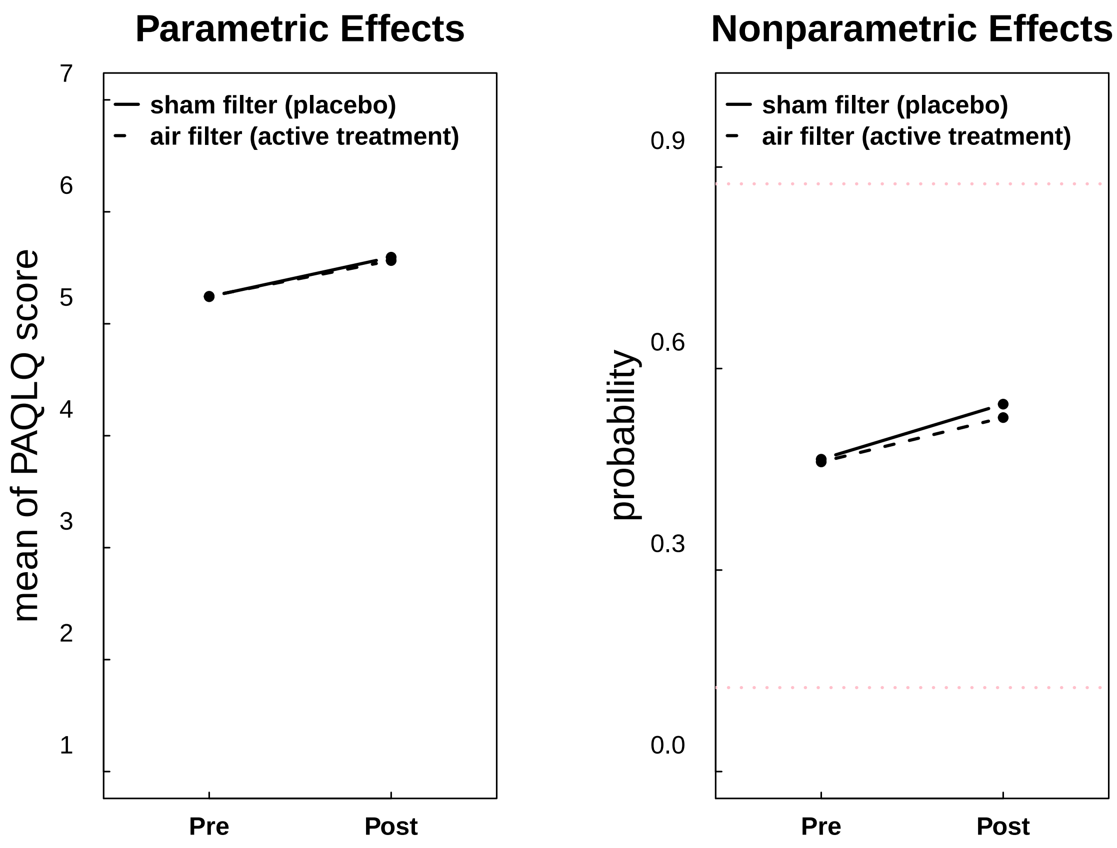


FIGURE 3.7: Profile plots of the means and relative summary effects of the Pediatric Asthma Quality of Life Questionnaire (PAQLQ) health outcome measure.

methods. The conclusion from the two nonparametric methods are also the same. The confidence intervals for their nonparametric effects are about the same.

GHP Dataset

The subjects for the study were students at University of Missouri–Columbia who were members of Greek letter organizations (fraternities and sororities). To be able to assess the efficacy of MI for smoking, the study employed MI based on fruit and vegetable consumption as a passive control. The study aimed to compare differences in smoking behavior outcomes between the treatment and control groups in which the treatment and control were matched on the number of counseling sessions, mode of delivery and counseling method. The treatments were administered in four one-on-one sessions lasting from 30 to 45 minutes with intersession period of about two weeks. Among the main outcomes of the study were number of days smoked in the past 30 days assessed at three time points; baseline, end of treatment (EOT), and six-month follow-up (FU). Unlike the ARTIS, the dataset in this example is subject to missing values at either of the two assessment occasions.

	$T_{M,\theta}$	$T^s_{M,\theta}$	T_{XH}	T^s_{XH}	T
Statistic	1.016	1.016	-0.034	-0.034	-0.069
p-value	0.856	0.856	0.922	0.922	0.399
95% CI	(0.847, 1.184)	(0.845, 1.187)	(-0.109, 0.041)	(-0.110, 0.043)	(-0.786, 0.648)

TABLE 3.8: The values of the test statistics, p-values and confidence intervals for the effect measure of the nonparametric methods proposed in the study ($T_{M,\theta}$: the new method and $T^s_{M,\theta}$: the new method with t approximation), as well as existing parametric and semi-parametric methods (T_{XH}: the method of Xu and Harrar (2012), T^s_{XH}: the method of Xu and Harrar (2012) with t approximation) and (T: complete-case t test).

Of the 452 participants recruited for the study, 306 were available at both EOT and FU whereas 49 and 25 participants did not show up at EOT and FU, respectively. The remaining 72 participants did not show up both at EOT and FU. The main objective of the study was to assess the differential effect of the treatment in causing change in smoking behavior at the EOT and whether that effect was maintained for a longer term. With that in mind, from here on we will only be concerned with analyzing the change from baseline in 30 days smoking rate at EOT and six-month FU. For more details for the GHP trial and the dataset we refer to Harris et al. (2010) and Xu and Harrar (2012).

The parametric method Xu and Harrar (2012) rejected the hypothesis of changes in the smoking reduction are the same in the two treatment groups: $T = -2.90$, p-value $= 0.001$ and 95% CI $(-3.91, -0.79)$. Likewise, the new nonparametric method rejected the null hypothesis of no tendency of the change in reduction in one group to be smaller or larger than that in the other group difference in the change: $T_{M,\theta} = -3.615$, p-value = 0.0003 and 95% CI $= (-0.173, -0.051)$. The results are coherent and complementary, and are also consistent with those of the other methods reported in Xu and Harrar (2012).

Chapter 4

Repeated Measures Design for Partially Complete Data: Three-Time Points

4.1 Introduction

The general repeated measure design model described in Chapter 2, equation (2.1), for one group situation, i.e., fixing $g = 1$, is the subject of this chapter. A nonparametric rank-based test procedure using relative effects for complete dataset is developed for this type of design by Konietschke et al. (2010). Here, we focus on developing a similar fully nonparametric procedure which can also handle missing data situation. This procedure is also robust to the Behrens-Fisher problem.

4.2 Statistical Model and Effects Size Measure

The current study extends the work of Konietschke et al. (2012a) to the case of *three* time points by developing a procedure for testing significance of effects over time and constructing a confidence interval for effects size.

Let

$$\boldsymbol{X}_k = (\Delta_{1k}X_{1k}, \Delta_{2k}X_{2k}, \Delta_{3k}X_{3k})',$$

$k = 1, \ldots, n,$ be a triplet of observations, where

$$X_{tk} \sim F_t,$$

for t = 1, 2, 3, and Δ_{tk} is an indicator defined as

$$\Delta_{tk} = \begin{cases} 1, & \text{if } X_{tk} \text{ is observed;} \\ 0, & \text{otherwise.} \end{cases}$$

As a consequence of the definition of Δ_{tk} there are

$$N_c = \sum_{k=1}^{n} \prod_{t=1}^{3} \Delta_{tk}$$

complete cases,

$$N_{jt} = \sum_{k=1}^{n} \Delta_{jk}\Delta_{tk}(1 - \Delta_{sk}), \quad \text{for} \quad s \neq \{j, t\},$$

subjects observed at j and t time points only, and

$$N_t = \sum_{k=1}^{n} \Delta_{tk}(1 - \Delta_{jk})(1 - \Delta_{sk}), \text{for} \quad j \neq t, s < j,$$

subjects observed at time point t only.

For $j < t = 1, 2, 3, s \in \{j, t\}$ let $n_c^{(jt)}$ denote the number of complete cases at time point s, when only time points j and t are considered. Similarly, $n_s^{(jt)}$ denotes the number of incomplete cases at time point s, when only time points j and t are considered. Then, for fixed j and t

$$n_c^{(jt)} = \sum_{k=1}^{n} \Delta_{jk}\Delta_{tk},$$

and

$$n_s^{(jt)} = \sum_{k=1}^{n} \Delta_{sk}(1 - \Delta_{rk}),$$

for $r \neq s \in \{j, t\}$. Incomplete cases refer to subjects where there are observations missing at least at one time point.

Furthermore,

$$n^{(jt)} = N_c + n_t^{(jt)} + n_j^{(jt)}$$

denotes the number of subjects observed at either of the time points j or t.

Notice that there are in total

$$n = N_c + \sum\sum_{1 \leq j < t \leq 3} N_{jt} + \sum_{t=1}^{3} N_t,$$

subjects and

$$N = 3N_c + 2 \sum\sum_{1 \le j < t \le 3} N_{jt} + \sum_{t=1}^{3} N_t,$$

observations for $j, t = 1, 2, 3$. The total number of subjects observed at time t with missing observation at least at one of the remaining time points is denoted as n_t and is given by

$$n_t = \sum_{k=1}^{n} \Delta_{tk}(1 - \Delta_{jk}\Delta_{sk}),$$

for $t \neq s$ and $j < s$.

The complete data are represented by

$$\boldsymbol{X}_k^{(c)} = (X_{1k}^{(c)}, X_{2k}^{(c)}, X_{3k}^{(c)})',$$

with marginal distributions

$$X_{tk}^{(c)} \sim F_t,$$

for k = 1, 2, ..., N_c subjects, and $t = 1, 2, 3$ time points. The incomplete cases are $X_{tk}^{(i)}$ are distributed as

$$X_{tk}^{(i)} \sim F_t,$$

for k = 1, 2, ..., n_t subjects, and $t = 1, 2, 3$ time points.

Note that we are assuming $X_{tk}^{(c)}$ and $X_{tk}^{(i)}$ have the same distribution. This assumption holds under MCAR.

The setup of the missing data problem we are addressing is illustrated in the Table 4.1 below.

TABLE 4.1: Schematic display of the dataset: "x" stands for available observation, "?" represents missing observation.

Subjects	Time 1	Time 2	Time 3
1	x	x	x
⋮	⋮	⋮	⋮
N_c	x	x	x
1	x	?	?
⋮	⋮	⋮	⋮
N_1	x	?	?
1	x	x	?
⋮	⋮	⋮	⋮
N_{12}	x	x	?
1	?	x	?
⋮	⋮	⋮	⋮
N_2	?	x	?
1	?	x	x
⋮	⋮	⋮	⋮
N_{23}	?	x	x
1	x	?	x
⋮	⋮	⋮	⋮
N_{13}	x	?	x
1	?	?	x
⋮	⋮	⋮	⋮
N_3	?	?	x

Our aim in this section is to develop a nonparametric test procedure for data with missing observations, specifically for the setup displayed in the above figure. For this purpose, we use marginal models to define relative effects. In nonparametric marginal models the underlying idea is to make use of the independence structure of the observations and marginal distributions in order to address questions of interest. Moreover, in order to account for ties and non-metric data, we will use the normalized version of the distribution function of a random variable in equation (1.2) in place of the usual (right continuous version) distribution function (Brunner et al., 2002). The same distribution function will be used for the purpose of defining time effects and constructing estimators.

Define the average distribution function, G, as

$$G = \frac{1}{3}\sum_{t=1}^{3} F_t.$$

Then the relative summary effect in equation (2.3) becomes

$$p_t = \int G dF_t = \frac{1}{3}\sum_{j=1}^{3} p_{jt}, \tag{4.1}$$

for t = 1, 2, 3. Accordingly, this summary effect, p_t, can be recognized as the average of all possible two-time points, say {j, t}, relative marginal effects (p_{jt}) with respect to F_t, where p_{jt} is defined as in equation (2.3).

This chapter has two objectives namely, (1) find an estimator for the time effects in terms of relative effects, and (2) develop test procedures for testing their significance. In the next section estimation procedures will be discussed.

4.3 Estimation of Relative Effects and their Properties

In this subsection, we propose a consistent and asymptotically unbiased estimator for the vector of nonparametric relative effect size measures, $\boldsymbol{p}$. Estimators are derived by replacing the marginal distribution function in equation (1.2) by its empirical counterpart. For this purpose we consider the following definitions and notations.

Let

$$\theta_s^{(jt)} := \frac{n_c^{(jt)}}{n_c^{(jt)} + n_s^{(jt)}},$$

for $j < t = 1, 2, 3, s \in \{j, t\}$be the relative sample sizes of the complete and incomplete dataset at each time point t; t = 1, 2, 3, in considering t^{th} and j^{th} time points. Define a weighted empirical distribution function $\hat{F}_{s,\theta}^{(jt)}(x)$ for the t^{th} and j^{th} time points as:

$$\hat{F}_{s,\theta}^{(jt)}(x) := \theta_s^{(jt)} \hat{F}_s^{(jt)(c)}(x) + (1 - \theta_s^{(jt)})\hat{F}_s^{(jt)(i)}(x), \tag{4.2}$$

where

$$\hat{F}_s^{(jt)(c)}(x) = \frac{1}{n_c^{(jt)}} \sum_{k=1}^{n_c^{(jt)}} c(x - X_{tk}^{(c)}) \qquad \hat{F}_s^{(jt)(i)}(x) = \frac{1}{n_s^{(jt)}} \sum_{k=1}^{n_s^{(jt)}} c(x - X_{tk}^{(i)}),$$

$c(\cdot)$ is as defined in equation (2.7). Consistent estimators, $\hat{F}_s^{(jt)(c)}$ and $\hat{F}_s^{(jt)(i)}$, of the distribution functions are analogous to those in Konietschke et al. (2012a).

