6-1-2023

Nonparametric Tests for Replicated Latin Squares

Joseph Yang

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Design of Experiments and Sample Surveys Commons, and the Statistical Methodology Commons

Recommended Citation

https://scholarworks.wmich.edu/dissertations/3973

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
Two classes of nonparametric procedures for a replicated Latin square design that test for both general and increasing alternatives are developed. The two classes of procedures are similar in the sense that both transform the data so that existing well-known tests for randomized complete block designs can be utilized. On the other hand, the two classes differ in the way that the data is transformed - one class essentially aggregates the data while the other class aligns the data. Within these contexts, the exact distributions and asymptotic distributions are discussed, when applicable. The exact distributions are easily computed using the R statistical software. Type I error rates and power estimates were computed via simulation for several design variations (including error terms that follow a contaminated normal distribution). The simulations show that the proposed methods have stable Type I error rates. Regarding power, the simulation results indicate that the proposed ordered alternative tests outperform their general alternative counterparts when the treatment effects are indeed ordered. Moreover, the proposed methods can outperform the parametric versions of the tests for heavily contaminated normal distributions; although the class of tests based on
aligning observations tend to have higher power than the tests based on aggregating the data. The proposed methods are illustrated with data from the existing literature.
NONPARAMETRIC TESTS FOR REPLICATED LATIN SQUARE DESIGNS

by

Joseph Yang

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Statistics
Western Michigan University
June 2023

Doctoral Committee:

Jeff Terpstra, Ph.D., Chair
Kevin Lee, Ph.D.
Joshua Naranjo, Ph.D.
Mike Tarn, Ph.D.
ACKNOWLEDGMENTS

First of all, I thank Professor Jeff Terpstra for his vast investment of time and effort in educating me. The completion of this dissertation would not be possible without his untiring guidance. I am also greatly indebted to Professors Kevin Lee, Josh Naranjo, and Mike Tarn, who served on my dissertation committee. Their insightful questions and invaluable suggestions are very much appreciated.

Throughout my five years at WMU, my parents have provided me with unfailing love and support. My father, Dr. Tian-Shiang Yang is a great source of inspiration and encouragement on my journey of education and heavily influenced my teaching style and research approach. My mother, Whei-Shan never forgot to set me on the right path in life, and I appreciate the wise advice from her. Shiuan-Ju, a longtime friend since college, has also given me encouragement and support throughout my journey as a student.

I would not be able to complete this degree without my colleagues and peers. I would especially like to thank Charles, who came to WMU the same semester as I did, and shared the journey from start to finish. Xi and Guohao also shared a lot of career and academic advice with me, and our gatherings during my last year at WMU have
been precious.

Lastly, I would like to thank the members of the Taiwanese Student Association. It was an honor to serve as president from 2018 to 2022, which provided me with a lot of leadership experience. Most notably, I thank Johnny and Jou-Chen for being with me during the tough times, and sharing many more good times with me.

Joseph Yang
# Table of Contents

**ACKNOWLEDGMENTS** ............................................................... ii

**LIST OF TABLES** ................................................................. vi

1 Introduction ................................................................. 1

2 Background and Literature Review ........................................... 4
  2.1 Review of Literature for Aggregation Methods ....................... 5
  2.2 Review of Literature for Alignment Methods ......................... 6

3 Methodology ................................................................. 9
  3.1 Existing Least-Squares Methods ......................................... 9
  3.2 Tests Based on Aggregating Observations ............................ 11
  3.3 Tests Based on Aligning Observations ................................. 13

4 Data Application Examples .................................................. 21
  4.1 Application to Apple Sales Data ....................................... 21
    4.1.1 Least-Squares Method Results ................................... 23
    4.1.2 Aggregated Test Class Results .................................. 24
    4.1.3 Aligned Test Results ............................................. 25
  4.2 Application to Crossover Designs ..................................... 30
# Table of Contents—Continued

5 Simulation Studies .................................................. 34
   5.1 Simulation under \( H_0 \) with Normally-Distributed Error .......... 36
   5.2 Simulation under \( H_A \) with Normally-Distributed Error .......... 37
   5.3 Simulation under \( H_0 \) with Contaminated-Normal Error .......... 39
   5.4 Simulation under \( H_A \) with Contaminated-Normal Error .......... 41

6 Future Work and Conclusion ........................................ 48
   6.1 Future Work .......................................................... 48
   6.2 Conclusion ............................................................ 50

7 Appendix ...................................................................... 52
   7.1 Proof that the MSE is a Consistent Estimator for \( \sigma^2 \) .......... 52
   7.2 Proof that \( \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \) for Aligned JT Test .... 56
   7.3 Proof that \( S_n(\sqrt{n}(\hat{\beta} - \beta)) - S_n(0) = o_p(1) \) for the Aligned JT Test 57
   7.4 Proof that \( S_n(\sqrt{n}(\hat{\beta} - \beta)) - S_n(0) = o_p(1) \) for Aligned Friedman Test 69
   7.5 A Uniform Convergence Theorem .................................... 80

BIBLIOGRAPHY ............................................................ 85
List of Tables

1.1 Example of a Latin Square Design ........................................ 1
3.1 Comparison of Exact and Normal Approximation Tail Probabilities for
the Aggregated Test Statistic ............................................ 13
4.1 Apple Sales Data .......................................................... 22
4.2 Least-squares Methods Results ......................................... 23
4.3 RCBD of Aggregated Observations .................................... 24
4.4 Ranks of Aggregated Observations .................................... 24
4.5 Test Results for Aggregated Test Class ............................... 25
4.6 Observations Aligned with Means ..................................... 27
4.7 Observations Aligned with Medians ................................... 28
4.8 Observations Aligned with Hodges-Lehmann ......................... 29
4.9 Test Results for Aligned Test Class ................................... 30
4.10 Cardiovascular Health Data as a Replicated Latin Square Design . 32
4.11 Least-squares Methods Results for Crossover Design ............... 32
4.12 Aggregated Class Test Results for Crossover Design .............. 33
4.13 Aligned Class Test Results for Crossover Design ................... 33
5.1 Type I Error Rate Comparison ........................................... 36
5.2 Power Comparison under Small Alternative ................. 37
5.3 Power Comparison under Medium Alternative ................. 38
5.4 Power Comparison under Large Alternative ................. 39
5.5 Type I Error Rate Comparison with Less-Contaminated-Normal Error . 40
5.6 Type I Error Rate Comparison with More-Contaminated-Normal Error 41
5.7 Power Comparison under Small Alternative with Less-Contaminated-Normal Error . ............................................. 42
5.8 Power Comparison under Small Alternative with More-Contaminated-Normal Error . ............................................. 43
5.9 Power Comparison under Medium Alternative with Less-Contaminated-Normal Error . ............................................. 44
5.10 Power Comparison under Medium Alternative with More-Contaminated-Normal Error . ............................................. 45
5.11 Power Comparison under Large Alternative with Less-Contaminated-Normal Error . ............................................. 46
5.12 Power Comparison under Large Alternative with More-Contaminated-Normal Error . ............................................. 47
7.1 Latin Square with Arranged Columns ......................... 60
Chapter 1

Introduction

In the study of experimental design, a common method to reduce the residual error of an experiment is to utilize the blocking principle to recognize controllable nuisance variables and remove variability. The Latin square design is used to eliminate two nuisance sources of variability. That is, it systematically allows blocking in two directions (rows and columns).

In general, a Latin square with \( p \) levels, or a \( p \times p \) Latin square, is a square containing \( p \) rows and \( p \) columns. Each of the resulting \( p^2 \) cells contains one of the \( p \) letter treatments, and each letter occurs once and only once in each row and column. An example of a \( 3 \times 3 \) Latin square design is shown in Table 1.1.

<table>
<thead>
<tr>
<th></th>
<th>Col. 1</th>
<th>Col. 2</th>
<th>Col. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Row 2</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Row 3</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
</tbody>
</table>
Consider a replicated Latin square design where we let $Y$ denote the response variable of interest, $R$ denote the row blocking variable, $C$ denote the column blocking variable, $S$ denote the square (or replicate), and $T$ the treatment variable. The basic main effects model (without nested effects) is essentially $Y_{ijkl} = \mu + R_i + C_j + S_k + T_l + \varepsilon_{ijkl}$, where $\mu$ is the overall location parameter (not necessarily the mean) and the $\varepsilon_{ijkl}$ are random errors. Note that $l = l(i, j, k)$ is actually a function of $i, j,$ and $k$, but is still included for notation convenience. Furthermore, $i, j,$ and $l$ are elements of $\{1, 2, \ldots, p\}$ ($p \geq 2$) and $k \in \{1, 2, \ldots, n\}$ ($n \geq 2$). All of the subsequent asymptotic results will correspond to $n \to \infty$. Thus, the data consists of $n \times p$ Latin squares for a total of $N = np^2$ observations.

With the consideration of nested effects, the model has four different versions. The different versions correspond to how $R$ and $C$ are treated for the different squares. The different versions, say V1, V2, V3, and V4, are as follows.

- **V1.** The $p$ levels of $R$ and $C$ remain the same across the different squares. In this case, the model is still denoted as $Y_{ijkl} = \mu + R_i + C_j + S_k + T_l + \varepsilon_{ijkl}$.

- **V2.** The $p$ levels of $R$ are different for each square. That is, they are nested within square. In this case, we might denote the levels of $R$ for the $k$th square as $R_{1(k)}, R_{2(k)}, \ldots, R_{p(k)}$ and rewrite the model as $Y_{ijkl} = \mu + R_{i(k)} + C_j + S_k + T_l + \varepsilon_{ijkl}$, where the notation $i(k)$ denotes that the $i$th level of $R$ is nested within the $k$th level of $S$.

- **V3.** The $p$ levels of $C$ are different for each square. That is, they are nested within square. In this case, we might denote the levels of $C$ for the $k$th square as $C_{1(k)}, C_{2(k)}, \ldots, C_{p(k)}$ and rewrite the model as $Y_{ijkl} = \mu + R_i + C_{j(k)} + S_k + T_l + \varepsilon_{ijkl}$, where the notation $j(k)$ denotes that the $j$th level of $C$ is nested within the $k$th
level of $S$. Note that $V3$ essentially corresponds to $V2$ upon interchanging $R$ and $C$. Thus, $V2$ and $V3$ can essentially be treated as the same case.

- $V4$. The $p$ levels of $R$ and $C$ are different for each square. That is, they are nested within square. In this case, we might denote the levels of $R$ as in $V2$ and the levels of $C$ as in $V3$ and rewrite the model as $Y_{ijkl} = \mu + R_i(k) + C_j(k) + S_k + T_l + \varepsilon_{ijkl}$.

The levels of $T$ remain the same in all four versions of the model. More advanced versions (e.g., those involving certain interactions) are possible, but we will not consider those here. Instead, we focus primarily on $V2$-$V3$ since $V4$ has been considered by Monga and Tardif (1994). We will assume that all effects are fixed. Additionally, $\sum_{k=1}^{n} S_k = \sum_{i=1}^{p} T_l = 0$ for all versions of the model and $\sum_{i=1}^{p} R_i(k) = 0, k = 1, 2, \ldots, n$ in $V2$, while $\sum_{j=1}^{p} C_j(k) = 0, k = 1, 2, \ldots, n$ for $V3$.