Using the weighted distribution function in equation (4.2), the relative effect, p_{jt}, in equation (4.1) can be estimated by its empirical counterpart $\hat{p}_{jt}^{\theta}$

$$\hat{p}_{jt}^{\theta} = \int \hat{F}_{j,\theta} d\hat{F}_{t,\theta}. \tag{4.3}$$

Konietschke et al. (2012a) show that $\hat{p}_{jt}^{\theta}$ is a consistent and asymptotically unbiased estimator of p_{jt}. Notice that, if $p_{jt} = \frac{1}{2}$, the observations in neither one of the two samples tend to be smaller or larger than in the other sample. On the other hand, if $p_{jt} < \frac{1}{2}$, the observations with the distribution F_t tend to be smaller than the observations with the distribution F_j and vice versa for $p_{jt} > \frac{1}{2}$. For two time points, say *j* and *t*, $\hat{p}_{jt}^{\theta}$ can be computed in terms of ranks as

$$\hat{p}_{jt}^{\theta} = \frac{1}{N}[\theta_t^{(jt)} \bar{R}_{t.}^{(c)} - \theta_j^{(jt)} \bar{R}_{j.}^{(c)} + (1-\theta_t^{(jt)})\bar{R}_{t.}^{(i)} + (1-\theta_j^{(jt)})\bar{R}_{j.}^{(i)}] - \frac{1}{2}, \tag{4.4}$$

where $\bar{R}_{t.}^{(c)}$ and $\bar{R}_{t.}^{(i)}$ are averages of ranks of X_{tk} among n^{jt} observations in a sample for complete and incomplete cases respectively at *j* and t time points. An estimator of p_t can be obtained by averaging $\hat{p}_{jt}^{\theta}$ over *j* as

$$\hat{p}_t^{\theta} = \frac{1}{3} \sum_{j=1}^{3} \int \hat{F}_{j,\theta} d\hat{F}_{t,\theta}. \tag{4.5}$$

Collecting these estimators, $\hat{p}_t^{\theta}$, in a vector yields the estimator $\hat{\boldsymbol{p}}^{\theta} = (\hat{p}_1^{\theta}, \hat{p}_2^{\theta}, \hat{p}_3^{\theta})'$ of the vector of relative effects $\boldsymbol{p} = (p_1, p_2, p_3)'$.

In Proposition 5 below we show consistency and asymptotic unbiasedness of $\hat{\boldsymbol{p}}^{\theta}$. For that we need the following assumption.

Assumption 3. $n_c^{(jt)} + n_s^{(jt)} \to \infty$ *such that* $\frac{n^{(jt)}}{n_c^{(jt)} + n_s^{(jt)}} \leq N_o < \infty$, $g, t = 1, 2, 3$, $s \in \{j, t\}$.

Assumptions 2 and 3 are satisfied if the number of available observation at each time point is large. This in fact, can happen, for instance, if either the complete cases or both of the incomplete cases are large at each time point. Specifically, for fixed *j* and *t*, either $n_c^{(jt)} \to \infty$ or $\min(n_j^{(jt)}, n_t^{(jt)}) \to \infty$ for each $j, t = 1, 2, 3$. In order to derive asymptotic results, in addition to the above Assumption 3, we will further assume that the observations are missing completely at random (MCAR).

Proposition 5. *Under Assumption 2 and 3, given earlier,* *1)* $\hat{\boldsymbol{p}}^{\theta}$ *is an asymptotically unbiased estimator of* $\boldsymbol{p}$*.*

2) $\hat{\boldsymbol{p}}^{\theta}$ *is a strongly consistent estimator of* $\boldsymbol{p}$*, i.e.,* $||\hat{\boldsymbol{p}}^{\theta} - \boldsymbol{p}|| \xrightarrow{a.s} \mathbf{0}$.

Proof. For part 1, taking the expectation of $\hat{p}_t^{\theta}$ in equation (4.5), we have

$$\begin{aligned} E(\hat{p}_t^{\theta}) &= \frac{1}{3}\sum_{j=1}^{3} E(\hat{p}_{jt}^{\theta}) \\ &= \frac{1}{3}\sum_{j=1}^{3}\{p_{jt} + O(\frac{1}{n^{(jt)}})\}, \qquad \text{(from Proposition 3 in Chapter 3 and by the Assumption 3 above)} \\ &= p_t + O(\frac{1}{n}). \end{aligned}$$

Hence,

$$E(\hat{\boldsymbol{p}}^{\theta}) = \boldsymbol{p} + O(\frac{1}{n}).$$

For the second part, the quantity $\hat{p}_t^{\theta}$ is given as

$$\hat{p}_t^{\theta} = \frac{1}{3}\sum_{j=1}^{3}\int \hat{F}_{j,\theta} d\hat{F}_{t,\theta}.$$

Using equation (4.5) we have,

$$\begin{aligned} |\hat{p}_t^{\theta} - p_t| &= |\frac{1}{3}\sum_{j=1}^{3}\int \hat{F}_{j,\theta} d\hat{F}_{t,\theta} - \frac{1}{3}\sum_{j=1}^{3}\int F_j dF_t| \\ &\leq \frac{1}{3}\sum_{j=1}^{3}|\int \hat{F}_{j,\theta} d\hat{F}_{t,\theta} - \int F_j dF_t|, \qquad \text{(triangle inequality)} \\ &= \frac{1}{3}\sum_{j=1}^{3}|\int \hat{F}_{j,\theta} d\hat{F}_{t,\theta} - \int F_j dF_t - \int F_j d\hat{F}_{t,\theta} + \int F_j d\hat{F}_{t,\theta}| \\ &= \frac{1}{3}\sum_{j=1}^{3}|\int (\hat{F}_{j,\theta} - F_j) d\hat{F}_{t,\theta} + \int F_j d(\hat{F}_{t,\theta} - F_t)| \\ &\leq \frac{1}{3}\sum_{j=1}^{3}\left(|\int (\hat{F}_{j,\theta} - F_j) d\hat{F}_{t,\theta}| + |\int F_j d(\hat{F}_{t,\theta} - F_t)|\right), \qquad \text{(triangle inequality)} \\ &= \frac{1}{3}\sum_{j=1}^{3}\left(|\int (\hat{F}_{j,\theta} - F_j) d\hat{F}_{t,\theta}| + |\int (\hat{F}_{t,\theta} - F_t) dF_j|\right) \\ &\leq \frac{1}{3}\sum_{j=1}^{3}\left(||\hat{F}_{j,\theta} - F_j||_{\infty} + ||\hat{F}_{t,\theta} - F_t||_{\infty}\right). \end{aligned}$$

By the Glivenko-Cantelli theorem, $||\hat{F}_{t,\theta} - F_t||_{\infty} \xrightarrow{a.s} 0$ for $t = 1, 2, 3$.

So,

$$\frac{1}{3}\sum_{j=1}^{3}\Big(||\hat{F}_{j,\theta}-F_j||_\infty-||\hat{F}_{t,\theta}-F_t||_\infty\Big)\xrightarrow{a.s}0$$

implying that

$$|\hat{p}_t^\theta-p_t|\xrightarrow{a.s}0,$$

for t= 1, 2, 3.

Therefore,

$$||\hat{\mathbf{p}}_t^\theta-\mathbf{p}_t||\xrightarrow{a.s}0.$$

□

To derive the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{p}}^\theta-\boldsymbol{p})$, we first consider

$$\sqrt{n}\boldsymbol{M}_\theta=\sqrt{n}(M_{\theta,1},M_{\theta,2},M_{\theta,3})',$$

where

$$\sqrt{n}M_{\theta,t}=\sqrt{n}\frac{1}{3}\sum_{j=1}^{3}\bar{Z}_{jt.}\quad and\quad \bar{Z}_{jt.}=\int F_jd\hat{F}_{\theta,t}-\int F_td\hat{F}_{\theta,j}+1-2\int F_jdF_t, \tag{4.6}$$

and show the asymptotic equivalence of the two quantities, for t=1, 2, 3. That is

$$\sqrt{n}(\hat{\boldsymbol{p}}^\theta-\boldsymbol{p})-\sqrt{n}\boldsymbol{M}_\theta=O_p(\frac{1}{n}).$$

To that end we prove the following theorem.

Theorem 4. *Under Assumption 2 and 3,*

$$||\sqrt{n}(\hat{\boldsymbol{p}}^\theta-\boldsymbol{p})-\sqrt{n}\boldsymbol{M}_\theta||^2=O(\frac{1}{n}).$$

Proof. To prove the theorem it suffices to show that

$$E[\sqrt{n}(\hat{p}_t^\theta-p_t)-\sqrt{n}M_{\theta,t}]^2=O(\frac{1}{n}),$$

for $t = 1, 2, 3$. Now,

$$\begin{aligned}
&E[\sqrt{n}(\hat{p}_t^{\theta} - p_t) - \sqrt{n}M_{\theta,t}]^2 \\
&= E\left[\sqrt{n}\frac{1}{3}\sum_{j=1}^{3}(\hat{p}_{jt}^{\theta} - p_{jt}) - \sqrt{n}\frac{1}{3}\sum_{j=1}^{3}\bar{Z}_{jt.}\right]^2 \\
&= E\left[\frac{1}{3}\sum_{j=1}^{3}(\sqrt{n}(\hat{p}_{jt}^{\theta} - p_{jt}) - \sqrt{n}\bar{Z}_{jt.})\right]^2 \\
&\leq \frac{1}{9}\sum_{j=1}^{3}E(\sqrt{n}(\hat{p}_{jt}^{\theta} - p_{jt}) - \sqrt{n}\bar{Z}_{jt.})^2 \\
&= O(\frac{1}{n}),
\end{aligned}$$

since $E(\sqrt{n}(\hat{p}_{jt}^{\theta} - p_{jt}) - \sqrt{n^{(jt)}}\bar{Z}_{jt.})^2 = O(\frac{1}{n})$, which follows by a similar method in the proof of Theorem 4.1 in Konietschke et al. (2012a). □

In the next theorem, we derive the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{p}}^{\theta} - \boldsymbol{p})$. For this purpose Theorem 4 plays a crucial role as it expresses the statistic of interest, $\sqrt{n}(\hat{\boldsymbol{p}}^{\theta} - \boldsymbol{p})$, as the sum of independent random variables. Application of the standard central limit theorem to these independent random variables together with Assumptions 2 and 3 helps to derive the asymptotic distribution.

Theorem 5. *Under Assumptions 2 and 3, $\sqrt{n}(\hat{\boldsymbol{p}}^{\theta} - \boldsymbol{p})$ has, asymptotically, a multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = Cov(\sqrt{n}\boldsymbol{M}_{\theta})$.*

Proof. Here we first express $\boldsymbol{M}_{\theta}$ as a linear combination of random variables, which can be expressed as sum of independent random variables, so that we can use Cramer-Wold's device and Lypanov's conditon to arrive at the desired result. Let $\mathbf{a} = (a_1, a_2, a_3)'$ be a vector with arbitrary constants.

$\sqrt{n}M_{\theta,t} = \sqrt{n}\frac{1}{3}\sum_{j=1}^{3}\bar{Z}_{jt.}$, where

$$\begin{aligned}
\bar{Z}_{jt.} &= \int F_j d\hat{F}_{\theta,t} - \int F_t d\hat{F}_{\theta,j} + 1 - 2\int F_j dF_t \\
&= \frac{\theta_2^{(jt)}}{n_c^{(jt)}}\sum_{k=1}^{n_c^{(jt)}} F_j(X_{tk}^{(c)}) + \frac{(1-\theta_2^{(jt)})}{n_2^{(jt)}}\sum_{k=1}^{n_2^{(jt)}} F_j(X_{tk}^{(i)}) - \frac{\theta_1^{(jt)}}{n_c^{(jt)}}\sum_{k=1}^{n_c^{(jt)}} F_t(X_{jk}^{(c)}) - \frac{(1-\theta_1^{(jt)})}{n_1^{(jt)}}\sum_{k=1}^{n_1^{(jt)}} F_t(X_{jk}^{(i)}) \\
&\quad + 1 - 2\int F_j dF_t \\
&= \frac{1}{n_c^{(jt)}}\sum_{k=1}^{n_c^{(jt)}} S_{jtk} + \frac{(1-\theta_2^{(jt)})}{n_2^{(jt)}}\sum_{k=1}^{n_2^{(jt)}} Y_{jt.2k} - \frac{(1-\theta_1^{(jt)})}{n_1^{(jt)}}\sum_{k=1}^{n_1^{(jt)}} Y_{jt.1k} + 1 - 2\int F_j dF_t.
\end{aligned}$$

In the above,

$$S_{jtk} := \theta_2^{(jt)} F_j(X_{tk}^{(c)}) - \theta_1^{(jt)} F_t(X_{jk}^{(c)}) \quad Y_{jt.1k} := F_t(X_{jk}^{(i)}) \quad and \quad Y_{jt.2k} := F_j(X_{tk}^{(i)}).$$

Now,

$$\begin{aligned}
\sqrt{n}\mathbf{a}'\mathbf{M}_\theta &= \sqrt{n}(a_1 M_{\theta,1} + a_2 M_{\theta,2} + a_3 M_{\theta,3}) \\
&= \frac{\sqrt{n}}{3}\sum_{j=1}^{3}\left[a_1 \bar{Z}_{j1.} + a_2 \bar{Z}_{j2.} + a_3 \bar{Z}_{j3.}\right] \\
&= \frac{\sqrt{n}}{3}\sum_{j=1}^{3}[a_1 \left(\frac{1}{n_c^{(j1)}} \sum_{k=1}^{n_c^{(j1)}} S_{j1k}^{(c)} + \frac{(1-\theta_2^{(j1)})}{n_2^{(j1)}} \sum_{k=1}^{n_2^{(j1)}} Y_{j1.2k}^{(i)} - \frac{(1-\theta_1^{(j1)})}{n_1^{(j1)}} \sum_{k=1}^{n_1^{(j1)}} Y_{j1.1k}^{(i)} + 1 - 2\int F_j dF_1 \right) \\
&\quad + a_2 \left(\frac{1}{n_c^{(j2)}} \sum_{k=1}^{n_c^{(j2)}} S_{j2k}^{(c)} + \frac{(1-\theta_2^{(j2)})}{n_2^{(j2)}} \sum_{k=1}^{n_2^{(j2)}} Y_{j2.2k}^{(i)} - \frac{(1-\theta_1^{(j2)})}{n_1^{(j2)}} \sum_{k=1}^{n_1^{(j2)}} Y_{j2.1k}^{(i)} + 1 - 2\int F_j dF_2 \right) \\
&\quad + a_3 \left(\frac{1}{n_c^{(j3)}} \sum_{k=1}^{n_c^{(j3)}} S_{j3k}^{(c)} + \frac{(1-\theta_2^{(j3)})}{n_2^{(j3)}} \sum_{k=1}^{n_2^{(j3)}} Y_{j3.2k}^{(i)} - \frac{(1-\theta_1^{(j3)})}{n_1^{(j3)}} \sum_{k=1}^{n_1^{(j3)}} Y_{j3.1k}^{(i)} + 1 - 2\int F_j dF_3 \right)].
\end{aligned}$$

Since $Z_{jtk} = 0$ for $j = t$ and sample $\{j, t\}$ and $\{t, j\}$ are symmetric,

$$\sqrt{n}\mathbf{a}'\mathbf{M}_\theta = \frac{2\sqrt{n}}{3} A,$$

where

$$\begin{aligned}
A = [a_1 &\left(\frac{1}{n_c^{(12)}} \sum_{k=1}^{n_c^{(12)}} S_{12k}^{(c)} + \frac{(1-\theta_2^{(12)})}{n_2^{(12)}} \sum_{k=1}^{n_2^{(12)}} Y_{12.2k}^{(i)} - \frac{(1-\theta_1^{(12)})}{n_1^{(12)}} \sum_{k=1}^{n_1^{(12)}} Y_{12.1k}^{(i)} + 1 - 2\int F_1 dF_2 \right) \\
&+ a_2 \left(\frac{1}{n_c^{(23)}} \sum_{k=1}^{n_c^{(23)}} S_{23k}^{(c)} + \frac{(1-\theta_2^{(23)})}{n_2^{(23)}} \sum_{k=1}^{n_2^{(23)}} Y_{23.2k}^{(i)} - \frac{(1-\theta_1^{(23)})}{n_1^{(23)}} \sum_{k=1}^{n_1^{(23)}} Y_{23.1k}^{(i)} + 1 - 2\int F_2 dF_3 \right) \\
&+ a_3 \left(\frac{1}{n_c^{(13)}} \sum_{k=1}^{n_c^{(13)}} S_{13k}^{(c)} + \frac{(1-\theta_2^{(13)})}{n_2^{(13)}} \sum_{k=1}^{n_2^{(13)}} Y_{13.2k}^{(i)} - \frac{(1-\theta_1^{(13)})}{n_1^{(13)}} \sum_{k=1}^{n_1^{(13)}} Y_{13.1k}^{(i)} + 1 - 2\int F_1 dF_3 \right)].
\end{aligned}$$

The terms on the right-hand side of the above equation in the square brackets, ignoring the constant terms, can be expressed as

$$a_1\left(\frac{1}{n_c^{(12)}}\sum_{k=1}^{n_c^{(12)}}\{S_{12k}^{1(c)}+S_{12k}^{2(c)}\}+\frac{(1-\theta_2^{(12)})}{n_2^{(12)}}\sum_{k=1}^{n_2^{(12)}}\{Y_{12.2k}^{1(i)}+Y_{12.2k}^{2(i)}\}-\frac{(1-\theta_1^{(12)})}{n_1^{(12)}}\sum_{k=1}^{n_1^{(12)}}\{Y_{12.1k}^{1(i)}+Y_{12.1k}^{2(i)}\}\right)$$
$$+a_2\left(\frac{1}{n_c^{(23)}}\sum_{k=1}^{n_c^{(23)}}\{S_{23k}^{1(c)}+S_{23k}^{2(c)}\}+\frac{(1-\theta_2^{(23)})}{n_2^{(23)}}\sum_{k=1}^{n_2^{(23)}}\{Y_{23.2k}^{1(i)}+Y_{23.2k}^{2(i)}\}-\frac{(1-\theta_1^{(23)})}{n_1^{(23)}}\sum_{k=1}^{n_1^{(23)}}\{Y_{23.1k}^{1(i)}+Y_{23.1k}^{2(i)}\}\right)$$
$$+a_3\left(\frac{1}{n_c^{(13)}}\sum_{k=1}^{n_c^{(13)}}\{S_{13k}^{1(c)}+S_{13k}^{2(c)}\}+\frac{(1-\theta_2^{(13)})}{n_2^{(13)}}\sum_{k=1}^{n_2^{(13)}}\{Y_{13.2k}^{1(i)}+Y_{13.2k}^{2(i)}\}-\frac{(1-\theta_1^{(13)})}{n_1^{(13)}}\sum_{k=1}^{n_1^{(13)}}\{Y_{13.1k}^{1(i)}+Y_{13.1k}^{2(i)}\}\right).$$

In the above, for instance,

$$S_{12k}^{(c)}=S_{12k}^{1(c)}+S_{12k}^{2(c)},$$

where $S_{12k}^{1(c)}$ represents observations from 1, ..., N_c and $S_{12k}^{2(c)}$ represents observations from N_c +1, ..., N_c+N_{12}. Here the variable splits into two to rearrange them into independent groups.

Collecting the missing pattern, we can form the sum of independent random variables as,

$$\sqrt{n}\mathbf{a}'\mathbf{M}_\theta=\left(\frac{1}{N_c}\sum_{k=1}^{N_c}\underbrace{\frac{2}{3}\{a_1S_{12k}^{1(c)}+a_2S_{23k}^{1(c)}+a_3S_{13k}^{1(c)}\}}_{H_{ck}}\right)$$
$$+\left(\frac{1}{N_{12}}\sum_{k=1}^{N_{12}}\underbrace{\frac{2}{3}\{a_1S_{12k}^{2(c)}+a_2(1-\theta_2^{(23)})Y_{23.1k}^{1(i)}-a_3(1-\theta_1^{(13)})Y_{13.1k}^{1(i)}\}}_{H_{12k}}\right)$$
$$+\left(\frac{1}{N_2}\sum_{k=1}^{N_2}\underbrace{\frac{2}{3}\{a_1(1-\theta_2^{(12)})Y_{12.2k}^{1(i)}-a_2(1-\theta_1^{(23)})Y_{23.1k}^{2(i)}\}}_{H_{2k}}\right)$$
$$+\left(\frac{1}{N_{23}}\sum_{k=1}^{N_{23}}\underbrace{\frac{2}{3}\{a_1(1-\theta_2^{(12)})Y_{12.2k}^{2(i)}+a_2S_{23k}^{2(c)}+a_3(1-\theta_2^{(13)})Y_{13.2k}^{1(i)}\}}_{H_{23k}}\right)$$
$$-\left(\frac{1}{N_{13}}\sum_{k=1}^{N_{13}}\underbrace{\frac{2}{3}\{a_1(1-\theta_1^{(12)})Y_{12.1k}^{1(i)}-a_2(1-\theta_2^{(23)})Y_{23.2k}^{1(i)}-a_3S_{13k}^{2(c)}\}}_{H_{13k}}\right)$$
$$-\left(\frac{1}{N_1}\sum_{k=1}^{N(1)}\underbrace{\frac{2}{3}\{a_1(1-\theta_1^{(12)})Y_{12.1k}^{2(i)}+a_3(1-\theta_1^{(13)})Y_{13.1k}^{2(i)}\}}_{H_{1k}}\right)$$

$$+\left(\frac{1}{N_3}\sum_{k=1}^{N_3}\underbrace{\frac{2}{3}\{a_2(1-\theta_2^{(23)})Y_{23.2k}^{2(i)}+a_3(1-\theta_2^{(13)})Y_{13.2k}^{2(i)}\}}_{H_{3k}}\right)$$

$$=\sqrt{n}\left(\sum_{k=1}^{N_c}\frac{H_{ck}}{N_c}+\sum_{k=1}^{N_{12}}\frac{H_{12k}}{N_{12}}+\sum_{k=1}^{N_{13}}\frac{H_{13k}}{N_{13}}+\sum_{k=1}^{N_{23}}\frac{H_{23k}}{N_{23}}+\sum_{k=1}^{N_1}\frac{H_{1k}}{N_1}+\sum_{k=1}^{N_2}\frac{H_{2k}}{N_2}+\sum_{k=1}^{N_3}\frac{H_{3k}}{N_3}\right) \quad (4.7)$$

Hence, the desired result can be obtained by verifying Lypanov's conditon in the same way as used in Theorem 2 on each term and applying the Cramer-Wold's theorem. □

The covariance matrix $\mathbf{\Sigma}$, can be expressed in terms of the covariance of the seven independent groups in equation (4.7).

$$\mathbf{\Sigma} = Cov(\sqrt{n}\mathbf{M}_\theta), \quad (4.8)$$

which can be written as

$$\mathbf{\Sigma} = nCov\begin{bmatrix} M_{\theta,1} \\ M_{\theta,2} \\ M_{\theta,3} \end{bmatrix} = n\begin{bmatrix} \sigma_{\theta,1}^2 & \sigma_{\theta,12} & \sigma_{\theta,13} \\ \sigma_{\theta,21} & \sigma_{\theta,2}^2 & \sigma_{\theta,23} \\ \sigma_{\theta,31} & \sigma_{\theta,32} & \sigma_{\theta,3}^2 \end{bmatrix}.$$

Notice that the vector $\boldsymbol{M}_\theta$ can be expressed as:

$$\boldsymbol{M}_\theta = \begin{bmatrix} \frac{1}{n_c^{(12)}}\sum_{k=1}^{n_c^{(12)}}\frac{2}{3}\{S_{12k}^{1(c)}+S_{12k}^{2(c)}\}+\frac{(1-\theta_2^{(12)})}{n_2^{(12)}}\sum_{k=1}^{n_2^{(12)}}\frac{2}{3}\{Y_{12.2k}^{1(i)}+Y_{12.2k}^{2(i)}\}-\frac{(1-\theta_1^{(12)})}{n_1^{(12)}}\sum_{k=1}^{n_1^{(12)}}\frac{2}{3}\{Y_{12.1k}^{1(i)}+Y_{12.1k}^{2(i)}\} \\ \frac{1}{n_c^{(23)}}\sum_{k=1}^{n_c^{(23)}}\frac{2}{3}\{S_{23k}^{1(c)}+S_{23k}^{2(c)}\}+\frac{(1-\theta_2^{(23)})}{n_2^{(23)}}\sum_{k=1}^{n_2^{(23)}}\frac{2}{3}\{Y_{23.2k}^{1(i)}+Y_{23.2k}^{2(i)}\}-\frac{(1-\theta_1^{(23)})}{n_1^{(23)}}\sum_{k=1}^{n_1^{(23)}}\frac{2}{3}\{Y_{23.1k}^{1(i)}+Y_{23.1k}^{2(i)}\} \\ \frac{1}{n_c^{(13)}}\sum_{k=1}^{n_c^{(13)}}\frac{2}{3}\{S_{13k}^{1(c)}+S_{13k}^{2(c)}\}+\frac{(1-\theta_2^{(13)})}{n_2^{(13)}}\sum_{k=1}^{n_2^{(13)}}\frac{2}{3}\{Y_{13.2k}^{1(i)}+Y_{13.2k}^{2(i)}\}-\frac{(1-\theta_1^{(13)})}{n_1^{(13)}}\sum_{k=1}^{n_1^{(13)}}\frac{2}{3}\{Y_{13.1k}^{1(i)}+Y_{13.1k}^{2(i)}\} \end{bmatrix}.$$

Rearranging the terms in the square brackets we can split the vector, $\boldsymbol{M}_\theta$, into seven independent pieces as given below

$$\boldsymbol{M}_\theta = \frac{1}{N_c}\frac{2}{3}\sum_{k=1}^{N_c}\underbrace{\begin{bmatrix} S_{12k}^{1(c)} \\ S_{23k}^{1(c)} \\ S_{13k}^{1(c)} \end{bmatrix}}_{A_{1k}} + \frac{1}{N_{12}}\frac{2}{3}\sum_{k=1}^{N_{12}}\underbrace{\begin{bmatrix} S_{12k}^{2(c)} \\ -(1-\theta_1^{(23)})Y_{23.1k}^{1(i)} \\ -(1-\theta_1^{(13)})Y_{13.1k}^{1(i)} \end{bmatrix}}_{A_{2k}} + \frac{1}{N_{23}}\frac{2}{3}\sum_{k=1}^{N_{23}}\underbrace{\begin{bmatrix} (1-\theta_2^{(12)})Y_{12.2k}^{2(i)} \\ S_{23k}^{2(c)} \\ (1-\theta_2^{(13)})Y_{13.2k}^{1(i)} \end{bmatrix}}_{A_{3k}}$$

$$+\frac{1}{N_{13}}\frac{2}{3}\sum_{k=1}^{N_{13}}\underbrace{\begin{bmatrix}-(1-\theta_1^{(12)})Y_{12.1k}^{1(i)}\\(1-\theta_2^{(23)})Y_{23.2k}^{1(i)}\\S_{13k}^{2(c)}\end{bmatrix}}_{A_{4k}}+\frac{1}{N_1}\frac{2}{3}\sum_{k=1}^{N_1}\underbrace{\begin{bmatrix}-(1-\theta_1^{(12)})Y_{12.1k}^{2(i)}\\0\\-(1-\theta_1^{(13)})Y_{13.1k}^{2(i)}\end{bmatrix}}_{A_{5k}}+\frac{1}{N_2}\frac{2}{3}\sum_{k=1}^{N_2}\underbrace{\begin{bmatrix}(1-\theta_2^{(12)})Y_{12.2k}^{1(i)}\\-(1-\theta_1^{(23)})Y_{23.1k}^{2(i)}\\0\end{bmatrix}}_{A_{6k}}$$

$$+\frac{1}{N_3}\frac{2}{3}\sum_{k=1}^{N_3}\underbrace{\begin{bmatrix}0\\(1-\theta_2^{(23)})Y_{23.2k}^{2(i)}\\(1-\theta_2^{(13)})Y_{13.2k}^{2(i)}\end{bmatrix}}_{A_{7k}}$$

$$=\frac{2}{3}\{\frac{1}{N_c}\sum_{k=1}^{N_c}A_{1k}+\frac{1}{N_{12}}\sum_{k=1}^{N_{12}}A_{2k}+\frac{1}{N_{23}}\sum_{k=1}^{N_{23}}A_{3k}\quad+\frac{1}{N_{13}}\sum_{k=1}^{N_{13}}A_{4k}+\frac{1}{N_1}\sum_{k=1}^{N_1}A_{5k}+\frac{1}{N_2}\sum_{k=1}^{N_2}A_{6k}+\frac{1}{N_3}\sum_{k=1}^{N_3}A_{7k}\}.$$

Since A_{mk}s are independent, the covariance matrix $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{\Sigma}=Cov(\sqrt{n}\boldsymbol{M}_\theta)=n\frac{2}{3}\{\frac{1}{N_c^2}\sum_{k=1}^{N_c}Cov(A_{1k})+\frac{1}{N_{12}^2}\sum_{k=1}^{N_{12}}Cov(A_{2k})+\frac{1}{N_{23}^2}\sum_{k=1}^{N_{23}}Cov(A_{3k})+\frac{1}{N_{13}^2}\sum_{k=1}^{N_{13}}Cov(A_{4k})+\frac{1}{N_1^2}\sum_{k=1}^{N_1}Cov(A_{5k})+\frac{1}{N_2^2}\sum_{k=1}^{N_2}Cov(A_{6k})+\frac{1}{N_3^2}\sum_{k=1}^{N_3}Cov(A_{7k})\}.$$

Consider the elements on the diagonal of the covariance matrix in equation (4.8), i.e., $\sigma^2_{\theta,t}$, corresponding to time $j < t = 1, 2, 3$.