In all versions, the $\varepsilon_{ijkl}$ are assumed to be independently and identically distributed (iid) according to a distribution function $F$ with density $f = F'$. For theoretical results, we assume that $f$ is bounded and uniformly continuous. Note that the distributions of $Y_{ijkl}$ (for $V1$) are connected through the relationship $F_{ijkl}(y) = F(y - \mu - R_i - C_j - S_k - T_l)$. For $V2$, we would have $F_{ijkl}(y) = F(y - \mu - R_i(k) - C_j - S_k - T_l)$. Similarly for $V3$.

Our analysis concerns the effectiveness of the treatment variable $T$. Both the general alternative (i.e. $H_A : T_l \neq 0$ for some $l$) and the ordered alternative (i.e. $H_A : T_1 \leq T_2 \leq \cdots \leq T_p$ (i.e. $T_1 < T_p$)) are considered. We will propose two new classes of tests for these hypotheses. These classes are based on aggregating observations with averages or aligning observations by subtracting coefficient estimates.
Chapter 2

Background and Literature Review

In this chapter, we review how our two proposed methods (i.e. aggregating or aligning observations) are used in different experimental designs. For the methodology used in the experimental designs, particularly the least-squares methods, we refer to the textbook by Montgomery (2013). The complete randomized design is discussed in Chapter 3, while blocking designs such as the randomized complete block design and Latin squares are discussed in Chapter 4. The relationship between crossover designs and Latin squares is also discussed in the book.

For the nonparametric methods pertaining to the one-way layout, we employ the test for general alternatives by Kruskal and Wallis (1952) and the test for ordered alternatives by Jonckheere (1954). Pertaining to the two-way layout, we employ the test for general alternatives by Friedman (1937) and the test for ordered alternatives by Page (1963). Fundamental results for the above tests are also listed in the textbook by Hollander et al. (2013).
2.1 Review of Literature for Aggregation Methods

The dissertation by Le (2019) introduces an averages-based Jonckheere-Terpstra test (Jonckheere (1954)) for a nested design. The method of averaging observations within the nesting factor to remove nested effects is similar to our aggregated Jonckheere-Terpstra test. The work mentions that the averages-based test is a multisample U-statistic $U_N$ where $E(U_N) = \frac{1}{2}$ and $Var(U_N)$ has to be estimated. The standardized version of their $U_N$ statistic is shown to have an asymptotic standard normal distribution under the null hypothesis with rejection for large values of $Z_{U_N}$.

When calculating the distribution of the aggregated test statistic, we refer to the paper regarding sums of discrete random variables with finite support by Terpstra (2005). The paper provides methods and R code for calculating the probability generating function. Relevant examples given in the paper include the use of the dsumdisc R function to calculate the exact Wilcoxon signed-rank distribution and the exact distribution of Page’s test statistic.

The article by Tachibana (1985) discusses the effects of aggregating data in open-field test experiments on rodents. The author mentions that the aggregation of data is useful in enhancing reliability, as it reduces the influence of error associated with incidental factors. In the author’s experiment, seven measures are taken on a sample of rats over four days, with observations aggregated by summing across days for analysis. The main discussion in the article is through the comparison of correlation coefficients between aggregated and nonaggregated data. Large correlations are obtained in aggregation since the aggregation technique cancels out incidental, uncontrolled factors and thereby enhances reliability.

We may also view the aggregation method as an example of the marginalization
methods discussed in Basu (1977). The paper presents the marginalization method as a means of eliminating nuisance parameters, where a suitable design replaces the original design through data reduction. The justification for this data reduction is made through a partial sufficiency argument.

2.2 Review of Literature for Alignment Methods

The dissertation by Le (2019) also introduces an aligned rank Jonckheere-Terpstra test (Jonckheere (1954)). The method of aligning observations by subtracting estimated effects is similar to our alignment methods. This work also discusses the theory for aligning with estimates of the coefficients, where it is shown that the standardized test statistic has an asymptotic standard normal distribution. Simulation studies in the work showed that the proposed aligned rank test (incorporating an a-priori ordering) outperform the general alternative hypothesis test based on the reduction-in-dispersion F-test.

The paper by Fawcett (1990) discusses the use of alignment methods in a blocking design. The alignment method used in their paper is similar to our method of aligning by least-squares methods. That is, they plot residuals associated with each treatment centered at treatment means. A given example is the alignment of responses from a Graeco-Latin square design where the least-squares estimates for the row, column, and Greek-letter block effects are subtracted from the responses. The benefits and costs of blocking in the original design are illustrated by comparing the dot diagrams of the original and aligned responses, where it is observed that the aligned responses have reduced variability and suggests that the null hypothesis should have been rejected more decisively with a smaller $p$-value. In comparison to the alignment methods used in our proposed test class, we do not fully reduce the design to a complete randomized
design (for V2 and V3).

The paper by Best and Rayner (2011) examines alignment procedures for Latin squares. Their methods mainly focus on the design with one single square replicate, but the replicated Latin square design is briefly discussed. They test the general alternative and compare four possible nonparametric statistics (Kruskal-Wallis ignoring row and column effects (KW), Kruskal-Wallis adjusting for row and column effects (AKW), rank transform, and aligned rank transform F statistic (ARTF)). The p-values for their tests can be found using the asymptotic $\chi^2$ or F distributions, or through Monte Carlo simulation. The method of alignment used in the paper is through subtracting row, column, and overall means, which is similar to the alignment of least-squares estimates in our dissertation. Their adjusted KW statistic adjusts for row and column effects by using the aligned ranks. A power study shows that the KW test ignoring row and column effects has poor power, while the F, ARTF, and AKW tests have less power when outliers are present. The authors do note that an alignment procedure based on medians rather than means may improve the results, which we investigate in our dissertation through alignment with medians and Hodges-Lehmann estimates.

For our proof on the asymptotic normality of the aligned statistic with estimated parameters, we present a theorem that combines the ideas of Sievers (1983) and Newey (1991). Newey’s paper presents a stochastic equicontinuity condition that, together with pointwise convergence, characterizes uniform convergence in probability to equicontinuous functions on a compact set. However, Newey’s condition is not obvious in our context. Instead, we propose a condition similar to that used in Theorem 5.1 of Sievers to obtain uniform convergence.

The paper by Randles (1982) also provides methods for showing the asymptotic normality of statistics with estimated parameters. The author points out that the
limiting distribution’s dependence of the estimator is reliant on the nonzero derivative with respect to that variable of the limiting mean. A section contains conditions for the asymptotic normality of U-statistics with estimated parameters.

The paper by Tardif (1981) discusses the almost sure convergence of the permutation distribution of aligned rank statistics in a randomized complete block design (RCBD). The observations of a RCBD are aligned by subtracting the block means. The vector of ranks for the aligned observations is not distribution free under $H_0$, so the authors introduce a conditional permutational argument so that the distribution of the aligned observations over the intrablock permutations is equally likely. The paper generalizes results of Hodges and Lehmann (2012) and Mehra and Sarangi (1967) that (under weak conditions of the block means for alignment), the conditional permutation distribution of the test statistic almost surely and uniformly converges to a $\chi^2$-distribution as the number of blocks goes to infinity.
Chapter 3

Methodology

3.1 Existing Least-Squares Methods

Before we discuss the methodology of our proposed test classes, we will explore some of the traditional least-squares methods that will be used for comparison. Here, we simply fit the 'full model' and obtain least-squares estimates of the treatment vector $\hat{T} = (T_1, T_2, \ldots, T_p)^\top$. The test statistics are based on (approximately standardized) linear combinations of these effects.

Under the full model of a V2 design where $Y_{ijkl} = S_k + R_i(k) + C_j + T_l + \varepsilon_{ijkl}$, the $np^2$ error terms are iid with $E(\varepsilon_{ijkl}) = 0$ and variance $\sigma^2$. Consider $\bar{Y}_{..kl} - \bar{Y}_{..k}$. which follows a one-way design, $\bar{Y}_{..kl} - \bar{Y}_{..k} = T_l + (\bar{\varepsilon}_{..kl} - \bar{\varepsilon}_{..k})$. This balanced one-way design is then used to calculate the least-square (mean) estimates of the treatment effects. That is, $\hat{T}_l = \bar{Y}_{..l} - \bar{Y}_{...}$, $l = 1, 2, \ldots, p.$
Now, it can be shown that the covariance matrix of the error terms \((\bar{\varepsilon}_{..k1} - \bar{\varepsilon}_{..k})\) is

\[
\Sigma = \sigma^2 \begin{bmatrix}
\frac{p-1}{p^2} & -\frac{1}{p^2} & \cdots & -\frac{1}{p^2} \\
-\frac{1}{p^2} & \frac{p-1}{p^2} & \cdots & -\frac{1}{p^2} \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{1}{p^2} & \cdots & -\frac{1}{p^2} & \frac{p-1}{p^2}
\end{bmatrix}_{p \times p}
\]

It should be noted that \(\Sigma\) is not positive definite. By the central limit theorem, 
\(\sqrt{n}(\hat{T} - T) \sim N_p(0, \Sigma)\) when \(n \to \infty\). Proof 7.1 shows that the MSE \((\sigma^2)\) from the full model is a consistent estimator for \(\sigma^2\).

For testing the general alternative \(H_0: T_1 = T_2 = \cdots = T_p = 0\) vs. \(H_A: T_l \neq 0\) for at least one \(l\), consider the \((p-1) \times p\) matrix \(K^\top = [I_{p-1}, 0]\) where \(0\) is a \(p-1\) vector of zeroes. Then, \(\sqrt{n}(K^\top \hat{T} - K^\top T) \sim N_{p-1}(0, K^\top \Sigma K)\). An asymptotic test follows where the test statistic \(n(K^\top \hat{T} - 0)^\top \hat{\Sigma}^{-1}(K^\top \hat{T} - 0)\) has an asymptotic \(\chi^2_{p-1}\) distribution \((H_0\) is rejected for large values of the test statistic). Note that \(\hat{\Sigma}\) is the covariance matrix \(\Sigma\) where \(\sigma^2\) is substituted with the consistent estimator \(\hat{\sigma}^2\). For an exact test which assumes normality, the adjusted test statistic \(\frac{1}{p-1} \cdot n(K^\top \hat{T} - 0)^\top \hat{\Sigma}^{-1}(K^\top \hat{T} - 0)\) follows an \(F\)-distribution with \(p-1\) and \(np^2 - (n+2)p + 2\) (for \(V2\) and \(V3\)) degrees of freedom.

For testing the ordered alternative \(H_A: T_1 \leq T_2 \leq \cdots \leq T_p\) \(\left( T_1 < T_p \right)\), the maximin contrast defined by Abelson and Tukey (1963) is used, where \(K_p = (K_1, K_2, \ldots, K_p)^\top\) and \(K_l = ((l-1) \cdot (1-(l-1)/p))^{0.5} - (l \cdot (1-l/p))^{0.5}, l = 1, \ldots, p\). So an asymptotic \(z\)-test can be done where the standardized test statistic is \(\frac{K_p^\top T - K_p^\top T_0}{\sqrt{K_p^\top \Sigma K_p}/n} \sim N(0, 1)\) and \(K_p^\top T_0 = 0\) under \(H_0\). An exact \(t\)-test (which assumes normality) can be done where the standardized test statistic follows a \(t\)-distribution with \(np^2 - (n+2)p + 2\) (for \(V2\) and \(V3\)) degrees of freedom under \(H_0\).
3.2 Tests Based on Aggregating Observations

The aggregation of observations will be done by calculating \( \bar{Y}_{-kl} \), the average of the observations for treatment \( l \) in square \( k \). It follows (from the structure of the Latin square design and the linear constraint assumptions) that \( \bar{Y}_{-kl} = \mu + S_k + T_l + \varepsilon_{-kl} \) for V1-V4. Note that \( \{\varepsilon_{-kl}\} \) are iid according to some distribution that depends on \( F \). This "new" model is that of a randomized complete block design, where Friedman’s test (Friedman (1937)) (for the general alternative) and Page’s test (Page (1963)) (for the ordered alternative) may be performed (without modification) to test for significance of \( T \).