$$\sigma^2_{\theta,t}=\frac{4}{9}nVar\left(\sum_{k=1}^{N_c}\frac{S_{jtk}^{1(c)}}{N_c}+\sum_{k=1}^{N_{12}}\frac{S_{jtk}^{2(c)}}{N_{12}}+(1-\theta_2^{(jt)})\{\sum_{k=1}^{N_{23}}\frac{Y_{jt.2k}^{1(i)}}{N_{23}}+\sum_{k=1}^{N_{13}}\frac{Y_{jt.2k}^{2(i)}}{N_{13}}\}\right)-\frac{4}{9}nVar\left((1-\theta_1^{(jt)})\{\sum_{k=1}^{N_1}\frac{Y_{jt.1k}^{1(i)}}{N_1}+\sum_{k=1}^{N_2}\frac{Y_{jt.1k}^{2(i)}}{N_2}\}\right)$$

$$=\frac{4}{9}n\left(\frac{\sigma^2_{jt.c1}}{N_c}+\frac{\sigma^2_{jt.c2}}{N_{12}}+(1-\theta_2^{(jt)})^2(\frac{\sigma^2_{jt.21}}{N_{23}}+\frac{\sigma^2_{jt.22}}{N_{13}})-(1-\theta_1^{(jt)})^2(\frac{\sigma^2_{jt.11}}{N_1}+\frac{\sigma^2_{jt.12}}{N_2})\right), \quad (4.9)$$

for $j < t = 1, 2, 3$. Next consider $\sigma_{\theta,12}$,

$$\sigma_{\theta,12} = \frac{4}{9}n\{\sum_{k=1}^{N_c} \frac{Cov(S_{12k}^{1(c)}, S_{23k}^{1(c)})}{N_c^2} + (1-\theta_2^{(23)}) \sum_{k=1}^{N_{12}} \frac{Cov(S_{12k}^{2(c)}, Y_{23.1k}^{1(i)})}{N_{12}^2} + (1-\theta_2^{(12)}) \sum_{k=1}^{N_{23}} \frac{Cov(S_{23k}^{1(c)}, Y_{12.2k}^{2(i)})}{N_{23}^2}$$
$$- (1-\theta_1^{(12)})(1-\theta_2^{(23)}) \sum_{k=1}^{N_{13}} \frac{Cov(Y_{12.1k}^{1(i)}, Y_{23.2k}^{1(i)})}{N_{13}^2} - (1-\theta_2^{(12)})(1-\theta_1^{(23)}) \sum_{k=1}^{N_2} \frac{Cov(Y_{12.2k}^{1(i)}, Y_{23.1k}^{2(i)})}{N_2^2}\},$$

which can be rewritten as

$$\sigma_{\theta,12} = \frac{4}{9}n \left(\frac{\sigma_{12.1}}{N_c} + (1-\theta_2^{(23)})\frac{\sigma_{12.2}}{N_{12}} + (1-\theta_2^{(12)})\frac{\sigma_{12.3}}{N_{23}} - (1-\theta_1^{(12)})(1-\theta_2^{(23)})\frac{\sigma_{12.4}}{N_{13}}) \right)$$
$$-\frac{4}{9}n \left((1-\theta_2^{(12)})(1-\theta_1^{(23)})\frac{\sigma_{12.5}}{N_2} \right). \quad (4.10)$$

In the above equation, for example, $\sigma_{12.1}$ in the first component is given by $\sigma_{12.1} = E(S_{121}^{1(c)} - E(S_{121}^{1(c)}))(S_{231}^{1(c)} - E(S_{231}^{1(c)}))$.

4.4 Estimation of Covariance Matrix

In this section, we derive a consistent estimator of the asymptotic covariance matrix $\mathbf{\Sigma}$. Note that it is sufficient to derive consistent estimators of elements of $\mathbf{\Sigma}$. If the random variables S_{jtk} and $Y_{jt.sk}$ were observable then, a natural estimator for the variance $\sigma_{\theta,1}^2$, will be

$$\tilde{\sigma}_{\theta,1}^2 = \frac{4}{9}n \left(\frac{\tilde{\sigma}_{12.c1}^2}{N_c} + \frac{\tilde{\sigma}_{12.c2}^2}{N_{12}} + (1-\theta_2^{(12)})^2(\frac{\tilde{\sigma}_{12.21}^2}{N_{23}} + \frac{\tilde{\sigma}_{12.22}^2}{N_{13}}) + (1-\theta_1^{(12)})^2(\frac{\tilde{\sigma}_{12.11}^2}{N_1} + \frac{\tilde{\sigma}_{12.12}^2}{N_2}) \right), \quad (4.11)$$

For example, the first term in the parentheses is given by

$$\tilde{\sigma}_{12.c1}^2 = \frac{1}{N_c - 1} \sum_{k=1}^{N_c} (S_{12k}^{1(c)} - \bar{S}_{12.}^{1(c)})^2.$$

The random variables S_{jtk} and $Y_{jt.sk}$, however, are not observable. For the computation of an estimator they must be replaced by observable values of random variables, which could be obtained by replacing distribution functions with their empirical counterparts. That is,

$$\hat{S}_{jtk} := \theta_2^{(jt)}\hat{F}_j(X_{tk}^{(c)}) - \theta_1^{(jt)}\hat{F}_t(X_{jk}^{(c)}), \quad \hat{Y}_{jt.jk}^{(i)} := \hat{F}_t(X_{jk}^{(i)}) \quad and \quad \hat{Y}_{jt.tk}^{(i)} := \hat{F}_j(X_{tk}^{(i)}).$$

The variance in equation (4.11) can be estimated by

$$\hat{\sigma}_{\theta,1}^2 = \frac{4}{9}n\left(\frac{\hat{\sigma}_{12.c1}^2}{N_c} + \frac{\hat{\sigma}_{12.c2}^2}{N_{12}} + (1-\theta_2^{(12)})^2(\frac{\hat{\sigma}_{12.21}^2}{N_{23}} + \frac{\hat{\sigma}_{12.22}^2}{N_{13}}) + (1-\theta_1^{(12)})^2(\frac{\hat{\sigma}_{12.11}^2}{N_1} + \frac{\hat{\sigma}_{12.12}^2}{N_2})\right). \tag{4.12}$$

For example,

$$\hat{\sigma}_{12.c1}^2 = \frac{1}{N_c-1}\sum_{k=1}^{N_c}(\hat{S}_{12k}^{1(c)} - \bar{\hat{S}}_{12.}^{1(c)})^2.$$

In a similar way, the first term of covariance, $\sigma_{\theta,12}$, in equation (4.10), $\sigma_{\theta,12.1}$ can be estimated by $\tilde{\sigma}_{\theta,12.1}$, where

$$\tilde{\sigma}_{\theta,12.1} = \frac{1}{N_c-1}\sum_{k=1}^{N_c}(S_{12k} - \bar{S}_{12.})(S_{23k} - \bar{S}_{23.}). \tag{4.13}$$

Since, S_{12k} and S_{23k} are unobservable, replacing them by their empirical counterparts results in the following estimator

$$\hat{\sigma}_{\theta,12.1} = \frac{1}{N_c-1}\sum_{k=1}^{N_c}(\hat{S}_{12k} - \bar{\hat{S}}_{12.})(\hat{S}_{23k} - \bar{\hat{S}}_{23.}). \tag{4.14}$$

Estimation for the remaining terms in equations (4.10) and (4.11) can be defined in the same way.

Theorem 6. *Under Assumptions 2 and 3, $\hat{\mathbf{\Sigma}}$ is a consistent estimator of $\mathbf{\Sigma}$, i.e., $\hat{\mathbf{\Sigma}} \xrightarrow{p} \mathbf{\Sigma}$.*

Proof. Consider the first element in the covariance matrix in equation (4.16). Note that from Theorem 3, $\hat{\sigma}_{\theta,1}^2$ in equation (4.12) is a consistent estimator of $\tilde{\sigma}_{\theta,1}^2$.

Now, we show that $\hat{\sigma}_{\theta,12.1}$ in equation (4.13) is a consistent estimator of $\tilde{\sigma}_{\theta,12.1}$ in equation (4.14).

$$\begin{aligned}
&E[\tilde{\sigma}_{\theta,12.1} - \hat{\sigma}_{\theta,12.1}]^2 \\
&= E\left[\frac{1}{N_c-1}\sum_{k=1}^{N_c}(S_{12k} - \bar{S}_{12.})(S_{23k} - \bar{S}_{23.}) - \frac{1}{N_c-1}\sum_{k=1}^{N_c}(\hat{S}_{12k} - \bar{\hat{S}}_{12.})(\hat{S}_{23k} - \bar{\hat{S}}_{23.})\right]^2 \\
&= \left(\frac{1}{N_c-1}\right)^2 E\left[\sum_{k=1}^{N_c}\{(S_{12k} - \bar{S}_{12.})(S_{23k} - \bar{S}_{23.}) - (\hat{S}_{12k} - \bar{\hat{S}}_{12.})(\hat{S}_{23k} - \bar{\hat{S}}_{23.})\}\right]^2 \\
&\leq \left(\frac{1}{N_c-1}\right)^2 \sum_{k=1}^{N_c} E\left[(S_{12k} - \bar{S}_{12.})(S_{23k} - \bar{S}_{23.}) - (\hat{S}_{12k} - \bar{\hat{S}}_{12.})(\hat{S}_{23k} - \bar{\hat{S}}_{23.})\right]^2 \\
&\leq 2N_c\left(\frac{1}{N_c-1}\right)^2 (E[(S_{121} - \bar{S}_{12.})(S_{231} - \bar{S}_{23.})]^2 + E[(\hat{S}_{121} - \bar{\hat{S}}_{12.})(\hat{S}_{231} - \bar{\hat{S}}_{23.})]^2) \\
&\leq 4N_c\left(\frac{1}{N_c-1}\right)^2 [E(S_{121} - \bar{S}_{12.})^2 + E(S_{231} - \bar{S}_{23.})^2 + E(\hat{S}_{121} - \bar{\hat{S}}_{12.})^2 + E(\hat{S}_{231} - \bar{\hat{S}}_{23.})^2] \\
&= O(\frac{1}{N_c^2}),
\end{aligned}$$

since all terms in the square brackets are of order $O(N_c^{-1})$. We can apply the same procedure to the remaining terms in equation (4.10). □

To facilitate the computation of the covariance matrix we employ a rank-transform based procedure. The empirical distributions are replaced by ranks as follows:

$$\hat{Y}_{jt.k}^{(w)} = \hat{F}_j(X_{tk}^{(w)}) = \frac{1}{N_{jt} - n_c^{(jt)} - n_t^{(jt)}} (R_{tk}^{(w)} - R_{tk}^{(w,t)}), \tag{4.15}$$

for $j \neq t$ and $w \in \{c, i\}$.

Using the relationship between empirical distribution and ranks in equation (4.15), therefore, a consistent estimator of the covariance matrix $\boldsymbol{\Sigma}$ is given by $\hat{\boldsymbol{\Sigma}}$,

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\sigma}_{\theta,1}^2 & \hat{\sigma}_{\theta,12} & \hat{\sigma}_{\theta,13} \\ \hat{\sigma}_{\theta,21} & \hat{\sigma}_{\theta,2}^2 & \hat{\sigma}_{\theta,23} \\ \hat{\sigma}_{\theta,31} & \hat{\sigma}_{\theta,32} & \hat{\sigma}_{\theta,3}^2 \end{bmatrix}, \tag{4.16}$$

where the elements are obtained in the way discussed in Section 4.4.

4.5 Test Procedures and Confidence Intervals

The null hypothesis of interest is that there is no statistically significant time effect versus the alternative that there is significant effect. Formally, we express this by

$$H_o : p_1 = p_2 = p_3 \quad vs. \quad H_1 : \text{at least one } p_t \text{ is different},$$

for $t = 1, 2, 3$. The above hypothesis can be rewritten as

$$H_o : \boldsymbol{Cp} = \boldsymbol{0} \quad vs. \quad H_o : \boldsymbol{Cp} \neq \boldsymbol{0}, \tag{4.17}$$

where **C** is the contrast matrix

$$C = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

and $\boldsymbol{p}$ is a vector of average relative effects, $\boldsymbol{p} = (p_1, p_2, p_3)'$.

Recall that in equation (4.17) for testing the hypothesis of no time effect, i.e., (H_o^p : $\boldsymbol{Cp} = \mathbf{0}$), the following two test statistics can be considered. Under the null hypothesis and Assumption 3, the Wald-type statistic (WTS)

$$WTS = n\hat{\boldsymbol{p}}^{\theta'}\boldsymbol{C'}[\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C'}]^{+}\boldsymbol{C}\hat{\boldsymbol{p}}^{\theta},$$

has asymptotically a χ^2_ν distribution with ν = rank(**C**) degrees of freedom, and $[A]^+$ denotes the Moore-Penrose generalized inverse of a matrix A (Brunner et al. (1999) and Brunner et al. (2002)). Also, under the null hypothesis, the ANOVA-type statistic (ATS)

$$ATS = n\frac{tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}})}{tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}}\boldsymbol{M}\hat{\boldsymbol{\Sigma}})}\hat{\boldsymbol{p}}^{\theta'}\boldsymbol{M}\hat{\boldsymbol{p}}^{\theta},$$

has approximately a χ^2_ν distribution with approximated degrees of freedom

$$\hat{\nu} = \frac{[tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}})]^2}{tr(\boldsymbol{M}\hat{\boldsymbol{\Sigma}}\boldsymbol{M}\hat{\boldsymbol{\Sigma}})}$$

and $\boldsymbol{M} = \boldsymbol{C'}(\boldsymbol{CC'})^{-}\boldsymbol{C}$, where $(\boldsymbol{CC'})^{-}$ is a generalized inverse of $\boldsymbol{CC'}$ (Brunner et al. (1999) and Brunner et al. (2002)).

It is known that the ATS and WTS are only useful in determining the validity of an overall hypothesis of no time effect; they do not provide more detailed information. Therefore, we derive a multiple contrast test procedure (MCTP) that allows explicit inference for arbitrary pairwise comparisons between samples.

4.5.1 Multiple Contrast Test Procedure

Since recent times there have been considerable work (Gao and Alvo (2008), Munzel and Hothorn (2001), Munzel and Tamhane (2002), Konietschke et al. (2010), and Konietschke et al. (2012b), Konietschke et al. (2018)) on the use of multiple comparison test procedures in a variety of designs. For instance, a nonparametric multiple comparison procedure based on the unweighted version of relative effects was developed by Gao and Alvo (2008). Munzel and Tamhane (2002) utilize a unified asymptotic theory of rank tests to derive large sample multiple comparison procedures (MCPs) for data with ties, particularly for at least ordinal data, whereas our procedure can handle nonmetric data as well. Munzel and Hothorn (2001) develop a multiple test procedure that is robust to the Behrens-Fisher problem. But their test procedure is slightly liberal in the many-to-one design which may be tolerated from a practical point of view. Based on the ideas discussed by Bretz et al.