For Friedman’s test, the test statistic \( \chi^2_r = \frac{12}{np(p+1)} \sum_{l=1}^{p} (\sum_{k=1}^{n} R_{kl} - \frac{n(p+1)}{2})^2 \) where \( R_{kl} \) denotes the rank of the aggregated observation \( (\bar{Y}_{-kl}) \) among the \( l = 1, 2, \ldots, p \) treatments within the \( k \)th square. An asymptotic \( \alpha \) level test rejects \( H_0 \) if \( \chi^2_r \geq \chi^2_{1-\alpha(p-1)} \). An exact test is based on the relative frequency distribution of \( W = \frac{1}{n(p-1)} \chi^2_r \) for each of the \( (p!)^n \) possible rankings. To calculate this distribution, we use the \textit{pFrd} R function from the \texttt{NSM3} package (Schneider et al. (2022)).

For Page’s test, the test statistic \( L = \sum_{l=1}^{p} \sum_{n}^{n} R_{kl} \). Under \( H_0 \), \( L \) has expectation \( E_{H_0}(L) = \frac{np(p+1)^2}{4} \) and variance \( Var_{H_0}(L) = \frac{np^2(p+1)(p^2-1)}{144} \). An exact null distribution corresponds to the relative frequency of \( L \) for each of the \( (p!)^n \) rank configurations. By the Central Limit Theorem, \( \frac{L-E_{H_0}(L)}{\sqrt{Var_{H_0}(L)}} \xrightarrow{d} N(0,1) \) as \( n \to \infty \) under \( H_0 \). See section 7.2 of Hollander et al. (2013) for details.

As a competitor to Page’s test, we propose a test for the ordered alternative based on Jonckheere-Terpstra (JT) test statistics \( J_1, J_2, \ldots, J_n \) for each block. The new proposed test statistic for testing the ordered alternative is then the sum of the \( n \) test statistics, \( AGJ = \sum_{k=1}^{n} J_k \). In a completely randomized design with \( p \) treatment groups,
the Jonckheere-Terpstra statistic is a sum of \( p(p - 1)/2 \) Mann-Whitney counts, i.e.
\[
J = \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} U_{ij}
\]
where \( U_{ij} = \sum_{a=1}^{n_i} \sum_{b=1}^{n_j} I(X_{ia} < X_{jb}) = \sum_{b=1}^{n_j} R(X_{jb}) - \frac{n_j(n_j+1)}{2} \).

For the individual JT statistics, the exact (null) distribution of \( J \) is obtained by considering all \( p! \) (since \( n_1 = n_2 = \cdots = n_p = 1 \)) permutations of \( \{1, 2, \ldots, p\} \) under \( H_0 \). Our user-defined \texttt{dJCK} R function calculates the exact distribution of \( J_k \) for each square, while the \texttt{dsumdisc} function written by Terpstra (2005) calculates the exact distribution for the sum of the \( n \) iid \( J_k \)'s across the \( n \) squares. Hence, an exact null distribution for \( AGJ \) can be calculated.

Regarding the asymptotic distribution of \( AGJ \), recall that each block contains \( p \) observations (1 observation per treatment level). Hence, it follows from the known formulas for the JT statistic that
\[
E_{H_0}(J_k) = \frac{1}{4}(p^2 - p) \quad \text{and} \quad Var_{H_0}(J_k) = \frac{p^2(2p+3)-5p}{72}.
\]
Since the statistics \( J_1, \ldots, J_n \) are iid with finite mean and variance, the classical Central Limit Theorem (Hollander et al. (2013)) shows that
\[
\frac{\sum_{k=1}^{n} J_k - \frac{n}{4}(p^2 - p)}{\sqrt{n(2p^3 + 3p^2 - 5p)/72}} \xrightarrow{d} N(0, 1).
\]
Table 3.1 gives a comparison of the tail probabilities (i.e. \( P(AGJ \geq x) \)) from the exact and normal-approximated distribution of the test statistic for a replicated \( 3 \times 3 \) Latin square with three square replicates.
Table 3.1: Comparison of Exact and Normal Approximation Tail Probabilities for the Aggregated Test Statistic

<table>
<thead>
<tr>
<th>AGJ</th>
<th>Exact</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0000</td>
<td>0.9967</td>
</tr>
<tr>
<td>1</td>
<td>0.9954</td>
<td>0.9826</td>
</tr>
<tr>
<td>2</td>
<td>0.9676</td>
<td>0.9342</td>
</tr>
<tr>
<td>3</td>
<td>0.8843</td>
<td>0.8171</td>
</tr>
<tr>
<td>4</td>
<td>0.7222</td>
<td>0.6185</td>
</tr>
<tr>
<td>5</td>
<td>0.5000</td>
<td>0.3815</td>
</tr>
<tr>
<td>6</td>
<td>0.2778</td>
<td>0.1829</td>
</tr>
<tr>
<td>7</td>
<td>0.1157</td>
<td>0.0658</td>
</tr>
<tr>
<td>8</td>
<td>0.0324</td>
<td>0.0174</td>
</tr>
<tr>
<td>9</td>
<td>0.0046</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

The tail probabilities show that the normal approximation is a suitable approximation for this particular exact distribution.

### 3.3 Tests Based on Aligning Observations

The aligned tests start with an estimate of some subset of model effects, say $\hat{\beta}$, that satisfies the following property: $\sqrt{n}(\hat{\beta} - \beta) = O_p(1)$. In what follows $\beta$ corresponds to the unnested row and/or column effects. The model parameters (excluding the treatment effects) that are not associated with $\hat{\beta}$ are then used to form 'blocks'. For the ordered alternative, JT statistics are then obtained on the aligned observations in the block. These individual JT statistics are then summed to obtain the new test statistic. For the general alternative, Friedman’s test (Friedman (1937)) is used where each treatment
effect occurs just once in each block.

More specifically, the method is summarized for the design versions below:

- **V2.** Here, let $\beta = (C_1, C_2, \ldots, C_p)^\top$ and note that this vector does not grow with $n$. Obtain estimates of $\beta$ (e.g., $\hat{C}_j$'s) by fitting the full model subject to the null hypothesis (i.e., reduced model). For example, the least squares estimates are given by $\hat{C}_j = \bar{Y}_{j..} - \bar{Y}...$. Define $\hat{Y}_{ikl}^* = Y_{ijkl} - \hat{C}_j$. Then, $ALJ = \sum_{t=1}^{np} J_t$, where $J_t$ is the JT test statistic for the $t$th $R_{i(k)}/S_k$ "block" combination, will be our new test statistic. For the general alternative, the test statistic is just Friedman’s test based on the $\hat{Y}_{ikl}^*$’s which follow a RCBD.

- **V3.** This case is similar to V2 except $\beta = (R_1, R_2, \ldots, R_p)^\top$.

For convenience, since the subscript $j$ is a function of $i, k,$ and $l$, suppose we rewrite the original model (V2, for example) as

\[
Y_{ikl} = \mu + R_{i(k)} + \lambda_{ikl}^\top C + S_k + T_l + \varepsilon_{ikl} \quad \text{iff} \quad Y_{ikl} - \lambda_{ikl}^\top C = \mu + (R_{i(k)} + S_k) + T_l + \varepsilon_{ikl},
\]

where $C = (C_1, C_2, \ldots, C_p)^\top$ and $\lambda_{ikl} = (0, \ldots, 1, \ldots, 0)^\top$ is a $p \times 1$ vector of zeroes except for a 1 in the $j$th position. The (true) aligned observations are then denoted as $Y_{ikl}^* = Y_{ikl} - \lambda_{ikl}^\top C$ and follow a RCBD with the $np$ combinations of $R_{i(k)} + S_k$ denoting the blocks.

Assuming $\beta$ (= $C$ here) is known, it follows (by the Central Limit Theorem) that the aligned JT statistic $ALJ = \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I(Y_{ikl_1}^* < Y_{ikl_2}^*)$ satisfies $\frac{ALJ - E(ALJ)}{\sqrt{Var(ALJ)}} \xrightarrow{d} N(0, 1)$ where $E(ALJ) = \frac{np}{t} [p^2 - \sum_{l_1 = 1}^{p} 1^2] = \frac{np}{t} (p^2 - p)$ and $Var(ALJ) = np \cdot \frac{p^2(2p+3) - \sum_{l_1 = 1}^{p} 1^2(2\cdot1+3)}{t^2} = np \cdot \frac{p^2(2p+3) - 5p}{t^2}$. For V2/V3, Friedman’s statistic is given by
\[ A L \chi^2_r = \frac{12np}{p(p+1)} \sum_{l=1}^{p} (\bar{R}_{-l} - \frac{p+1}{2})^2 \] where \( \bar{R}_{-l} \) denotes the average of ranks for treatment \( l \). Indeed, this follows an asymptotic \( \chi^2 \) distribution with \( p - 1 \) degrees of freedom.

For the more practical case estimates of the coefficients are needed for alignment. For example, consider the following calculations for \( V_2 \):

\[
\bar{Y}_{jk} = \frac{1}{p} \sum_{i=1}^{p} (\mu + R_{i(k)} + C_j + S_k + T_l + \varepsilon_{ijkl})
\]
\[
= \mu + C_j + S_k + \bar{\varepsilon}_{jk},
\]

\[
\bar{Y}_{..k} = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\mu + R_{i(k)} + C_j + T_l + \varepsilon_{ijkl})
\]
\[
= \mu + S_k + \bar{\varepsilon}..k, \text{ and}
\]

\[
\bar{Y}_{jk} - \bar{Y}_{..k} = (\mu + C_j + S_k + \bar{\varepsilon}_{jk}) - (\mu + S_k + \bar{\varepsilon}..k)
\]
\[
= C_j + \bar{\varepsilon}_{jk} - \bar{\varepsilon}..k.
\]
\[
= C_j + \bar{\varepsilon}_{jk} - \bar{\varepsilon}..k - \theta + \theta
\]
\[
= (\theta + C_j) + (\bar{\varepsilon}_{jk} - \bar{\varepsilon}..k - \theta)
\]
\[
= g_j + e_{jk},
\]

where \( \theta \) is the location parameter for \( \bar{\varepsilon}_{jk} - \bar{\varepsilon}..k \), \( g_j = \theta + C_j \), and \( e_{jk} = \bar{\varepsilon}_{jk} - \bar{\varepsilon}..k - \theta \).