(2001), regarding the numerical availability of multiple comparison procedures, Konietschke et al. (2012b) derived a multiple contrast test procedure for testing a general hypothesis H_0^p.

In what follows we take the following into consideration. Let $\Sigma = (\sigma_{jt})_{jt}$ for j, t = 1, 2, 3 be the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{p}}^{\theta} - \boldsymbol{p})$ and let $\hat{\Sigma} = (\hat{\sigma}_{\theta,jt})_{jt}$ for j, t = 1, 2, 3 be its estimator.

Under the null hypothesis, $H_o^p : \cap_{t=1}^{3}\{p_t = \frac{1}{2}\}$, and Assumption 3, the vector

$$J_n = \left(\sqrt{n}(\frac{\hat{p}_1^{\theta} - p_1}{\sqrt{\hat{\sigma}_{\theta,11}}}), \sqrt{n}(\frac{\hat{p}_2^{\theta} - p_2}{\sqrt{\hat{\sigma}_{\theta,22}}}), \sqrt{n}(\frac{\hat{p}_3^{\theta} - p_3}{\sqrt{\hat{\sigma}_{\theta,33}}}) \right)' \tag{4.18}$$

of the studentized effects $\hat{p}_t^{\theta}$, has, asymptotically, as $n \to \infty$, a trivariate normal distribution with expectation **0** and correlation matrix **R**, where elements of **R** are $r_{jt} = \frac{\sigma_{jt}}{\sqrt{\sigma_{jj}\sigma_{tt}}}$. Further, let $z_{1-\alpha/2,2,R}$ denote the two-sided equicoordinate quantile of the trivariate normal distribution with expectation **0** and correlation matrix **R** (Bretz et al., 2001). In the subsequent consideration, the unknown correlation matrix **R** is replaced with a consistent estimator $\hat{\mathbf{R}}$ of **R**. Here, we use

$$\hat{\mathbf{R}} = (\hat{r}_{\theta,jt})_{jt}; \quad \text{where} \quad \hat{r}_{\theta,jt} = \frac{\hat{\sigma}_{\theta,jt}}{\sqrt{\hat{\sigma}_{\theta,jj}\hat{\sigma}_{\theta,tt}}}.$$

For large sample sizes, the null hypothesis $H_0^p : \boldsymbol{Cp} = \mathbf{0}$ will be rejected at α level of significance, if

$$max\{|J_n|\} \geq z_{1-\alpha/2,2,\hat{R}}.$$

4.5.2 Simultaneous Confidence Intervals

Konietschke et al. (2012b) give a rank-based multiple contrast test procedure which is purely nonparametric that leads to a simultaneous confidence interval by taking the correlation between the test statistics. Besides, the test is robust in the presence of the Behrens-Fisher problem. In their previous work, Konietschke et al. (2010) came up with a procedure for testing the hypothesis of no time effect, $H_o^p : \boldsymbol{Cp} = \mathbf{0}$, and estimation of purely nonparametric effects in repeated measures designs in general.

Simultaneous confidence intervals for the effects p_t are obtained from

$$P\left(\bigcap_{t=1}^{3}\left\{p_t \in \left[\hat{p}_t^{\theta} \mp \frac{z_{\alpha/2,\hat{R}_n} \times \sqrt{\hat{\sigma}_{\theta,tt}}}{\sqrt{n}}\right]\right\}\right) \to 1-\alpha, \quad \text{as} \quad n \to \infty. \tag{4.19}$$

Also based on Theorems 5 and 6 , one can derive simultaneous confidence intervals for linear combinations of the effects p_t. Let $\boldsymbol{C}_m$ be a contrast matrix with row vectors C_l', and let $\Sigma^{(C)} = (\sigma^{\theta}_{l,h})_{l,h}$ $l, h = 1,\ldots, m$ be the asymptotic covariance matrix of $\sqrt{n}\boldsymbol{C}(\hat{\boldsymbol{p}}^{\boldsymbol{\theta}} - \boldsymbol{p})$. Furthermore, let $\hat{\Sigma}^{(C)} = (\hat{\sigma}^{\theta}_{l,h})_{l,h}$ $l, h = 1,\ldots, m$ be a consistent estimator of $\Sigma^{(\boldsymbol{C})} = \boldsymbol{C}\Sigma\boldsymbol{C}'$ and, let $\hat{\mathbf{R}}^{(\boldsymbol{C})} = (\hat{r}_{\theta,lh})_{lh}$ $l, h = 1,\ldots, m$ denote the empirical correlation matrix. We obtain simultaneous confidence intervals for the linear combinations $\boldsymbol{C}_l'\boldsymbol{p}$ from

$$P\left(\bigcap_{l=1}^{m}\left\{\boldsymbol{C}_l'\boldsymbol{p} \in \left[\boldsymbol{C}_l'\hat{\boldsymbol{p}}^{\boldsymbol{\theta}} \mp \frac{z_{\alpha/2,\hat{R}_n^{(\boldsymbol{C})}} \times \sqrt{\hat{\sigma}_{\theta,tt}}}{\sqrt{n}}\right]\right\}\right) \to 1 - \alpha, \quad \text{as n} \quad \to \infty. \tag{4.20}$$

4.6 Simulation Results

In this section we study the empirical properties of our asymptotic test procedures discussed in Section 4.5 via a simulation study. We make comparisons in terms of the size and power of tests. In the simulation study, the number of subjects and the degree of correlation between measurements will be allowed to vary. In addition, when testing the methods that are developed to work with missing data, we will vary the percentage of the missing data. By varying MP, sample size and within-pair correlation, we aim to identify the test procedure that works better in terms of maintaining the pre-determined level and having high power.

4.6.1 Size Simulation Results

To test the hypothesis of no time effect ($H_o : \boldsymbol{Cp} = \boldsymbol{0}$), various settings for missing percentage (MP), the distribution of data, sample sizes (10, 20, 30) and within-pair dependence were considered. The effect of missing percentage is evaluated by varying MP in $\{0\%, 5\%, 10\%, 25\%\}$. The covariance matrix was generated putting unity on the diagonal with different values on the off diagonal, i.e. (ρ = 0, 0.5 and 0.85). Furthermore, together with multivariate normal (MVN), both multivariate lognormal (MVLN) (as an example of a skewed distribution), and multivariate Cauchy (MVC) (as an example of a heavy-tailed distribution), are used for generating data. In all simulations, the run size is $10,000$. The proposed methods are compared, with respect to the pre-defined probability of type I error (0.05), are:

WTS: The Wald-Type statistic, as discussed in Brunner et al. (2002).

ATS: The ANOVA-Type statistic, the way discussed in Brunner et al. (2002).

$MCTP$: The Multiple Comparison Test Procedure as discussed in Konietschke et al. (2010).

$MCTPAdj.$: This is similar to MCTP where the MCTP is approximated by a t-distribution with approximated degrees of freedom $\hat{\nu}$.

As discussed in Section 4.5 the degrees of freedom for WTS is two which is the rank of contrast matrix C defined in equation (4.17). For the ATS method the degrees of freedom for Chisquare distribution is computed in the way discussed in Section 4.5. For the two methods, MCTP and MCTPAdj., the critical values to which the calculated values of test statistics are compared using equicoordinated quantiles, $z_{1-\alpha/2,2,\boldsymbol{R}}$, as discussed in Bretz et al. (2001), where $\boldsymbol{R}$ is estimated by sample correlation matrix $\hat{\boldsymbol{R}}$. The results of the size simulation for the tests are shown in Tables 4.2, 4.3 and 4.4 corresponding to the three distributions namely, Normal, Cauchy and Lognormal.

In Table 4.2, under normality, we see that WTS shows improvement with increase in the sample size when the missing percentages are up to 10% and then decline as the missing percent increases. This behavior of WTS is similar for both Cauchy and Lognormal distributions.

	n	%	WTS			ATS			MCTP			MCTP Adj.		
ρ			0	0.5	0.85	0	0.5	0.85	0	0.5	0.85	0	0.5	0.85
Normal	10	0	10.12	6.23	3.2	3.75	2.63	0.96	2.35	2.34	2.78	2.35	2.34	2.8
		5	11.46	8.16	5.06	4.83	3.28	1.72	2.26	2.44	2.66	2.27	2.45	2.67
		10	13	9.05	6.05	4.97	3.31	1.98	2.26	2.46	2.69	2.26	2.47	2.71
		25	13.61	11.46	9.86	4.8	3.68	3.2	2.75	2.69	2.5	2.77	2.72	2.5
	20	0	5.29	4.28	1.82	2.92	2.29	1.14	2.42	2.51	3.14	2.42	2.52	3.14
		5	7.17	5.51	3.39	3.63	3.29	1.85	2.26	2.3	2.69	2.27	2.3	2.69
		10	6.93	5.63	4.22	3.27	2.8	2.08	2.87	2.95	2.65	2.87	2.97	2.65
		25	4.91	4.27	5.09	1.79	1.57	2.1	4.1	3.94	3.94	4.11	3.96	3.96
	30	0	4.57	3.33	1.84	3.11	2.29	1.39	2.39	2.76	2.65	2.39	2.76	2.65
		5	5.69	4.35	3.6	3.46	2.68	2.38	2.97	2.49	2.93	2.99	2.49	2.93
		10	4.81	3.98	3.24	2.66	2.5	1.88	4.18	4.08	3.54	4.18	4.1	3.55
		25	2.29	1.82	2.9	0.79	0.77	1.04	5.76	6.26	5.14	5.79	6.29	5.15

TABLE 4.2: Achieved size ($\times 100$) at $\alpha = 0.05$ by WTS, ATS, MCTP and MCTP Adj.. Trivariate data are generated from Normal distribution with unstructured covariance. The simulation size is $10,000$.

Looking at Table 4.2, when data are from normal distribution, the simulation results show that the WTS method is liberal for small sample sizes while the ATS method preserved the pre-defined type-I error almost accurately when within-pair dependence is weak. On the other hand, for small sample sizes both MCTP and MCTP Adj. are conservative.

	n	%	WTS			ATS			MCTP			MCTP Adj.		
ρ			0	0.5	0.85	0	0.5	0.85	0	0.5	0.85	0	0.5	0.85
Cauchy	10	0	10.98	5.96	2.58	4.4	2.38	0.70	2.38	2.48	2.36	2.38	2.50	2.36
		5	12.08	8.04	4.34	5.08	2.78	1.66	2.12	2.22	2.44	2.12	2.22	2.44
		10	12.72	8.54	5.54	5.38	2.90	1.84	2.92	1.90	2.02	2.92	1.90	2.02
		25	14.02	11.2	9.48	5.54	3.58	2.98	2.44	2.44	2.08	2.44	2.44	2.10
	20	0	6.18	4.36	1.26	3.04	2.44	0.80	2.52	2.26	2.44	2.54	2.26	2.44
		5	7.64	5.64	3.66	3.72	3.04	2.20	2.72	1.92	2.38	2.72	1.94	2.38
		10	8.04	6.06	4.12	3.72	3.64	2.38	2.88	2.72	2.32	2.90	2.74	2.32
		25	6.16	4.72	4.74	2.52	1.70	2.12	3.88	3.62	3.20	3.90	3.62	3.22
	30	0	5.54	3.48	1.48	3.02	2.38	1.28	2.76	2.28	2.82	2.76	2.28	2.82
		5	6.36	5.16	3.66	3.26	2.82	2.22	2.54	3.00	2.94	2.56	3.00	2.94
		10	5.00	4.60	4.06	2.60	2.28	2.14	3.92	3.42	3.56	3.92	3.42	3.56
		25	2.16	1.88	3.36	0.68	0.68	1.40	5.92	6.48	4.58	5.94	6.52	4.58

TABLE 4.3: Achieved size ($\times 100$) at $\alpha = 0.05$ by WTS, ATS, MCTP and MCTP Adj.. Trivariate data are generated from Cauchy distribution with unstructured covariance. The simulation size is $10,000$.

In Table 4.3, for data generated from Cauchy distribution, we see that WTS shows improvement with increase in the sample size when the missing percentages are up to 10% and then decline as the missing percent increases. This behavior of WTS is similar to that of data from normal distribution. ATS performed better than WTS for small sample size and when within-pair dependence is weak while both MCTP and MCTPAdj. are conservative. Furthermore, we note that WTS performs better in the presence of medium (size of correlation) within-pair dependence. On the other hand, the ATS performs better than WTS in maintaining the predefined type-I error for small sample sizes. The two statistics MCTP and MCTPAdj., show an unusual behavior with increasing rejection rate as the missing percentage increases. However, the improvement in performance has been achieved in maintaining the pre-defined size when the missing percentage is large. This is also true for the lognormal distribution.

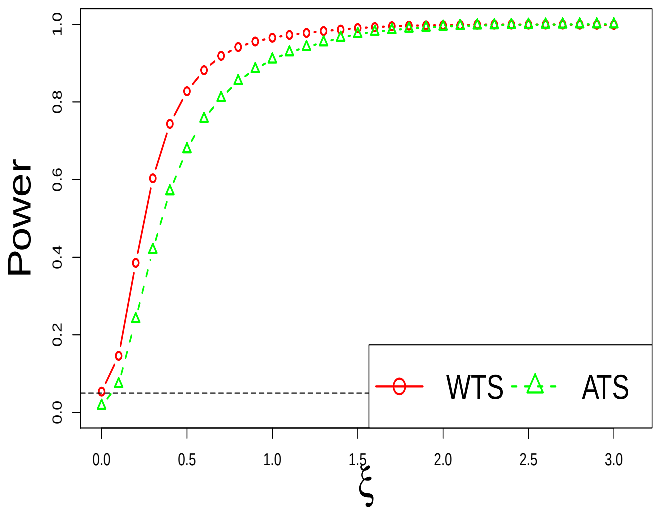
	n	%	WTS			ATS			MCTP			MCTP Adj.		
ρ			0	0.5	0.85	0	0.5	0.85	0	0.5	0.85	0	0.5	0.85
Lognormal	10	0	5.05	5.05	5.05	1.98	1.98	1.98	2.44	2.44	2.44	2.45	2.45	2.45
		5	6.80	6.80	6.80	2.70	2.70	2.70	2.41	2.41	2.41	2.41	2.41	2.41
		10	8.17	8.17	8.17	2.91	2.91	2.91	2.22	2.22	2.22	2.22	2.22	2.22
		25	10.40	10.40	10.40	3.61	3.61	3.61	2.25	2.25	2.25	2.26	2.26	2.26
	20	0	3.23	3.23	3.23	1.97	1.97	1.97	2.50	2.50	2.50	2.51	2.51	2.51
		5	4.51	4.51	4.51	2.60	2.60	2.60	2.13	2.13	2.13	2.13	2.13	2.13
		10	5.20	5.20	5.20	2.82	2.82	2.82	2.75	2.75	2.75	2.76	2.76	2.76
		25	4.18	4.18	4.18	1.48	1.48	1.48	4.28	4.28	4.28	4.31	4.31	4.31
	30	0	2.74	2.74	2.74	1.97	1.97	1.97	2.56	2.56	2.56	2.56	2.56	2.56
		5	4.53	4.53	4.53	2.67	2.67	2.67	2.62	2.62	2.62	2.62	2.62	2.62
		10	3.88	3.88	3.88	2.01	2.01	2.01	3.95	3.95	3.95	3.96	3.96	3.96
		25	2.01	2.01	2.01	0.84	0.84	0.84	6.03	6.03	6.03	6.04	6.04	6.04

TABLE 4.4: Achieved size ($\times 100$) at $\alpha = 0.05$ by WTS, ATS, MCTP and MCTP Adj.. Trivariate data are generated from Lognormal distribution with unstructured covariance. The simulation size is $10,000$.

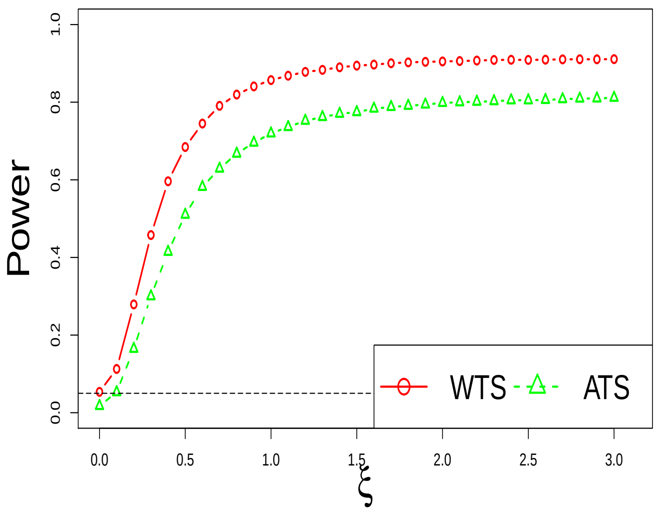
As can be seen in Table 4.4, for data generated from a lognormal distribution, WTS becomes more liberal as percentage of missingness increases for small sample sizes. MCTP and MCTPAdj. are conservative except when the sample size is large.

4.6.2 Power Simulation Results

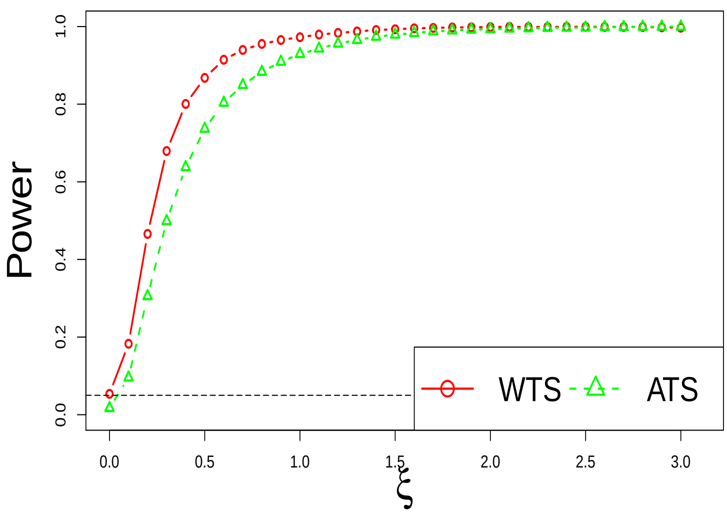
In Figures 4.1, 4.2 and 4.3 below power simulation for the new method is displayed for the three distributions namely, Normal, Cauchy and Generalized extreme value. To show an example of the power advantage of the new method, we set the sample size allocation to $n = 25$, missing percentage to $\mathrm{MP} = 20\%$ and $\alpha = 0.05$.



(A)



(B)



(C)

FIGURE 4.1: Power curves for the WTS and ATS. The alternatives considered are $F_1 = F_2$ whereas F_3 varies with respect to location and/or shape. The data is generated from skew-normal (SN) distribution and the panels (A), (B) and (C) are for location, shape and location-shape, respectively, alternatives. For all plots the sample size $n = 25$, the missing percentage is 20% and the simulation of size is $10,000$.

For the alternative hypothesis, we take two of the distributions to be the same but change the location, shape or both to the third one. We choose the third distribution in such a way that when the location and/or shape parameters are zeros, the power will be the same as the other two distributions. This scenario reduces to the null case and a desirable value of power would be 0.05 or less. The values of the scale parameter in all the three distributions is set to unity.

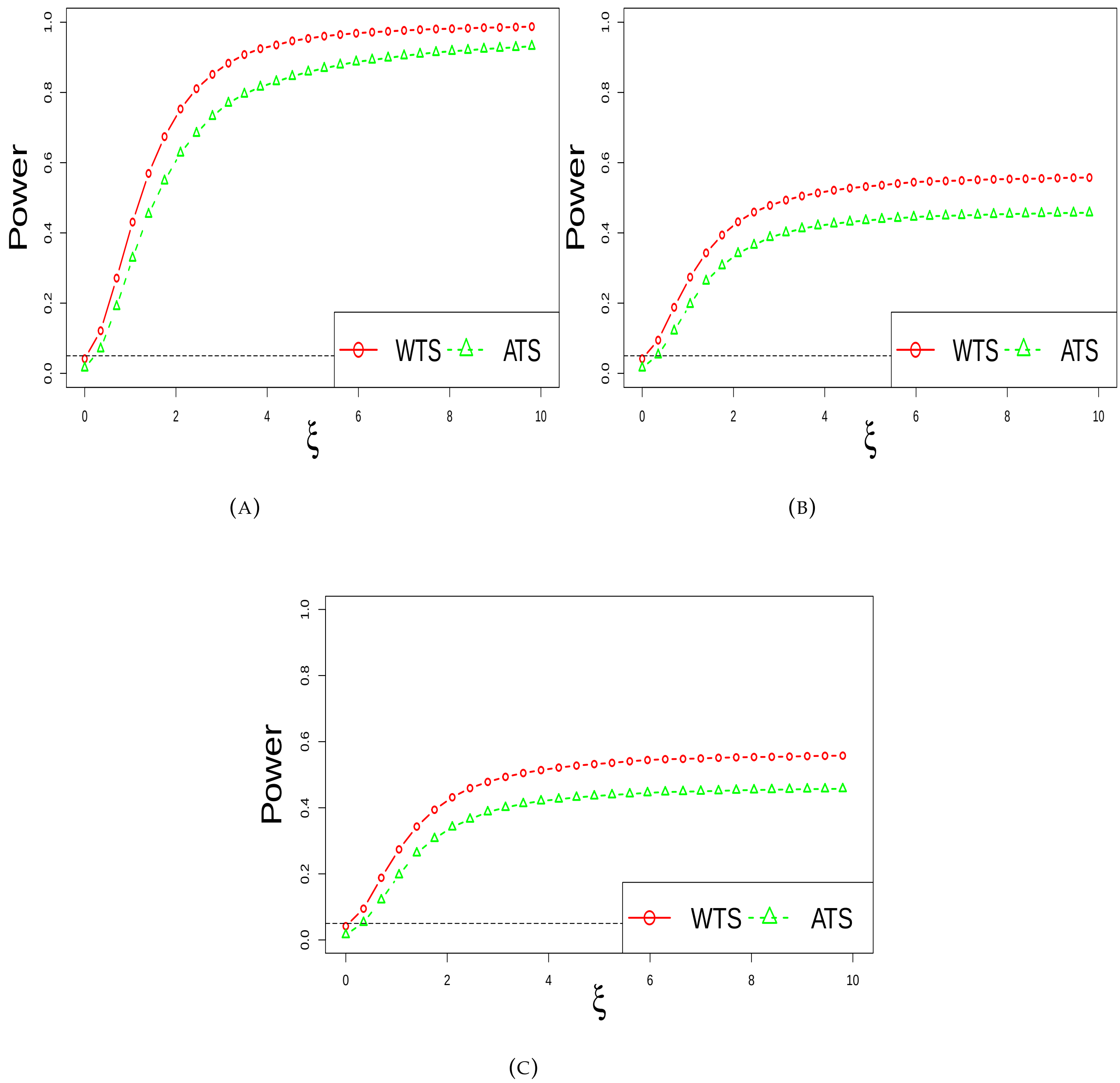
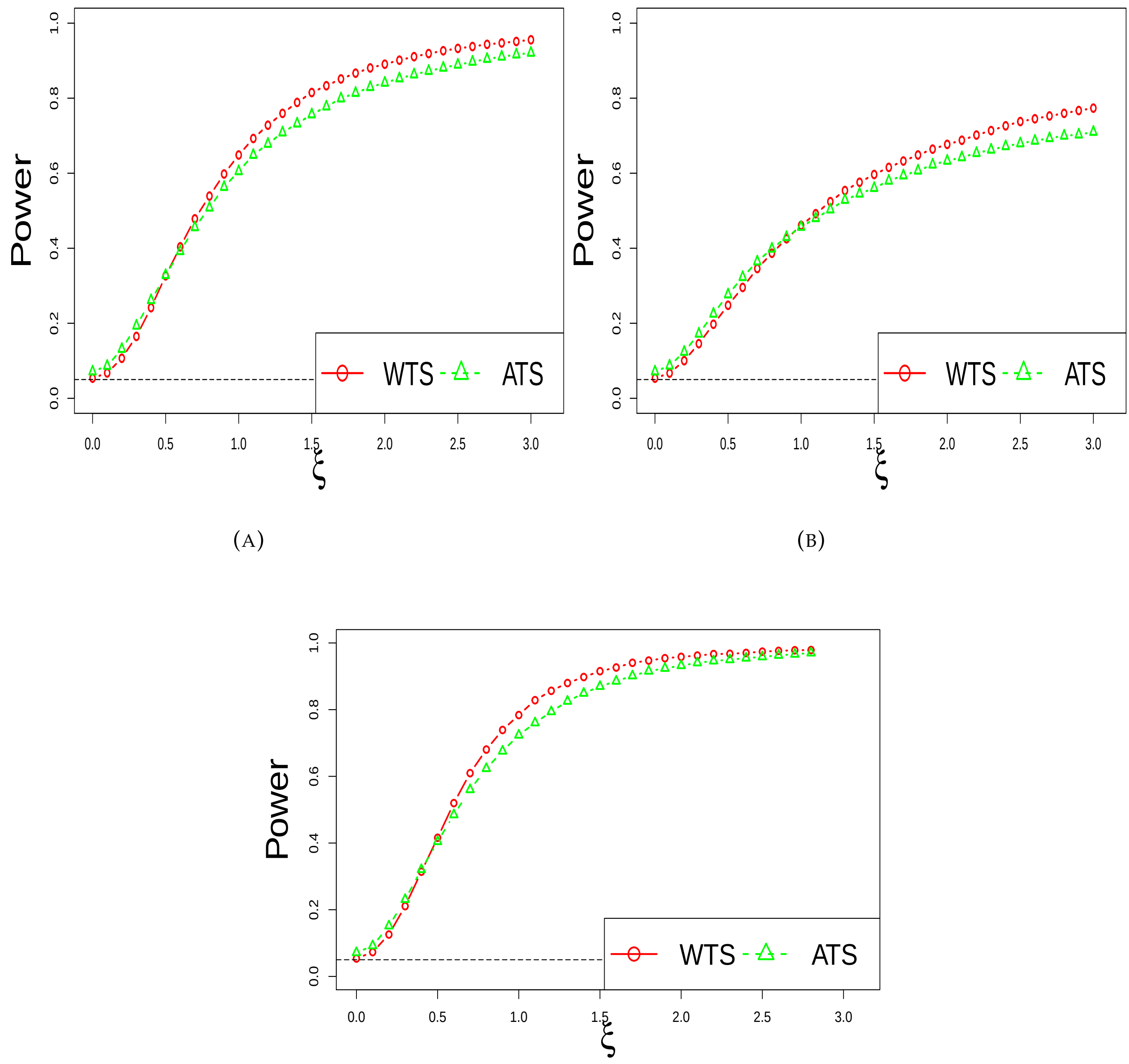


FIGURE 4.2: Power curves for the WTS and ATS. The alternatives considered are $F_1 = F_2$ whereas F_3 varies with respect to location and/or shape. The data is generated from skew-Cauchy (SC) distribution and the panels (A), (B) and (C) are for location, shape and location-shape, respectively, alternatives. For all plots the sample size $n = 25$, the missing percentage is 20% and the simulation of size is $10,000$.

In order to induce a measured amount of dependence within the paired observations, we generate data from the respective marginal distributions and couple them with Clayton's copula (Nelsen, 2006) setting the parameter value at 8. This value of the parameter induces a dependence of Kendal's $\tau = 0.8000$, and Spearman's $\rho = 0.9408$.



(A)

(B)

(C)

FIGURE 4.3: Power curves for the WTS and ATS. The alternatives considered are $F_1 = F_2$ whereas F_3 varies with respect to location and/or shape. The data is generated from generalized extreme value (GEV) distribution and the panels (A), (B) and (C) are for location, shape and location-shape, respectively, alternatives. For all plots the sample size $n = 25$, the missing percentage is 20% and the simulation of size is $10,000$.

The marginal distributions we consider are skew-normal (SN) of Azzalini (1985), (skew-t (ST) of Azzalini and Capitanio (2003), which is equivalent to Skew Cauchy (SC) when the degrees of freedom equals to 1), and generalized extreme value (GEV) of McFadden (1978).

Regarding the power simulation we limit our discussion to only two situations namely, WTS and ATS methods. The other two are dropped from our discussion as they displayed extremely poor performance in power. The results from the power simulations shown in Figures 4.1, 4.2 and 4.3 tell a similar story regarding the two test statistics (WTS and ATS). For location alternatives (panel (A) in Figures 4.1, 4.2 and 4.3), the new nonparametric method namely, WTS has a clear advantage over ATS when data come from heavy tailed distributions. The power curve for WTS starts at 0.05 and stays above that of ATS for panel (A) in Figures 4.2 and 4.3. For panels (A), (B) and (C) in Figure 4.3, although the power curve for WTS starts at value very close to 0.05, it attains a slightly higher power than ATS after a while. Overall, in the settings of these panels (panel (A) in Figures 4.1, 4.2 and 4.3), ATS performs comparably well but not as good as WTS. On the other hand, WTS performs better than ATS for shape-alone alternatives (panel (B) Figures 4.1 and 4.2). In panel (B) of Figure 4.3 the performances of the two are about the same. For data from heavy tailed distributions both methods (WTS and ATS) are inefficient for shape-alone alternatives. The results with data from GEV distribution (because the existence of moments depends on the value the shape parameter ξ), lead to poor power performance. In the case of shape and location alternatives both WTS and ATS perform almost similarly. Except the case in panel (C) of Figures 4.2, WTS has more power than ATS. Apart from GEV distribution (panels (A), (B) and (C) in Figures 4.3) in all the other cases ATS starts at power below the threshold value 0.05, showing that ATS is conservative in those settings. Finally, except for data from GEV distribution (panels (A), (B) and (C) in Figures 4.3), in all others the WTS method starts at values above and very close to 0.05. Hence ATS is liberal for data from this type of distribution. Overall, in the scenarios considered in the current simulation study WTS performs better than ATS.