Thus, the model corresponding to \( \{\bar{Y}_{jk} - \bar{Y}_{..k}\} \) is a multivariate location model. The \( \bar{Y}_{jk} - \bar{Y}_{..k} \) may be arranged and expressed as a \( n \times p \) matrix \( \mathbb{C} \) where:
\[
C = \begin{pmatrix}
\bar{Y}_{11} - \bar{Y}_{11} & \bar{Y}_{21} - \bar{Y}_{11} & \cdots & \bar{Y}_{p1} - \bar{Y}_{11} \\
\bar{Y}_{12} - \bar{Y}_{12} & \bar{Y}_{22} - \bar{Y}_{12} & \cdots & \bar{Y}_{p2} - \bar{Y}_{12} \\
\vdots & \vdots & \ddots & \vdots \\
\bar{Y}_{1n} - \bar{Y}_{1n} & \bar{Y}_{2n} - \bar{Y}_{1n} & \cdots & \bar{Y}_{pn} - \bar{Y}_{1n}
\end{pmatrix}
\]

To estimate \( C \) then, let \( m_j \) be the column-wise means, medians, or Hodges-Lehmann (Hodges and Lehmann (2012)) estimates of \( C \) (e.g. \( m_j = \text{median}\{\bar{Y}_{jk} - \bar{Y}_{-k}\} \)). However, in general, \( m_j \) does not estimate \( C_j \) and \( \sum_{j=1}^{p} m_j \neq 0 \) for the median and Hodges-Lehmann estimates. Therefore, we let

\[
\hat{C}_j = m_j - \bar{m}
\]

where \( \bar{m} = \frac{1}{p} \sum_{j=1}^{p} m_j \), so that \( \sum_{j=1}^{p} \hat{C}_j = 0 \) is satisfied and \( \hat{C}_j \) estimates \( C_j \). For our asymptotic theory, we will require \( \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \) (\( \sqrt{n}(\hat{C} - C) = O_p(1) \) here) for the estimates used to align the observations. Proof 7.2 in the appendix shows that the least squares estimates satisfy \( \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \). For the median and Hodges-Lehmann estimates, we refer to Theorem 6.1.2 in Hettmansperger and McKean (2010) to show that \( \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \).

We now show that the asymptotic distribution property holds when the alignment coefficients are estimated. More specifically, for the ordered alternative, we need to show that for the aligned JT statistic and its estimate \( \hat{ALJ} \), the difference between the standardized statistics satisfies

\[
\frac{\hat{ALJ} - E(\hat{ALJ})}{\sqrt{Var(\hat{ALJ})}} - \frac{ALJ - E(ALJ)}{\sqrt{Var(ALJ)}} = \frac{1}{\sqrt{Var(ALJ)}} (\hat{ALJ} - ALJ) = o_p(1).
\]
It is only necessary to show that $\frac{1}{\sqrt{n}}(\tilde{ALJ} - ALJ) = o_p(1)$, since $\text{Var}(ALJ) = O(n)$.

Similarly, for the general alternative, we need to show that $\tilde{ALX}_r^2 - ALX_r^2 = o_p(1)$.

For the ordered alternative,

$$\frac{1}{\sqrt{n}}(ALJ - ALJ) = \frac{1}{\sqrt{n}}\left[\sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I(Y^*_{ikl_1} < Y^*_{ikl_2}) - \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I(Y^*_{ikl_1} < Y^*_{ikl_2})\right]$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (I[Y_{ikl_1} - \lambda^\top_{ikl_1} \hat{C} < Y_{ikl_2} - \lambda^\top_{ikl_2} \hat{C}] - I[Y_{ikl_1} - \lambda^\top_{ikl_1} C < Y_{ikl_2} - \lambda^\top_{ikl_2} C])$$.

The common terms $\mu, R_{i(k)}, S_k$ between $Y_{ikl_1}$ and $Y_{ikl_2}$ cancel out on both sides of the inequality. Under $H_0$, $T_1 = T_2 = \cdots = T_p = 0$, so $\frac{1}{\sqrt{n}}(ALJ - ALJ)$ can be written as:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \sqrt{n}(\hat{C} - C) - I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0)).$$

Next, we define the function $S_n(\Delta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta]$. Then, our conjecture is that $\frac{1}{\sqrt{n}}(ALJ - ALJ) = S_n(\sqrt{n}(\hat{C} - C)) - S_n(0) = o_p(1)$.

For the general alternative, define a function $W_n(x) = \frac{12np}{p(p+1)} \sum_{i=1}^{p} (x_i - \frac{p+1}{2})^2$ so that $ALX_r^2 = W_n(\bar{R})$, where $\bar{R} = (\bar{R}_1, \bar{R}_2, \ldots, \bar{R}_p)\top$, and the test statistic obtained with estimates of the alignment coefficients is $\tilde{ALX}_r^2 = W_n(\hat{R})$. Then, by the multivariate Taylor theorem, $W_n(\hat{R}) = W_n(\bar{R}) + \nabla^\top W_n(g)(\hat{R} - \bar{R})$ where

$$\nabla W_n(x) = \frac{24np}{p(p+1)}(x - \frac{p+1}{2}1_p) \quad \text{and} \quad g = \lambda \hat{R} + (1 - \lambda) \bar{R} \quad \text{for some} \quad \lambda \in (0, 1).$$
Therefore,

\[
\overline{AL\chi_r^2} - \overline{AL\chi_r^2} = W_n(\hat{R}) - W_n(\bar{R})
\]

\[
= \nabla^\top W_n(g)(\hat{R} - \bar{R})
\]

\[
= \frac{24np}{p(p+1)}(g - \frac{p+1}{2}1_p)^\top(\hat{R} - \bar{R})
\]

\[
= \frac{24}{p(p+1)}[\sqrt{np}(g - \frac{p+1}{2}1_p)]^\top[\sqrt{np}(\hat{R} - \bar{R})].
\]

From the theory in Example 29, page 388 of Lehmann and D’Abrera (1975), we know that

\[
\sqrt{np}(\bar{R} - (\frac{p+1}{2}1_p)) \xrightarrow{d} N_p\left(0_p, \begin{pmatrix}
\frac{p^2-1}{12} & \ldots & \frac{-(p+1)}{12} \\
\frac{-(p+1)}{12} & \ddots & \\
\frac{-(p+1)}{12} & \ldots & \frac{p^2-1}{12}
\end{pmatrix}_{p \times p}\right).
\]

So, \(\sqrt{np}(\hat{R} - (\frac{p+1}{2}1_p)) = O_p(1)\). Next, consider the following:

\[
\sqrt{np}(g - \frac{p+1}{2}1) = \sqrt{np}(\lambda \hat{R} + (1 - \lambda)\bar{R}) - \lambda \frac{p+1}{2}1 - (1 - \lambda) \frac{p+1}{2}1
\]

\[
= \lambda \sqrt{np}(\hat{R} - \frac{p+1}{2}1) + (1 - \lambda) \sqrt{np}(\bar{R} - \frac{p+1}{2}1)
\]

\[
= \lambda \sqrt{np}(\hat{R} - \bar{R}) + \lambda \sqrt{np}(\bar{R} - \hat{R} - \frac{p+1}{2}1) + (1 - \lambda) \sqrt{np}(\bar{R} - \frac{p+1}{2}1)
\]

\[
= \lambda \sqrt{np}(\hat{R} - \bar{R}) + \sqrt{np}(\bar{R} - \frac{p+1}{2}).
\]
So, to show that \( \sqrt{n p} \hat{AL}^2 - AL \chi^2 = o_p(1) \) it suffices to show that \( \sqrt{n p} (\hat{R} - R) = o_p(1) \) since \( \sqrt{n p} (\hat{R} - (\text{proj}_1 \mathbf{1}_p)) = O_p(1) \) and \( \lambda \) is bounded by 1. Moreover, since \( \sqrt{n p} (\hat{R} - R) = o_p(1) \) if and only if each component converges, it will suffice to show that \( \sqrt{n p} (\hat{R}_{-1} - R_{-1}) = o_p(1) \). To that end, consider the following:

\[
\sqrt{n p} (\hat{R}_{-1} - R_{-1}) = \sqrt{n p} \left( \frac{1}{np} \sum_{k=1}^n \sum_{i=1}^p \hat{R}_{ik1} - \frac{1}{np} \sum_{k=1}^n \sum_{i=1}^p R_{ik1} \right)
\]

\[
= \frac{1}{\sqrt{np}} \sum_{k=1}^n \sum_{i=1}^p \left( \sum_{l=1}^p I(\hat{Y}^*_l < \hat{Y}^*_{ik1}) - \sum_{l=1}^p I(\hat{Y}^*_l < Y^*_{ik1}) \right)
\]

Define \( S_n(\Delta) = \frac{1}{\sqrt{np}} \sum_{k=1}^n \sum_{i=1}^{p-1} \sum_{l=2}^p I(\varepsilon_{ikl} < \varepsilon_{ik1} - (\lambda_{ikl} - \lambda_{ik1})^\top \frac{1}{\sqrt{n}} \sqrt{n}(\hat{C} - C)) \). Then, our conjecture is that \( \sqrt{n p} (\hat{R}_{-1} - R_{-1}) = S_n(\sqrt{n}(\hat{C} - C)) - S_n(0) = o_p(1) \).

For both expressions of \( S_n(\sqrt{n}(\hat{C} - C)) - S_n(0) \) (i.e. for \( \frac{1}{\sqrt{n}} (AL\hat{J} - ALJ) \) and \( \sqrt{n p}(\hat{R}_{-1} - R_{-1}) \)), our conjecture requires the following three conditions to be shown:

1. \( S_n(\Delta) - S_n(0) = o_p(1) \) for a fixed \( \Delta \),

2. \( \sup_{||\Delta|| \leq C} |S_n(\Delta) - S_n(0)| = o_p(1) \) for a fixed \( C > 0 \),
3. \(|\sqrt{n}(\hat{C} - C)|| = O_p(1)|.

These conditions imply that \(S_n(\sqrt{n}(\hat{C} - C)) - S_n(0) = o_p(1)|. Choose \(\varepsilon > 0, \delta > 0\).

Since \(\sqrt{n}(\hat{C} - C) = O_p(1)|, \exists B_\delta such that \(P(||\sqrt{n}(\hat{C} - C)|| > B_\delta) < \frac{\delta}{2}\) for \(n > N_1(\delta)|.

Next, let

\[
A_n = \{\omega : |S_n(\sqrt{n}(\hat{C} - C)) - S_n(0)| > \varepsilon\},
\]

\[
B_n = \{\omega : \sup_{||\Delta|| \leq B_\delta} |S_n(\Delta) - S_n(0)| > \varepsilon\}, \text{ and}
\]

\[
C_n = \{\omega : ||\sqrt{n}(\hat{C} - C)|| > B_\delta\}
\]

and note that \(P(B_n) < \frac{\delta}{2}\) if \(n > N_2(\varepsilon, \delta)|.

\[
P(A_n) = P(A_n \cap C_n^c) + P(A_n \cap C_n)
\]

\[
\leq P(B_n) + P(C_n) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta \text{ if } n > \max(N_1, N_2).\]

Proofs 7.3 and 7.4 show that the conditions are valid for the aligned Jonckheere-Terpstra test and the aligned Friedman test, respectively.