4.7 Real Data Example: The Fluvoxamine Study

In this section we demonstrate the application of the new method using a real dataset. The Fluvoxamine study data are from a Multicentre Psychiatric Study (MPS) which includes a total of 315 patients. The patients were treated with a fluvoxamine for psychiatric symptoms that are believed possibly to result from abnormality of serotonin in the brain. These patients were treated at three subsequent visits to the clinic and at each visit the intensity of both therapeutic-effect and side-effects were recorded, each on a four-category ordinal scale. Side-effect is coded as: $1 \equiv$ no; $2 \equiv$ not interfering with functionality of patient; $3 \equiv$ interfering significantly with functionality of patient; $4 \equiv$ side-effect surpasses therapeutic effect. Likewise, the effect of therapy is recorded on a four-point ordinal scale: $1 \equiv$ no improvement or worsening; $2 \equiv$ minimal improvement, not changing functionality; $3 \equiv$ moderate improvement, partial disappearance of symptoms; and $4 \equiv$ important improvement, almost disappearance of symptoms. Thus, a side-effect occurs if new symptoms occur, whereas there is therapeutic effect if old symptoms disappear. The dataset is a good example of partially repeated measures as there was a considerable dropout at each visit. From a total of 315 subjects, 14 were not observed at all after the start, 31 were observed on the first time only, 44 patients showed up at first and second occasions, and 224 had complete observations. Two patients were dropped out from the analyses as the missing pattern is non-monotone. Consequently, 299 subjects remained for analyses. These data were previously used, among others, by Molenberghs and Lesaffre (1994), Molenberghs et al. (1997), Lapp et al. (1998), and Molenberghs and Verbeke (2004). For a detailed description of these data one can refer to Molenberghs et al. (1997). A non-parametric analysis would be an appropriate choice as the two outcomes therapeutic-effect and side-effect are a measure with an ordinal scale.

In Table 4.5 and Figure 4.4 results for therapeutic effect are shown. Table 4.5 shows a summary of the four test statistics (Wald-Type Statistic (WTS), Anova-Type Statistic (ATS), Multiple Comparison Test Procedure (MCTP) and Multiple Comparison Test Procedure approximated by *t*-distribution (MCTP Adj.) with adjustment for degrees of freedom. Figure 4.4 shows profile plots for the therapeutic-effect outcome variable in the Fluvoxamine study.

	WTS	ATS	MCTP	MCTP Adj.
Statistic	361.183	304.457	9.6667	9.6667
Critical-value	7.378	7.214	2.397	2.411

TABLE 4.5: The values of the test statistics and critical-values for WTS, ATS, MCTP and MCTP Adj.(the new method with approximated degrees of freedom ($\hat{\nu}$)) for the outcome variable 'therapeutic-effect'.

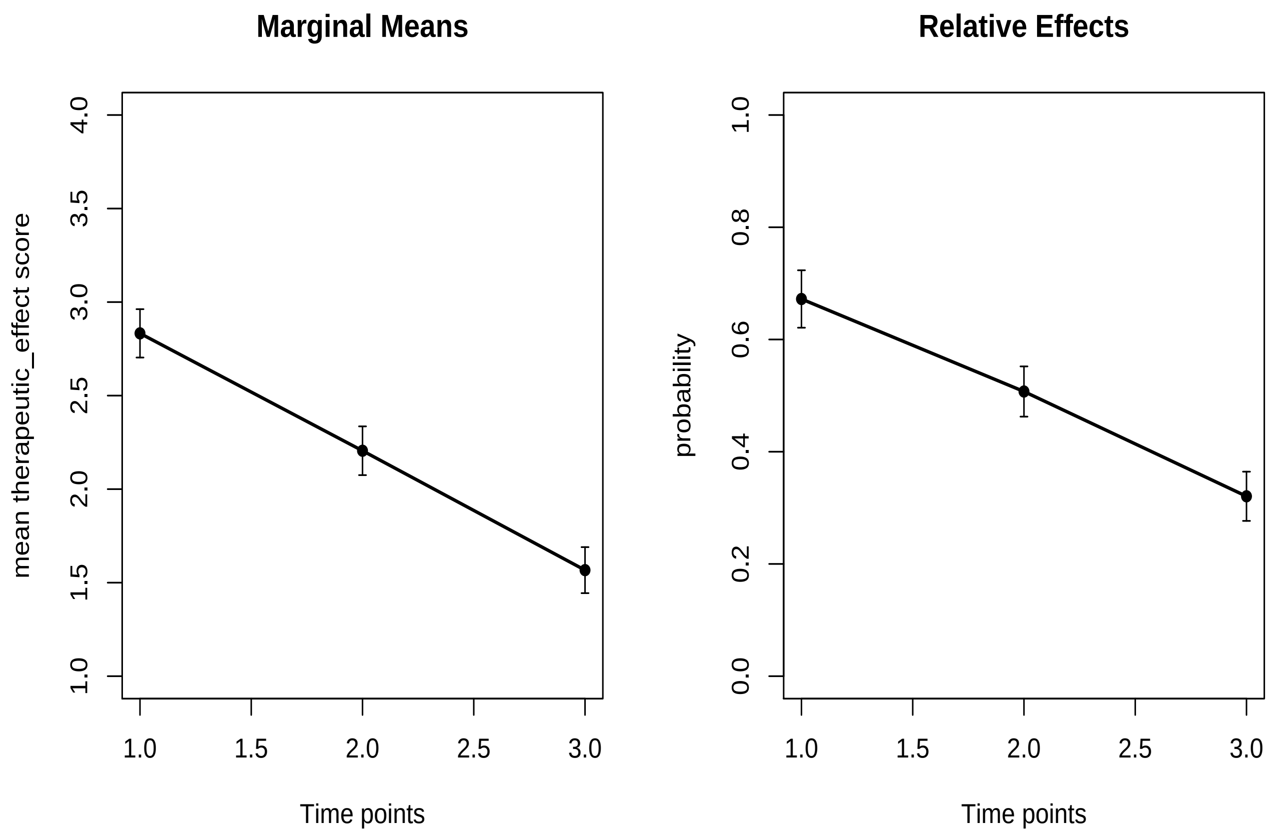


FIGURE 4.4: Profile plots for therapeutic-effect outcome variable.

In the analysis of Fluvoxamine data Table 4.6 below contains a summary of the results for four test statistics. All lead to the same conclusion: significant change over time. Figure 4.5 below shows the profile plots of the marginal means and relative effects for the side-effect outcome variable. They also confirm change over time.

	WTS	ATS	MCTP	MCTP Adj.
Statistic	41.876	19.738	3.215	3.215
Critical-value	7.378	6.324	2.344	2.363

TABLE 4.6: The values of the test statistics and critical-values for WTS, ATS, MCTP and MCTP Adj.(the new method with approximated degrees of freedom ($\hat{\nu}$)) for the outcome variable 'side-effect'.

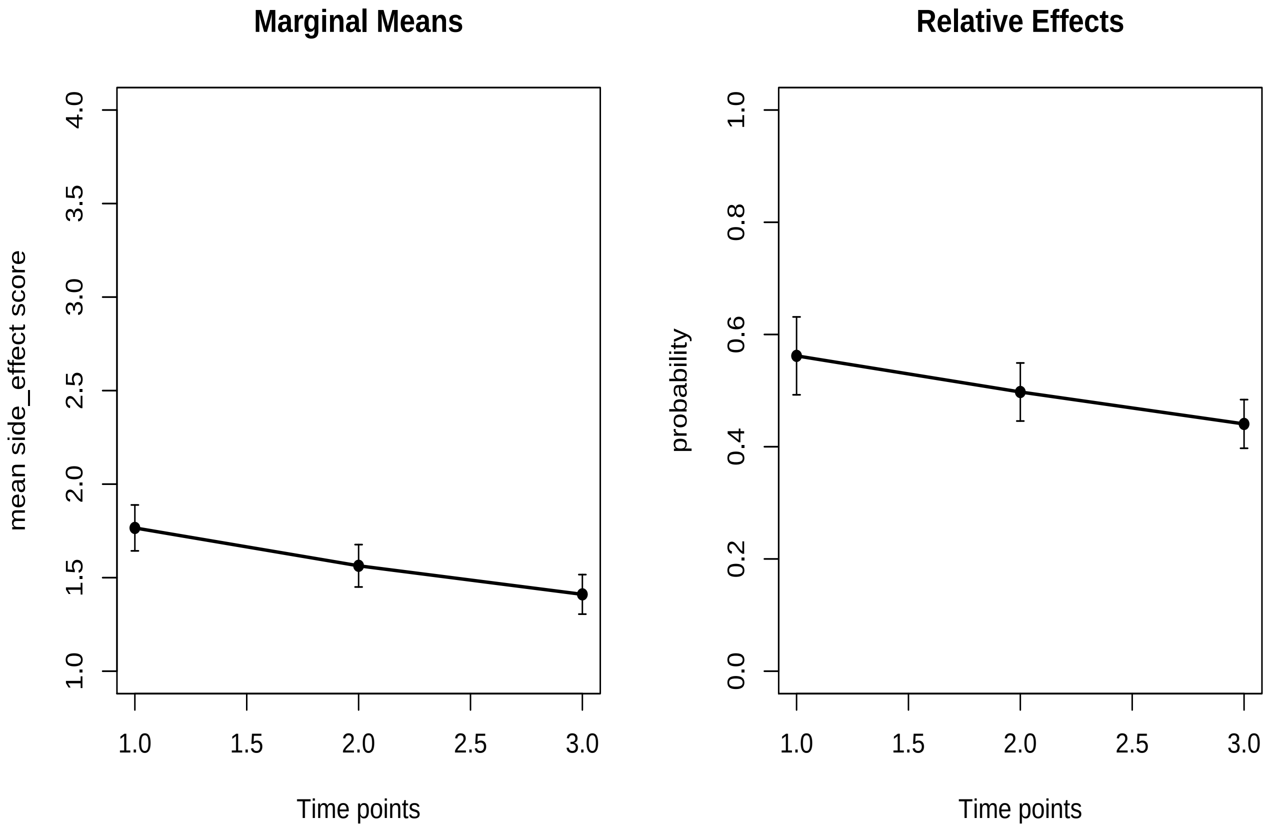


FIGURE 4.5: Profile plots of marginal means and relative effects for side-effect outcome.

As shown in Table 4.6 above, all the four test statistics lead to the same conclusion unequivocally. The results imply that there is a statistically significant change in the side-effect over time. Moreover, the profile plots in Figure 4.5 depict that there is a decline in the side-effects over time. This result is in agreement with that in Molenberghs and Verbeke (2004).

Chapter 5

Discussion , Conclusion and Future Directions

5.1 Discussion and Conclusion

We developed nonparametric methods for partially paired (correlated) data for two groups and multiple time situations in repeated measures design setup. Chapter 3 discussed inference procedures for the nonparametric relative effect for partially paired data in two groups. The method is particularly suited to the situation where two treatments are compared by assessing outcomes on two occasions. The main contribution is accommodation of data that are incomplete at either one of the occasions and in either of the two groups. In this connection, the procedures proposed are geared towards testing significance and constructing a confidence interval for the difference in the change induced by two treatments. The method can also be used in a 2×2 factorial design where one factor is a within-subject factor and the other factor is a between-subject factor. In this context, the design covered by the methods is an instance of incomplete block design. The method is appropriate for outcome variables that are binary, ordinal, discrete and continuous. When none of the subjects are assessed at both time points, the design reduces to a 2×2 cross-classified design. On the other hand, when all the subjects are assessed at both time points, the new method reduces to the repeated measures design discussed in Chapter 2. In both special cases the theoretical as well as numerical results firmly coincide and corroborate that of Brunner et al. (2017). Although our theory focuses on testing the interaction effect, the tests for the main effect follow along the same lines by considering the appropriate contrast (linear combination) of the relative effects.

Other recent methods for analyzing similar type of data compute the values of test statistics for the complete and incomplete cases separately and combine the two by taking a weighted average. This approach fails to compare directly, for example, when the observations at occasion 1 with complete cases do not match with those of the incomplete cases at occasion 2 and vice versa. Proposition 2 makes it clear that the methods in this study compares the complete and incomplete groups directly at individual level to make a more intuitive and powerful inference.

The estimator and test developed enjoy desirable asymptotic properties, such as asymptotic unbiasedness and consistency. The asymptotic variance is consistent whether or not the null hypothesis is true. This allows for the construction of a confidence interval. However, the test and confidence interval derived are asymptotic. The quality of approximation of the asymptotic result may not be adequate for small sample sizes. In such a cases, a t approximation with estimated degrees of freedom is proposed for the studentized version of the nonparametric measure of effect. Several authors have demonestrated the favorable utility of this approach for nonparametric methods.

In practical applications, we recommend carrying out group-wise marginal descriptive analysis to get a rough feel for the nature of the marginal distributions. When the descriptive statistics suggest mild or no departure from normality and the missing percentage is small, not much will be lost in terms of inferential efficiency by using the simple complete case analysis. On the other hand, if the departures from normality are moderate to severe but the tails of the distribution are not too heavy to violate the moment assumptions, the use of the semi-parametric methods may be appropriate, in particular, if only a location-type alternative is expected. However, when the marginal distributions are suspected to be different or are heavy tailed, the new nonparametric method should be used in the case of both small and larger sample sizes, when there is strong within-pair dependence. For smaller to moderate sample sizes, the t-based approximation should be considered when there is either negative or weak within-pair dependence. We should caution that one needs to stay away from the semi-parametric methods if shape-type alternative are indicated or suspected.

The second contribution of this dissertation is the theory developed in Chapter 4. We proposed and theoretically investigated a fully nonparametric test procedure for partially complete repeated measured data. Likewise in Chapter 3, the proposed method is not limited to metric data. In addition, it does not require the assumption of continuity or symmetry for the underlying distribution functions. The theory accommodates ties in the data in a seamless fashion. Unlike several other marginal models for repeated measures, the use of nonparametric effects effectively addresses the Behrens-Fisher problem in the sense that hypotheses are in terms of a nonparametric quantity such that the marginal distributions can still be different even under the null hypothesis. The proposed estimators for the relative effects are asymptotically unbiased and consistent. The technique used in developing the test and deriving the associated theory makes a nontrivial application of existing results for paired samples with missing data and repeated measures for complete data. This formulation allows visualizing a formal extension beyond three time points. The asymptotic theory of vector of relative effects is utilized to construct a Wald-Type Statistics (WTS), an ANOVA-Type Statistic and Multiple Comparison Test Procedure (MCTP) for testing purpose. An approximate degrees of freedom test was also proposed based on MCTP (MCTPAdj).

The empirical results, albeit limited, show that WTS performs better compared to the other three for all distributions and settings considered. ATS performs fairly as good as WTS in many of the cases. The two statistics, MCTP and MCTPAdj exhibited unexpectedly strange behavior in the size simulations. To make things worse, their performance in terms of power was unacceptably poor. In summary, considering the excelling performance of WTS observed in the simulation, we recommend its use with caution until more comprehensive empirical or theoretical evidence become available.

5.2 Future Directions

Finally, we close this chapter by providing an itemized list of future directions.

1. All the test procedures developed in this research assumed that incompleteness (missing data) happens completely at random. Although this assumption can be justified in many cases, it could be unrealistic in some applications. In particular, the reason data are missing may depend on observed variables or, even on the missed values themselves. These translate into the possibility of having different marginal distributions at a time point depending on whether observations are available at other time points. A possible avenue to tackle the former mechanism of missing data could potentially involve nonparametric estimation of the conditional distribution at a time point given that other time points are observed. Existence of additional information on the missing mechanism outside of the data is needed. Incorporating this information in the nonparametric inference is not yet clear at this point.

2. It would be of interest from practical stand point to extend the procedures developed for testing group-time interaction for two-group and two-time point to multiple groups and multiple time points. This extension amounts to extending the methods in Chapter 4 to multiple groups and multiple time points. The vector of relative effects may be defined in a similar way. The techniques employed in that chapter could be formally adapted. However, the calculation of the estimator, the estimated asymptotic covariance matrix and investigation

of their theoretical properties will be prohibitive. A matrix-based approach where the missing data are technically avoided by using some missing incidence matrix could be a viable approach.

3. Data arising in randomized trials and survey designs could have a more complex layout than the crossed-factor design this dissertation focused on. In some home-based interventions, for example, houses (possibly with multiple inhabitants) are randomly assigned to interventions. This group randomization induces cluster correlation which needs to be accounted for in statistical analysis. One of the challenges is that the group sizes will typically be unequal, which leads to cluster correlation matrices whose dimensions vary from cluster to cluster. Practically, the techniques employed in Chapter 4 could be applied in this complex situation as well but the resulting calculations will be a far too involved.

Author's address:
Merga Belina FEYASA
Department of Statistics
Addis Ababa University
Addis Ababa, Ethiopia.
email: merga.belina@aau.edu.et

Appendix A

Appendix

Riemann-Stieltjes and Integration by Parts

In the evaluation of relative summary effects which is expressed in terms integrals with respect to distribution function, we use the result given below by Hewitt (1960) on Riemann-Stieltjes integration by parts.

Let f and g be finite real-valued functions defined on the closed interval [a, b] of the real number system $\boldsymbol{R}$. Let $\int_a^b f(x)dg(x)$ denote the Riemann-Stieltjes integral of f with respect to g over [a, b]. Then, if $\int_a^b g(x)df(x)$ exists, the integral $\int_a^b g(x)df(x)$ also exists, and

$$\int_a^b g(x)df(x) + \int_a^b f(x)dg(x) = f(b)g(b) - f(a)g(a).$$

We use this result for derivation of interaction effect size, relative effects and relative summary effects.

Bibliography

Akritas, Michael, Efi Antoniou, and Jouni Kuha (2006). "Nonparametric Analysis of Factorial Designs With Random Missingness: Bivariate Data". In: *Journal of the American Statistical Association* 101.476, pp. 1513–1526.

Akritas, Michael G. and Steven F. Arnold (1994). "Fully Nonparametric Hypotheses for Factorial Designs I: Multivariate Repeated Measures Designs". In: *Journal of the American Statistical Association* 89.425, pp. 336–343.

Akritas, Michael G. and Edgar Brunner (1997). "A unified approach to rank tests for mixed models". In: *Journal of Statistical Planning and Inference* 61.2, pp. 249 –277.

Akritas, Michael G., Jouni Kuha, and D. Wayne Osgood (2002). "A Nonparametric Approach to Matched Pairs with Missing Data". In: *Sociological Methods & Research* 30.3, pp. 425–454.

Amro, Lubna and Markus Pauly (2017). "Permuting incomplete paired data: a novel exact and asymptotic correct randomization test". In: *Journal of Statistical Computation and Simulation* 87.6, pp. 1148–1159.

Azzalini, A. (1985). "A Class of Distributions Which Includes the Normal Ones". In: *Scandinavian Journal of Statistics* 12.2, pp. 171–178.

Azzalini, Adelchi and Antonella Capitanio (2003). "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 65.2, pp. 367–389.

Billingsley, P. (1995). *Probability and Measure*. New York: John Wiley & Sons.

Bretz, Frank, Alan Genz, and Ludwig A. Hothorn (2001). "On the Numerical Availability of Multiple Comparison Procedures". In: *Biometrical Journal* 43.5, pp. 645–656.

Browne, Richard H. (2010). "The t-Test p Value and Its Relationship to the Effect Size and P(X > Y)". In: *The American Statistician* 64.1, pp. 30–33.

Brunner, E. and N. Neumann (1982). "Rank tests for correlated random variables". In: *Biometrical Journal* 24, pp. 373–389.

Brunner, Edgar and Ullrich Munzel (2000). "The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation". In: *Biometrical Journal* 42.1, pp. 17–25.

Brunner, Edgar and Madan L. Puri (1996). “19 Nonparametric methods in design and analysis of experiments”. In: *Design and Analysis of Experiments*. Vol. 13. Handbook of Statistics. Elsevier, pp. 631 –703.

– (2001). “Nonparametric methods in factorial designs”. In: *Statistical Papers* 42.1, pp. 1–52.

Brunner, Edgar, Holger Dette, and Axel Munk (1997). “Box-Type Approximations in Nonparametric Factorial Designs”. In: *Journal of the American Statistical Association* 92.440, pp. 1494–1502.

Brunner, Edgar, Ulrich Munzel, and Madan L Puri (1999). “Rank-Score Tests in Factorial Designs with Repeated Measures”. In: *Journal of Multivariate Analysis* 70.2, pp. 286 –317.

Brunner, Edgar, Sebastian Domhof, and Frank Langer (2002). *Nonparametric analysis of longitudinal data in factorial experiments*. New York: John Wiley & Sons.

Brunner, Edgar, Frank Konietschke, Markus Pauly, and Madan L. Puri (2017). “Rank-based procedures in factorial designs: hypotheses about non-parametric treatment effects”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.5, pp. 1463–1485.

Conover, W. J. and Ronald L. Iman (1981). “Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics”. In: *The American Statistician* 35.3, pp. 124–129.

Davis, Charles S (2002). *Statistical methods for the analysis of repeated measurements*. New York: Springer.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1, pp. 1–38.

Derrick, Ben, Deirdre Toher, and Paul White (2017). “How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017)”. In: *The Quantitative Methods for Psychology* 13.2, pp. 120–126.

Diggle, P., P. Heagerty, K.Y. Liang, and S. Zeger (2013). *Analysis of Longitudinal Data*. 2nd ed. New York: Oxford University Press Inc.

Fligner, Michael A. and George E. Policello (1981). “Robust Rank Procedures for the Behrens-Fisher Problem”. In: *Journal of the American Statistical Association* 76.373, pp. 162–168.

Fong, Youyi, Ying Huang, Maria P. Lemos, and M. Juliana Mcelrath (2017). “Rank-based two-sample tests for paired data with missing values”. In: *Biostatistics*, kxx039.

Fuchs, Nicole, Werner Pölz, and Arne C. Bathke (2017). “Confidence intervals for population means of partially paired observations”. In: *Statistical Papers* 58.1, pp. 35–51.

Gao, Xin (2007). “A Nonparametric Procedure for the Two-Factor Mixed Model with Missing Data”. In: *Biometrical Journal* 49.5, pp. 774–788.

Gao, Xin and Mayer Alvo (2005). “A nonparametric test for interaction in two-way layouts”. In: *Canadian Journal of Statistics* 33.4, pp. 529–543.

– (2008). “Nonparametric multiple comparison procedures for unbalanced two-way layouts”. In: *Journal of Statistical Planning and Inference* 138.12, pp. 3674 –3686.

Harris, Kari Jo, Delwyn Catley, Glenn E. Good, Nikole J. Cronk, Solomon Harrar, and Karen B. Williams (2010). “Motivational interviewing for smoking cessation in college students: A group randomized controlled trial”. In: *Preventive Medicine* 51.5, pp. 387 –393.

Hewitt, Edwin (1960). “Integration by Parts for Stieltjes Integrals”. In: *The American Mathematical Monthly* 67.5, pp. 419–423.

Hollander, Myles, Douglas A. Wolfe, and Eric Chicken (2015). *Non-parametric Statistical Methods*. New Jersey: John Wiley & Sons, Inc.

Konietschke, F., A.C. Bathke, L.A. Hothorn, and E. Brunner (2010). “Testing and estimation of purely nonparametric effects in repeated measures designs”. In: *Computational Statistics & Data Analysis* 54.8, pp. 1895 –1905.

Konietschke, F., S.W. Harrar, K. Lange, and E. Brunner (2012a). “Ranking procedures for matched pairs with missing data – Asymptotic theory and a small sample approximation”. In: *Computational Statistics & Data Analysis* 56.5, pp. 1090 –1102.

Konietschke, Frank and Edgar Brunner (2009). “Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes”. In: *Computational Statistics & Data Analysis* 53.3, pp. 730 –741.

Konietschke, Frank, Ludwig A. Hothorn, and Edgar Brunner (2012b). “Rank-based multiple test procedures and simultaneous confidence intervals”. In: *Electron. J. Statist.* 6, pp. 738–759.

Konietschke, Frank, Lubna Amro, and Markus Pauly (2018). “Multiplication-Combination Tests for Incomplete Paired Data”. In:

Lapp, Krista, Geert Molenberghs, and Emmanuel Lesaffre (1998). “Models for the association between ordinal variables”. In: *Computational Statistics & Data Analysis* 28.4, pp. 387 –411.

Little, Roderick J. A. and Donald B. Rubin (2002). *Statistical Analysis with Missing Data*. New Jersey: John Wiley & Sons, Inc.

Matthews, J. N. S. (1993). “A refinement to the analysis of serial data using summary measures”. In: *Statistics in Medicine* 12.1, pp. 27–37.

McFadden, Daniel (1978). “Modelling the Choice of Residential Location”. In: *Transportation forecasting and travel behavior*, pp. 72–77.

Molenberghs, G., M. G. Kenward, and E. Lesaffre (1997). “The Analysis of Longitudinal Ordinal Data with Nonrandom Drop-Out”. In: *Biometrika* 84.1, pp. 33–44.

Molenberghs, Geert and Emmanuel Lesaffre (1994). "Marginal Modeling of Correlated Ordinal Data Using a Multivariate Plackett Distribution". In: *Journal of the American Statistical Association* 89.426, pp. 633–644.

Molenberghs, Geert and Geert Verbeke (2004). "Meaningful Statistical Model Formulations for Repeated Measures". In: *Statistica Sinica* 14.3, pp. 989–1020.

Munzel, U. (1999). "Nonparametric methods for paired samples". In: *Statistica Neerlandica* 53.3, pp. 277–286.

Munzel, Ullrich and Ludwig A. Hothorn (2001). "A Unified Approach to Simultaneous Rank Test Procedures in the Unbalanced One-way Layout". In: *Biometrical Journal* 43.5, pp. 553–569.

Munzel, Ullrich and Ajit C. Tamhane (2002). "Nonparametric Multiple Comparisons in Repeated Measures Designs for Data with Ties". In: *Biometrical Journal* 44.6, pp. 762–779.

Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer.

Noonan, Curtis W. and Tony J. Ward (2012). "Asthma randomized trial of indoor wood smoke (ARTIS): Rationale and methods". In: *Contemporary Clinical Trials* 33.5, pp. 1080 –1087.

Noonan, Curtis W., Erin O. Semmens, Paul Smith, Solomon W. Harrar, Luke Montrose, Emily Weiler, Marcy McNamara, and Tony J. Ward (2017). "Randomized Trial of Interventions to Improve Childhood Asthma in Homes with Wood - burning Stoves." In: *Environmental Health Perspectives* 125.9, pp. 1–9.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Ruymgaart, F. H. (1980). "A unified approach to the asymptotic distribution theory of certain midrank statistics". In: *j-LECT-NOTES-MATH* 821, pp. 1–18.

Ryu, Euijung (2009). "Simultaneous confidence intervals using ordinal effect measures for ordered categorical outcomes". In: *Statistics in Medicine* 28.25, pp. 3179–3188.

Ryu, Euijung and Alan Agresti (2008). "Modeling and inference for an ordinal effect size measure". In: *Statistics in Medicine* 27.10, pp. 1703–1717.

Samawi, Hani M. and Robert Vogel (2014). "Notes on two sample tests for partially correlated (paired) data". In: *Journal of Applied Statistics* 41.1, pp. 109–117.

– (2015). "On the Inference of Partially Correlated Data with Applications to Public Health Issues". In: *Innovative Statistical Methods for Public Health Data*. Springer International Publishing, pp. 31–55.

Thompson, G. L. (1991). "A Unified Approach to Rank Tests for Multivariate and Repeated Measures Designs". In: *Journal of the American Statistical Association* 86.414, pp. 410–419.

Thompson, G.L (1990). "Asymptotic distribution of rank statistics under dependencies with multivariate application". In: *Journal of Multivariate Analysis* 33.2, pp. 183 –211.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Vossoughi, Mehrdad, SMT Ayatollahi, Mina Towhidi, and Farzaneh Ketabchi (2012). "On summary measure analysis of linear trend repeated measures data: performance comparison with two competing methods". In: *BMC Medical Research Methodology* 12.1, p. 33.

Xu, Jin and Solomon W. Harrar (2012). "Accurate mean comparisons for paired samples with missing data: An application to a smoking-cessation trial". In: *Biometrical Journal* 54.2, pp. 281–295.