In what follows, to distinguish between the notations for different alignment estimates, we denote the aligned JT statistic as \(ALJ_{\text{mean}}, ALJ_{\text{median}}, \text{ and } ALJ_{\text{HL}}\) when aligning with mean, median, or Hodges-Lehmann estimates, respectively. The aligned Friedman statistic is denoted as \(AL\chi^2_{r\text{mean}}, AL\chi^2_{r\text{median}}, AL\chi^2_{r\text{HL}}\) for the three different estimates, respectively.
Chapter 4

Data Application Examples

4.1 Application to Apple Sales Data

For our example of applying the proposed test class methods, we begin with the apple sales dataset given in the paper by Hoofnagle (1965). The author cites that the Latin square design and variations, along with factorial designs, are experimental designs that have been adopted most readily to problem solving in the field of promotion and advertising. In the original analysis by the author, an $F$-test used to determine the difference in sales between color ranges was highly significant. More specifically, the three color ranges of apples serve as the levels of the treatment $T$, with treatment A being highly colored, treatment B partly red, and treatment C a mixture between treatments A and B. We will test if highly colored apples yield higher sales through the ordered effects alternative, $(H_A : T_B \leq T_C \leq T_A$ with at least one strict inequality), as well as the general alternative $(H_A : \text{At least one } T_i \neq 0)$. The data for this experiment is given in Table 4.1.
Table 4.1: Apple Sales Data

<table>
<thead>
<tr>
<th>Row (within Square)</th>
<th>Column 1 C</th>
<th>Column 2 C</th>
<th>Column 3 C</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_{1(1)}</td>
<td>A, 779</td>
<td>B, 496</td>
<td>C, 424</td>
</tr>
<tr>
<td>R_{2(1)}</td>
<td>B, 312</td>
<td>C, 314</td>
<td>A, 238</td>
</tr>
<tr>
<td>R_{3(1)}</td>
<td>C, 803</td>
<td>A, 599</td>
<td>B, 314</td>
</tr>
<tr>
<td>R_{1(2)}</td>
<td>A, 703</td>
<td>C, 416</td>
<td>B, 319</td>
</tr>
<tr>
<td>R_{2(2)}</td>
<td>B, 376</td>
<td>A, 458</td>
<td>C, 276</td>
</tr>
<tr>
<td>R_{3(2)}</td>
<td>C, 623</td>
<td>B, 397</td>
<td>A, 556</td>
</tr>
<tr>
<td>R_{1(3)}</td>
<td>A, 557</td>
<td>B, 382</td>
<td>C, 346</td>
</tr>
<tr>
<td>R_{2(3)}</td>
<td>B, 313</td>
<td>C, 489</td>
<td>A, 396</td>
</tr>
<tr>
<td>R_{3(3)}</td>
<td>C, 170</td>
<td>A, 211</td>
<td>B, 85</td>
</tr>
</tbody>
</table>

The model follows a V2 design

\[
Y_{ijkl} = \mu + R_{i(k)} + C_j + S_k + T_l + \epsilon_{ijkl}
\]

where \(\mu\) denotes an overall location parameter, \(R_{i(k)}\) is the \(i^{th}\) store row effect nested within the \(k^{th}\) square replication, \(C_j\) is the experiment period column effect, \(S_k\) is the \(k^{th}\) replication square effect (may be interpreted as the region that contains the three stores), \(T_l\) is the color range treatment effect, and \(\epsilon_{ijkl}\) the random error term. All indices are elements of \{1, 2, 3\} and for the treatment effect 1 = B, 2 = C, and 3 = A.
4.1.1 Least-Squares Method Results

Before we discuss the results of our proposed test classes, we will present results from traditional least-squares methods for comparison. The least-squares estimates of the treatment effects are

\[ \hat{T} = (\bar{Y}_{-1} - \bar{Y}, \bar{Y}_{-2} - \bar{Y}, \bar{Y}_{-3} - \bar{Y})^\top = (-87.7778, 8.5556, 79.2222)^\top. \]

For the general alternative,

\[ K^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \text{ and } K^\top \hat{T} = \begin{pmatrix} -87.7778 \\ 8.5556 \end{pmatrix}. \]

The covariance matrix of \( K^\top \hat{T} \) is \( \hat{\Sigma} \) where

\[ \hat{\Sigma} = \hat{\sigma}^2 \begin{pmatrix} \frac{p-1}{p^2} & \frac{-1}{p^2} \\ \frac{-1}{p^2} & \frac{p-1}{p^2} \end{pmatrix} = 5795.079 \cdot \begin{pmatrix} \frac{2}{9} & \frac{-1}{9} \\ \frac{-1}{9} & \frac{2}{9} \end{pmatrix}. \]

Therefore, the asymptotic \( \chi^2 \) statistic is \( n(K^\top \hat{T})^\top \hat{\Sigma}^{-1}(K \hat{T}) = 21.1694 \) with a \( p \)-value \( P(\chi^2_2 > 21.1694) < 0.0001 \). The exact \( F \) statistic is \( \frac{1}{p-1} \cdot n(K^\top \hat{T})^\top \hat{\Sigma}^{-1}(K \hat{T}) = 10.5847 \) with a \( p \)-value \( P(F_{2,14} > 10.5847) = 0.0016 \).

For the ordered alternative, the Abelson-Tukey contrast is \((-\sqrt{\frac{2}{3}}, 0, \sqrt{\frac{2}{3}})^\top\), so the contrast of treatments is \( K^\top \hat{T} = 136.3549 \). Therefore, the standardized test statistic is

\[ \frac{K^\top \hat{T}}{\sqrt{K^\top \Sigma K/n}} = 4.583. \]

An exact \( p \)-value from a \( t \)-distribution is \( P(t_{14} > 4.583) = 0.0002 \), while an asymptotic \( p \)-value from a standard normal distribution is \( P(Z > 4.583) < 0.0001 \). A summary of the least-squares methods results are given in Table 4.2.

<table>
<thead>
<tr>
<th>Table 4.2: Least-squares Methods Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Alternative</td>
</tr>
<tr>
<td>Asymptotic ( p )-value</td>
</tr>
<tr>
<td>Exact ( p )-value</td>
</tr>
</tbody>
</table>
4.1.2 Aggregated Test Class Results

A sketch of the randomized complete block design as a result of aggregating the observations is given in Table 4.3.

Table 4.3: RCBD of Aggregated Observations

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Blocks (squares)</th>
<th>B</th>
<th>C</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>(\bar{Y}_{-11} = 374)</td>
<td>(\bar{Y}_{-12} = 513.6667)</td>
<td>(\bar{Y}_{-13} = 538.6667)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(\bar{Y}_{-21} = 364)</td>
<td>(\bar{Y}_{-22} = 438.3333)</td>
<td>(\bar{Y}_{-23} = 572.3333)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(\bar{Y}_{-31} = 260)</td>
<td>(\bar{Y}_{-32} = 335.0000)</td>
<td>(\bar{Y}_{-33} = 388.0000)</td>
</tr>
</tbody>
</table>

In Table 4.4, the ranks of the aggregated observations are given in order to calculate the test statistics.

Table 4.4: Ranks of Aggregated Observations

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Blocks (squares)</th>
<th>B</th>
<th>C</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>(R_{11} = 1)</td>
<td>(R_{12} = 2)</td>
<td>(R_{13} = 3)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(R_{21} = 1)</td>
<td>(R_{22} = 2)</td>
<td>(R_{23} = 3)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(R_{31} = 1)</td>
<td>(R_{32} = 2)</td>
<td>(R_{33} = 3)</td>
</tr>
<tr>
<td>Average ranks</td>
<td></td>
<td>(\bar{R}_{-1} = 1)</td>
<td>(\bar{R}_{-2} = 2)</td>
<td>(\bar{R}_{-3} = 3)</td>
</tr>
<tr>
<td>Rank sums</td>
<td></td>
<td>(R_1 = 3)</td>
<td>(R_2 = 6)</td>
<td>(R_3 = 9)</td>
</tr>
</tbody>
</table>

For the general alternative, Friedman’s test returned a test statistic of \(\chi^2_r = \frac{12}{np(p+1)} \cdot \frac{1}{n} \sum_{l=1}^{p} (R_l - \frac{n(p+1)}{2})^2 = 6\) with a \(p\)-value of \(P(\chi^2_r > 6) = 0.0498\). The exact \(p\)-value is calculated with the \texttt{pFrd} R function from the \texttt{NSM3} package specifying for an exact method and is equal to 0.0278. Under \(\alpha = 0.05\), we reject \(H_0\) and conclude that
there is significant evidence that at least one of the treatments are different from the others.

For the ordered alternative, we perform Page’s test treating the aggregated observations as an RCBD as well as our proposed aggregated JT test. That is, we are testing the hypotheses: $H_0 : T_A = T_B = T_C$ vs. $H_A : T_B < T_C < T_A$. Page’s statistic is $L = \sum_{l=1}^p \sum_{k=1}^n R_{kj} = 42$ with an exact $p$-value of 0.0046 and asymptotic $p$-value of $P(Z > \frac{42 - 36}{\sqrt{6}}) = 0.0072$. Since the $p$-value is less than 0.05, we reject the null hypothesis and conclude that there is an increasing difference between the three treatments. For the aggregated JT test, the exact test statistic is $AGJ = 9$ with an exact $p$-value of 0.0046 and asymptotic $p$-value of 0.0033 (see Table 3.1). Therefore, the null hypothesis is rejected which concludes that there is significant evidence of an increasing difference between the three treatments. A summary of results from the aggregated test class is given in Table 4.5.

<table>
<thead>
<tr>
<th>Table 4.5: Test Results for Aggregated Test Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Asymptotic $p$-value</td>
</tr>
<tr>
<td>Exact $p$-value</td>
</tr>
</tbody>
</table>

4.1.3 Aligned Test Results

We will start this section by outlining how the observations are aligned for this example. Since the apple sales data follows a V2 design (rows are nested within squares), the estimates of the column effects are to be subtracted (aligned) from the original observations. The matrix $C$, which contains $Y_{jk} - \bar{Y}_{.,k}$, is given below:
Using $C$, the mean estimates of the column effect are

$$C = \begin{pmatrix} \bar{Y}_{11} - \bar{Y}_{-1} = 155.8889 & \bar{Y}_{21} - \bar{Y}_{-1} = -5.7778 & \bar{Y}_{31} - \bar{Y}_{-1} = -150.1111 \\ \bar{Y}_{12} - \bar{Y}_{-2} = 109.1111 & \bar{Y}_{22} - \bar{Y}_{-2} = -34.5556 & \bar{Y}_{32} - \bar{Y}_{-2} = -74.5556 \\ \bar{Y}_{13} - \bar{Y}_{-3} = 19.0000 & \bar{Y}_{23} - \bar{Y}_{-3} = 33.0000 & \bar{Y}_{33} - \bar{Y}_{-3} = -52.0000 \end{pmatrix}.$$ 

With these estimates, the observations aligned with means, medians, and Hodges-Lehmann estimates are given in Tables 4.6, 4.7, and 4.8, respectively. These tables also include the row ranks, JT statistics for each row, the average ranks, and rank sums needed to calculate our test statistics for testing $H_A : T_B \leq T_C \leq T_A$. 

Using $C$, the mean estimates of the column effect are

$$\hat{C}^T = (m_1 - \bar{m}, m_2 - \bar{m}, m_3 - \bar{m}) = (\frac{1}{n} \sum_{k=1}^{n} (\bar{Y}_{1k} - \bar{Y}_{-k}) - \bar{m}, \frac{1}{n} \sum_{k=1}^{n} (\bar{Y}_{2k} - \bar{Y}_{-k}) - \bar{m}, \frac{1}{n} \sum_{k=1}^{n} (\bar{Y}_{3k} - \bar{Y}_{-k}) - \bar{m}) = (94.6667, -2.4444, -92.2222).$$ 

The median estimates of the column effect are

$$\hat{C}_{\text{median}}^T = (m_1 - \bar{m}, m_2 - \bar{m}, m_3 - \bar{m}) = (\text{median}_{1 \leq k \leq n} \{\bar{Y}_{1k} - \bar{Y}_{-k}\} - \bar{m}, \text{median}_{1 \leq k \leq n} \{\bar{Y}_{2k} - \bar{Y}_{-k}\} - \bar{m}, \text{median}_{1 \leq k \leq n} \{\bar{Y}_{3k} - \bar{Y}_{-k}\} - \bar{m}) = (99.5185, -15.3704, -84.1481).$$ 

The Hodges-Lehmann estimates of the column effect are

$$\hat{C}_{\text{HL}}^T = (m_1 - \bar{m}, m_2 - \bar{m}, m_3 - \bar{m}) = (\text{HL}_{1 \leq k \leq n} \{\bar{Y}_{1k} - \bar{Y}_{-k}\} - \bar{m}, \text{HL}_{1 \leq k \leq n} \{\bar{Y}_{2k} - \bar{Y}_{-k}\} - \bar{m}, \text{HL}_{1 \leq k \leq n} \{\bar{Y}_{3k} - \bar{Y}_{-k}\} - \bar{m}) = (95.8796, -5.6759, -90.2037).$$
Table 4.6: Observations Aligned with Means

<table>
<thead>
<tr>
<th>Column</th>
<th>Row(Square)</th>
<th>$C_1$</th>
<th>Rank</th>
<th>$C_2$</th>
<th>Rank</th>
<th>$C_3$</th>
<th>Rank</th>
<th>$J_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_{1(1)}$</td>
<td>A, 684.3333,</td>
<td>3</td>
<td>B, 498.4444,</td>
<td>1</td>
<td>C, 516.2222,</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$R_{2(1)}$</td>
<td>B, 217.3333,</td>
<td>1</td>
<td>C, 316.4444,</td>
<td>2</td>
<td>A, 330.2222,</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$R_{3(1)}$</td>
<td>C, 708.3333,</td>
<td>3</td>
<td>A, 601.4444,</td>
<td>2</td>
<td>B, 406.2222,</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$R_{1(2)}$</td>
<td>A, 608.3333,</td>
<td>3</td>
<td>C, 418.4444,</td>
<td>2</td>
<td>B, 411.2222,</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$R_{2(2)}$</td>
<td>B, 281.3333,</td>
<td>1</td>
<td>A, 460.4444,</td>
<td>3</td>
<td>C, 368.2222,</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$R_{3(2)}$</td>
<td>C, 528.3333,</td>
<td>2</td>
<td>B, 399.4444,</td>
<td>1</td>
<td>A, 648.2222,</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$R_{1(3)}$</td>
<td>A, 462.3333,</td>
<td>3</td>
<td>B, 384.4444,</td>
<td>1</td>
<td>C, 438.2222,</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$R_{2(3)}$</td>
<td>B, 218.3333,</td>
<td>1</td>
<td>C, 491.4444,</td>
<td>3</td>
<td>A, 488.2222,</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$R_{3(3)}$</td>
<td>C, 75.3333,</td>
<td>1</td>
<td>A, 213.4444,</td>
<td>3</td>
<td>B, 177.2222,</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Average ranks: B, 1.1111, C, 2.1111, A, 2.7778

Rank sums: B, 10, C, 19, A, 25
Table 4.7: Observations Aligned with Medians

<table>
<thead>
<tr>
<th>Row(Square)</th>
<th>$C_1$</th>
<th>Rank</th>
<th>$C_2$</th>
<th>Rank</th>
<th>$C_3$</th>
<th>Rank</th>
<th>$J_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{1(1)}$</td>
<td>A, 679.4815,</td>
<td>3</td>
<td>B, 511.3704,</td>
<td>2</td>
<td>C, 508.1481,</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$R_{2(1)}$</td>
<td>B, 212.4815,</td>
<td>1</td>
<td>C, 329.3704,</td>
<td>3</td>
<td>A, 322.1481,</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$R_{3(1)}$</td>
<td>C, 703.4815,</td>
<td>3</td>
<td>A, 614.3704,</td>
<td>2</td>
<td>B, 398.1481,</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$R_{1(2)}$</td>
<td>A, 603.4815,</td>
<td>3</td>
<td>C, 431.3704,</td>
<td>2</td>
<td>B, 403.1481,</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>$R_{2(2)}$</td>
<td>B, 276.4815,</td>
<td>1</td>
<td>A, 473.3704,</td>
<td>3</td>
<td>C, 360.1481,</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$R_{3(2)}$</td>
<td>C, 523.4815,</td>
<td>2</td>
<td>B, 412.3704,</td>
<td>1</td>
<td>A, 640.1481,</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$R_{1(3)}$</td>
<td>A, 457.4815,</td>
<td>3</td>
<td>B, 397.3704,</td>
<td>1</td>
<td>C, 430.1481,</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$R_{2(3)}$</td>
<td>B, 213.4815,</td>
<td>1</td>
<td>C, 504.3704,</td>
<td>3</td>
<td>A, 480.1481,</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$R_{3(3)}$</td>
<td>C, 70.4815,</td>
<td>1</td>
<td>A, 226.3704,</td>
<td>3</td>
<td>B, 169.1481,</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Average ranks</strong></td>
<td>B, 1.2222</td>
<td>C, 2.1111</td>
<td>A, 2.6667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rank sums</strong></td>
<td>B, 11</td>
<td>C, 19</td>
<td>A, 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.8: Observations Aligned with Hodges-Lehmann

<table>
<thead>
<tr>
<th>Column</th>
<th>Row(Square)</th>
<th>$C_1$</th>
<th>Rank</th>
<th>$C_2$</th>
<th>Rank</th>
<th>$C_3$</th>
<th>Rank</th>
<th>$J_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{1(1)}$</td>
<td>$A, 683.1204, 3$</td>
<td>$B, 501.6759, 1$</td>
<td>$C, 514.2037, 2$</td>
<td>$3$</td>
<td>$3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{2(1)}$</td>
<td>$B, 216.1204, 1$</td>
<td>$C, 319.6759, 2$</td>
<td>$A, 328.2037, 3$</td>
<td>$3$</td>
<td>$3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{3(1)}$</td>
<td>$C, 707.1204, 3$</td>
<td>$A, 604.6759, 2$</td>
<td>$B, 404.2037, 1$</td>
<td>$2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{1(2)}$</td>
<td>$A, 607.1204, 3$</td>
<td>$C, 421.6759, 2$</td>
<td>$B, 409.2037, 1$</td>
<td>$3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{2(2)}$</td>
<td>$B, 280.1204, 1$</td>
<td>$A, 463.6759, 3$</td>
<td>$C, 366.2037, 2$</td>
<td>$3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{3(2)}$</td>
<td>$C, 527.1204, 2$</td>
<td>$B, 402.6759, 1$</td>
<td>$A, 646.2037, 3$</td>
<td>$3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{1(3)}$</td>
<td>$A, 461.1204, 3$</td>
<td>$B, 387.6759, 1$</td>
<td>$C, 436.2037, 2$</td>
<td>$3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{2(3)}$</td>
<td>$B, 217.1204, 1$</td>
<td>$C, 494.6759, 3$</td>
<td>$A, 486.2037, 2$</td>
<td>$2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{3(3)}$</td>
<td>$C, 74.1204, 1$</td>
<td>$A, 216.6759, 3$</td>
<td>$B, 175.2037, 2$</td>
<td>$2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average Ranks B, 1.1111 C, 2.1111 A, 2.7778

Rank sums B, 10 C, 19 A, 25

For the general alternative, the asymptotic Friedman’s test with observations aligned by mean estimates gave a test statistic of $AL\chi^2_{\text{mean}} = 12.6667$ with a $p$-value of 0.0018. For observations aligned by median estimates, the $AL\chi^2_{\text{median}}$ test statistic is 9.5556 with a $p$-value of 0.0084. Lastly, when observations are aligned by Hodges-Lehmann estimates, $AL\chi^2_{\text{HL}} = 12.6667$ with a $p$-value of 0.0018. Therefore, for all three alignment methods, the null hypothesis for the general alternative is rejected, which indicates that there is evidence of a difference between the three treatments.

The ordered-alternative test aligning by mean estimates has a test statistic of $ALJ_{\text{mean}} = 24$ with an asymptotic $p$-value 0.0001. When aligning by median estimates, the test statistic is $ALJ_{\text{median}} = 22$ with an asymptotic $p$-value of 0.0015.
aligning by Hodges-Lehmann estimates, the test statistic is $AL_{HL} = 24$ with an asymptotic $p$-value 0.0001. From these results, the null hypothesis is rejected for all variations of the test, which indicates that there is an increasing difference between the three treatments. A summary ($p$-values) of results from the aligned test class is given in Table 4.9.

<table>
<thead>
<tr>
<th>Alignment Estimate</th>
<th>General Alternative</th>
<th>Ordered Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0018</td>
<td>0.0001</td>
</tr>
<tr>
<td>Median</td>
<td>0.0084</td>
<td>0.0015</td>
</tr>
<tr>
<td>Hodges-Lehmann</td>
<td>0.0018</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

### 4.2 Application to Crossover Designs

The paper by Tardif et al. (2005) proposes nonparametric tests for balanced crossover designs of $n$ replications, each formed by $m$ Latin squares of order $p$. Their tests are permutation tests based on the $n$ vectors of least-squares estimators of the parameter of interest. The exact and limiting distribution of their test statistics show that the tests (asymptotically) have the same power as the $F$-ratio test. Accordingly, in this section, we examine how our proposed methods may also be utilized with crossover designs. We will use the cardiovascular health data measuring the duration of exercise in seconds (after subtracting a baseline value) from Tudor and Koch (1994), which is a two-treatment, two-period, two sequence crossover design. We may view this design as a V2 replicated $2 \times 2$ Latin square where the treatment has two levels (AIR and CO), rows would correspond to patients nested within square, columns (the same for each
square) would essentially be the time period (e.g., 1st and 2nd), and squares are the
$n = 14$ patient dyads where we essentially matched the 14 AIR:CO patients to 14 of
the CO:AIR patients via baseline values.

The techniques we employ to match observations between patients is detailed in the
journal article by Sekhon (2008). Briefly, the matching method is based on propensity
scores or multivariate matching based on Mahalanobis distance. In our example, we
do a 1-to-1 matching without replacement and estimate ATT, the sample average
treatment effect for the treated (matched) observations. The resulting replicated
Latin square design is given in Table 4.10. For the ordered alternative, we test that
the CO treatment has shorter (adjusted) exercise duration than the AIR treatment
($H_A : T_{CO} < T_{AIR}$).
### Table 4.10: Cardiovascular Health Data as a Replicated Latin Square Design

<table>
<thead>
<tr>
<th>Dyad</th>
<th>Patient</th>
<th>Time Period</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Dyad 1</td>
<td>1</td>
<td>AIR 180</td>
<td>CO 219</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>CO -60</td>
<td>AIR 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>AIR -360</td>
<td>CO -360</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>CO 0</td>
<td>AIR 0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AIR -241</td>
<td>CO -105</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>CO -180</td>
<td>AIR -102</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>AIR 625</td>
<td>CO 385</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>CO -38</td>
<td>AIR 150</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>AIR -30</td>
<td>CO 10</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>CO -20</td>
<td>AIR -20</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>AIR 270</td>
<td>CO 270</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>CO 65</td>
<td>AIR -13</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>AIR -40</td>
<td>CO -60</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>CO 0</td>
<td>AIR 0</td>
</tr>
<tr>
<td>Dyad 8</td>
<td>8</td>
<td>AIR -120</td>
<td>CO -350</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>CO -30</td>
<td>AIR -30</td>
</tr>
<tr>
<td>Dyad 9</td>
<td>9</td>
<td>AIR 180</td>
<td>CO 70</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>CO 18</td>
<td>AIR 18</td>
</tr>
<tr>
<td>Dyad 10</td>
<td>10</td>
<td>AIR -55</td>
<td>CO -80</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>CO -80</td>
<td>AIR -44</td>
</tr>
<tr>
<td>Dyad 11</td>
<td>11</td>
<td>AIR 155</td>
<td>CO 80</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>CO 90</td>
<td>AIR 118</td>
</tr>
<tr>
<td>Dyad 12</td>
<td>12</td>
<td>AIR -5</td>
<td>CO -5</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>CO -50</td>
<td>AIR -50</td>
</tr>
<tr>
<td>Dyad 13</td>
<td>13</td>
<td>AIR -80</td>
<td>CO -105</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>CO 30</td>
<td>AIR 205</td>
</tr>
<tr>
<td>Dyad 14</td>
<td>14</td>
<td>AIR -20</td>
<td>CO -90</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>CO 45</td>
<td>AIR 30</td>
</tr>
</tbody>
</table>

### Table 4.11: Least-squares Methods Results for Crossover Design

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Ordered Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic p-value</td>
<td>0.0236</td>
</tr>
<tr>
<td>Exact p-value</td>
<td>0.0321</td>
</tr>
</tbody>
</table>
We present the results from our proposed tests in Tables 4.11, 4.12, and 4.13. From the least-squares methods and the aligned test class, the null hypotheses are rejected for both the general and ordered alternative tests at the 0.05 level. However, for the aggregated test class, the null hypotheses are not rejected. This discrepancy might be explained via inferior power of the aggregation method, due to the fewer number of blocks caused by aggregating the data. We shall explore this through simulations in the next chapter.
Chapter 5

Simulation Studies

In this chapter, we conduct a simulation study and compare our proposed tests in terms of power. Our user-defined \texttt{rLS.sim1} makes use of the \texttt{repLSD} function based on the \texttt{agricolae} package (de Mendiburu (2021)) to generate simulated Latin square designs, in particular, V2 designs. A vector for the effects \((S, R, C, T)\) must be declared by the user to generate the simulated dataset. More specifically, \(S\) has \(n - 1\) elements, \(R\) has \((p - 1)n\) elements, and \(C\) and \(T\) each have \(p - 1\) elements. In general, we perform 10,000 simulations with a design including 15 square replications and \(p = 3\) rows/columns/treatments under a V2 design where rows are nested within squares. The effects vector \((T)\) differs when simulating under the null or alternative hypothesis. On the other hand, the effects for \(S\), \(R\), and \(C\) were generated from a uniform distribution and remain constant over all simulations. That is,
In Sect. 5.1, a simulation under the null hypothesis is performed, while a simulation under the alternative hypothesis is performed in Sect. 5.2. The normal distribution is assumed in both of these sections. In Sects. 5.3 and 5.4, we consider simulations where the error distribution follows a contaminated normal distribution (i.e. $\varepsilon_{ijkl} \sim (1 - \gamma)N(\mu_1, \sigma^2_1) + \gamma N(\mu_2, \sigma^2_2)$). The methods of sampling from a contaminated normal distribution that we use in our simulation studies are discussed in Tukey (1959). In the journal article by Kafadar (2003), the author cites Tukey’s findings from several graphs that show the degradation in performance of traditional estimators for the Gaussian distribution under the contaminated normal model. Thus, our goal is to showcase this difference through our proposed test class methods under the contaminated normal model. For our simulations, with error terms generated from a contaminated normal distribution, we will fix $\mu_1 = \mu_2 = 0$ and $\sigma^2_1 = 1$ and vary $\gamma$ (the proportion of contamination) and $\sigma^2_2$ in different simulations. In all simulations, we use $\alpha = 0.05$. 

$$
(S, R, C) = \begin{pmatrix}
-3.7874, & 3.4633, & -3.9448, & -3.9462, & 0.7319, & 2.8116, \\
-0.6518, & 3.7489, & 4.5739, & -3.0371, & -3.4856, & -1.2871, \\
5.3874, & -0.8415, & -4.1503, & 4.1503, & -3.8956, & 3.8956, \\
1.5290, & -1.5290, & -3.9212, & 3.9212, & -0.7592, & 0.7592, \\
3.6112, & -3.6112, & 4.0457, & -4.0457, & 1.3733, & -1.3733, \\
-4.0352, & 4.0352, & -3.3330, & 3.3330, & 1.6412, & -1.6412, \\
-1.6573, & 1.6573, & -3.4285, & 3.4285, & 4.5566, & -4.5566, \\
3.9276, & -3.9276, & 4.9536, & -4.9536 & &
\end{pmatrix}.
$$
5.1 Simulation under $H_0$ with Normally-Distributed Error

Here, the error distribution is set to follow a standard normal distribution. The effects vector is set up according to the null hypothesis where $T_1 = T_2 = T_3 = 0$ and the remaining effects have been previously discussed. Since this is under $H_0$, we expect the proportion of times (out of 10,000) that the null hypothesis is rejected to be close to $\alpha = 0.05$. In table 5.1 and all subsequent tables in this section, we will abbreviate asymptotic tests as "asymp. ".

Table 5.1: Type I Error Rate Comparison

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Ordered Alternative</th>
<th>Type I Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.0507</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.0553</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>Aggregated Page Exact</td>
<td>0.0575</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>Aggregated JT Exact</td>
<td>0.0467</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>Aggregated Page Asymp.</td>
<td>0.0302</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>Aligned Mean JT</td>
<td>0.0494</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>Aligned Median JT</td>
<td>0.0496</td>
</tr>
</tbody>
</table>

Table 5.1 shows that most of our proposed tests are slightly conservative, but still have acceptable Type I error rates around 0.05. The exact aggregated JT test and
Friedman’s test with observations aligned with means had the lowest Type I error rates at 0.0302.

5.2 Simulation under $H_A$ with Normally-Distributed Error

We also perform the simulations under the alternative hypothesis. To illustrate the differences between the power of each test method, we run simulations under three different effects vectors for the alternative hypothesis with varying differences between treatment levels: \((T_1, T_2, T_3) = (-0.2, 0, 0.2)\) (we will call this 'small'), \((-0.3, 0, 0.3)\) (we will call this 'medium'), and \((-0.5, 0, 0.5)\) (we will call this 'large').

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ((F))</td>
<td>0.3665</td>
<td>Least-Squares Exact ((t))</td>
<td>0.5950</td>
</tr>
<tr>
<td>Least-Squares Asymp. ((\chi^2))</td>
<td>0.3827</td>
<td>Least-Squares Asymp. ((z))</td>
<td>0.6018</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.2941</td>
<td>Aggregated Page Exact</td>
<td>0.4425</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Exact</td>
<td>0.3730</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.2568</td>
<td>Aggregated Page Asymp.</td>
<td>0.4425</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Asymp.</td>
<td>0.4820</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.2746</td>
<td>Aligned Mean JT</td>
<td>0.4538</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.2732</td>
<td>Aligned Median JT</td>
<td>0.4528</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.2765</td>
<td>Aligned HL JT</td>
<td>0.4547</td>
</tr>
</tbody>
</table>

Table 5.2 presents the results under the "small" alternative hypothesis. For a given
alternative type, $H_0$ is correctly rejected at a similar rate for all tests. The tests for an ordered alternative correctly reject $H_0$ at a higher rate than the respective general alternative counterparts. The least-squares methods typically reject $H_0$ correctly at a higher proportion than the nonparametric methods. Indeed, this is as it should be since the data is normally distributed. We also note that there is very little difference between the three alignment methods.

Table 5.3: Power Comparison under Medium Alternative

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.7042</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.8816</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.7198</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.8855</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.5654</td>
<td>Aggregated Page Exact</td>
<td>0.7343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Exact</td>
<td>0.6652</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.5163</td>
<td>Aggregated Page Asymp.</td>
<td>0.7343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Asymp.</td>
<td>0.7619</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.5549</td>
<td>Aligned Mean JT</td>
<td>0.7480</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.5570</td>
<td>Aligned Median JT</td>
<td>0.7453</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.5580</td>
<td>Aligned HL JT</td>
<td>0.7510</td>
</tr>
</tbody>
</table>

When testing under the "medium" alternative (see Table 5.3), we see an increase in the simulated power due to the larger effect size. Once more, the tests for the ordered alternative still have higher simulated power than the general alternative, the least-squares methods are typically outperforming the nonparametric methods, and there are only minor differences between the different alignment procedures.
Table 5.4: Power Comparison under Large Alternative

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.9932</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.9990</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.9944</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.9993</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.9478</td>
<td>Aggregated Page Exact</td>
<td>0.9854</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.9297</td>
<td>Aggregated JT Exact</td>
<td>0.9739</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.9297</td>
<td>Aggregated Page Asymp.</td>
<td>0.9854</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.9539</td>
<td>Aligned Mean JT</td>
<td>0.9904</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.9538</td>
<td>Aligned Median JT</td>
<td>0.9898</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.9547</td>
<td>Aligned HL JT</td>
<td>0.9902</td>
</tr>
</tbody>
</table>

Lastly, consider the results under the 'large' alternative hypothesis given in Table 5.4. While the conclusions are generally consistent with those given in Tables 5.2 and 5.3, they are not as obvious since all power estimates exceed 0.90. Of course, these high powers are expected given that the effect size is large.

5.3 Simulation under $H_0$ with Contaminated-Normal Error

Regarding the methods under investigation, only those based on aggregating the data are truly distribution free under $H_0$. Thus, in this section, we perform simulations under the null hypothesis where the error terms are generated from a contaminated normal distribution. We consider two contaminated normal distributions: $(\gamma, \sigma^2_2) = (0.10, 9)$.
and \((\gamma, \sigma^2) = (0.25, 100)\). In what follows we refer to \((0.10, 9)\) as 'less' contaminated and \((0.25, 100)\) as 'more' contaminated. The results are given in Tables 5.5 and 5.6.

Table 5.5: Type I Error Rate Comparison with Less-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Type I Error Rate</th>
<th>Ordered Alternative</th>
<th>Type I Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ((F))</td>
<td>0.0497</td>
<td>Least-Squares Exact ((t))</td>
<td>0.0505</td>
</tr>
<tr>
<td>Least-Squares Asymp. ((\chi^2))</td>
<td>0.0551</td>
<td>Least-Squares Asymp. ((z))</td>
<td>0.0521</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.0588</td>
<td>Aggregated Page Exact</td>
<td>0.0388</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.0461</td>
<td>Aggregated JT Exact</td>
<td>0.0268</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.0479</td>
<td>Aggregated Page Asymp.</td>
<td>0.0388</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.0501</td>
<td>Aggregated JT Asymp.</td>
<td>0.0492</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.0483</td>
<td>Aligned Mean JT</td>
<td>0.0423</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aligned Median JT</td>
<td>0.0437</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aligned HL JT</td>
<td>0.0433</td>
</tr>
</tbody>
</table>

From these simulations, we observe that the results are not too different from the null-hypothesis simulation with no contamination. That is, the least-squares methods retain their Type I error rates around 0.05, while the nonparametric methods have Type I error rates close or slightly lower than 0.05. That being said, all error rates are elements of \((0.025, 0.075)\), which is the benchmark for an acceptable Type I error rate when \(\alpha = 0.05\) noted in Bradley (1978).
Table 5.6: Type I Error Rate Comparison with More-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Type I Error Rate</th>
<th>Ordered Alternative</th>
<th>Type I Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.0468</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.0541</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.0532</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.0558</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.0610</td>
<td>Aggregated Page Exact</td>
<td>0.0433</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.0477</td>
<td>Aggregated JT Exact</td>
<td>0.0301</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.0477</td>
<td>Aggregated Page Asymp.</td>
<td>0.0433</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.0376</td>
<td>Aligned Mean JT</td>
<td>0.0407</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.0401</td>
<td>Aligned Median JT</td>
<td>0.0425</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.0386</td>
<td>Aligned HL JT</td>
<td>0.0419</td>
</tr>
</tbody>
</table>

5.4 Simulation under $H_A$ with Contaminated-Normal Error

In this section, we simulate the power of our tests with the three effect vectors (small, medium, and large differences between treatment levels). Once more, for each simulation of the three effect vectors, we run two variations of the error term distribution from a contaminated normal distribution ($\gamma = 0.1, \sigma_2^2 = 9$ and $\gamma = 0.25, \sigma_2^2 = 100$). Since the data is generated from a contaminated normal distribution, we expect our nonparametric methods to eventually perform better than the least-squares methods when the error distribution sufficiently deviates from the normal distribution. The results are given in Tables 5.7 - 5.12.
Table 5.7: Power Comparison under Small Alternative with Less-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.2164</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.4063</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.2323</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.4142</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.2136</td>
<td>Aggregated Page Exact</td>
<td>0.3322</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Exact</td>
<td>0.2699</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.1817</td>
<td>Aggregated Page Asymp.</td>
<td>0.3322</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Asymp.</td>
<td>0.3690</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.2149</td>
<td>Aligned Mean JT</td>
<td>0.3751</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.2164</td>
<td>Aligned Median JT</td>
<td>0.3759</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.2177</td>
<td>Aligned HL JT</td>
<td>0.3777</td>
</tr>
</tbody>
</table>
Table 5.8: Power Comparison under Small Alternative with More-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.0558</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.1030</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.0633</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.1059</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.0813</td>
<td>Aggregated Page Exact</td>
<td>0.1156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Exact</td>
<td>0.0856</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.0652</td>
<td>Aggregated Page Asymp.</td>
<td>0.1156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Asymp.</td>
<td>0.1378</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.0915</td>
<td>Aligned Mean JT</td>
<td>0.1839</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.1007</td>
<td>Aligned Median JT</td>
<td>0.1959</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.0976</td>
<td>Aligned HL JT</td>
<td>0.1926</td>
</tr>
</tbody>
</table>
Table 5.9: Power Comparison under Medium Alternative with Less-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.4565</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.6786</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.4724</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.6857</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.4120</td>
<td>Aggregated Page Exact</td>
<td>0.5689</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Exact</td>
<td>0.4964</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.3688</td>
<td>Aggregated Page Asymp.</td>
<td>0.5689</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aggregated JT Asymp.</td>
<td>0.6031</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.4443</td>
<td>Aligned Mean JT</td>
<td>0.6504</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.4489</td>
<td>Aligned Median JT</td>
<td>0.6468</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.4487</td>
<td>Aligned HL JT</td>
<td>0.6521</td>
</tr>
</tbody>
</table>
Table 5.10: Power Comparison under Medium Alternative with More-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact ($F$)</td>
<td>0.0737</td>
<td>Least-Squares Exact ($t$)</td>
<td>0.1354</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.0807</td>
<td>Least-Squares Asymp. ($z$)</td>
<td>0.1391</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.1128</td>
<td>Aggregated Page Exact</td>
<td>0.1658</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.0949</td>
<td>Aggregated JT Exact</td>
<td>0.1276</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.0949</td>
<td>Aggregated Page Asymp.</td>
<td>0.1658</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.1682</td>
<td>Aligned Mean JT</td>
<td>0.3067</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.1812</td>
<td>Aligned Median JT</td>
<td>0.3261</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.1778</td>
<td>Aligned HL JT</td>
<td>0.3222</td>
</tr>
</tbody>
</table>
Table 5.11: Power Comparison under Large Alternative with Less-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact (F)</td>
<td>0.8866</td>
<td>Least-Squares Exact (t)</td>
<td>0.9662</td>
</tr>
<tr>
<td>Least-Squares Asymp. (χ²)</td>
<td>0.8948</td>
<td>Least-Squares Asymp. (z)</td>
<td>0.9669</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.8178</td>
<td>Aggregated Page Exact</td>
<td>0.9173</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.7834</td>
<td>Aggregated JT Exact</td>
<td>0.8831</td>
</tr>
<tr>
<td>Aggregated Page Asymp.</td>
<td>0.9173</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregated JT Asymp.</td>
<td>0.9315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.8861</td>
<td>Aligned Mean JT</td>
<td>0.9604</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.8848</td>
<td>Aligned Median JT</td>
<td>0.9603</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.8855</td>
<td>Aligned HL JT</td>
<td>0.9613</td>
</tr>
</tbody>
</table>
Table 5.12: Power Comparison under Large Alternative with More-Contaminated-Normal Error

<table>
<thead>
<tr>
<th>General Alternative</th>
<th>Power</th>
<th>Ordered Alternative</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least-Squares Exact (F)</td>
<td>0.1229</td>
<td>Least-Squares Exact (t)</td>
<td>0.2492</td>
</tr>
<tr>
<td>Least-Squares Asymp. ($\chi^2$)</td>
<td>0.1307</td>
<td>Least-Squares Asymp. (z)</td>
<td>0.2538</td>
</tr>
<tr>
<td>Aggregated Friedman Exact</td>
<td>0.2014</td>
<td>Aggregated Page Exact</td>
<td>0.3161</td>
</tr>
<tr>
<td>Aggregated Friedman Asymp.</td>
<td>0.1723</td>
<td>Aggregated Page Asymp.</td>
<td>0.3161</td>
</tr>
<tr>
<td>Aligned Mean Friedman</td>
<td>0.4072</td>
<td>Aligned Mean JT</td>
<td>0.6137</td>
</tr>
<tr>
<td>Aligned Median Friedman</td>
<td>0.4425</td>
<td>Aligned Median JT</td>
<td>0.6426</td>
</tr>
<tr>
<td>Aligned HL Friedman</td>
<td>0.4279</td>
<td>Aligned HL JT</td>
<td>0.6309</td>
</tr>
</tbody>
</table>

From the simulations with less contamination, the least-squares methods still outperform the nonparametric methods. This may be due to the ability of least-squares methods to retain power when the data is in a neighborhood "close" to the normal distribution. However, when more contamination is introduced in the data, we see the simulated powers of the least-squares methods fall below the simulated powers of the nonparametric methods. Comparing the aggregation and aligned methods across all the simulations, we generally observe that the aligned test class has higher simulated power than the aggregated test class. For each of the contaminated normal distributions, we observe an increasing simulated power as the effect size increases. Also, for a fixed alternative (effect size), we observe a decreasing power when moving from less-contaminated to more-contaminated normal distributions for the error terms.
Chapter 6

Future Work and Conclusion

6.1 Future Work

Our methods mainly apply to the V2-V3 designs, but can be easily extend to the V1 design. For example, the asymptotic properties of the aligned class test statistic are slightly different due to the different number of blocks caused by aligning with estimates of $R$ and $C$. More specifically, there are $n$ blocks with $p^2$ observations for a V1 design whereas V2 and V3 designs have $np$ blocks with $p$ observations. Hence, $E(\text{ALJ}) = \frac{n}{4}[(p^2)^2 - \sum_{i=1}^{p} p^2] = \frac{n}{4}(p^4 - p^3)$ and $\text{Var}(\text{ALJ}) = n \cdot \frac{(p^2)^2(2p^2+3) - \sum_{i=1}^{p} p^2(2p+3)}{72}$ for V1. For testing the general alternative in a V1 design, the Mack-Skillings test (Skillings and Mack (1981)) is used instead of Friedman’s test since there are replicates in each block. On the other hand, the aligned test class is not applicable with the V4 design since the dimension of $\beta$ (which includes $R_i(k)$ and $C_j(k)$) increases with $n$. However, to circumvent this, permutation/bootstrap tests could be explored for the V4 case.

For the aligned test class, we can also consider other estimates for the model coefficients such as the rank-based estimates (Rfit) by Kloke and McKean (2012) or
the weighted Wilcoxon estimates discussed in McKean and Terpstra (2005); although our simulations suggest minimal differences for different estimates. These robust estimates can also be used in place of the least-squares estimates used in the parametric methods, essentially forming a separate test class based on contrasts of the effects vector. Shao (2015) has applied this technique to certain designs.

The test classes we developed may still be applicable when \( R, C, \) and/or \( S \) are random. For instance, in V2, if \( S \) was random then we would assume that \( S_1, S_2, \ldots, S_n \) are iid \((0, \sigma^2_S)\) random variables with cdf \( F_S \). However, inspection of the proposed procedures suggest that the procedures should remain valid. On the other hand, if \( C_1, C_2, \ldots, C_p \) are iid \((0, \sigma^2_C)\) random variables with cdf \( F_C \), then it is not obvious how to proceed with an alignment method. That said, the aggregate method should remain valid since the aggregated model would simply be \( \bar{Y}_{.kl} = \mu + \bar{C} + S_k + T_l + \bar{\varepsilon}_{.kl} \), where \( \bar{C} \) is just the average random effect of \( C_1, \ldots, C_p \). That is, \( \bar{C} + S_k \) behaves like a random block effect here.

Test classes that test for the umbrella alternative (i.e. \( T_1 \leq T_2 \leq \cdots \leq T_a \geq T_{a+1} \geq \cdots \geq T_p \) where \( a \neq 1 \) and \( a \neq p \)) may also be explored. That is, it is possible to substitute the JT or Friedman tests with appropriate umbrella-alternative tests for the aggregated and aligned methods (provided that the uniform convergence theorem applies for the test). For example, a test of the umbrella alternative may capture the context of the apple sales data from Hoefnagle (1965) better since treatment \( C \) is a mixture of equal quantities of treatments A and B.

Lastly, more advanced experimental designs related to the replicated Latin square can also be explored. For example, the journal article by Choi (2012) compares the power of rank-transformed statistics testing the main effect in a \( 4 \times 4 \) Graeco-Latin square design to parametric methods. Since a Graeco-Latin square design essentially
adds another blocking variable to the model, consideration of this design using our methods suggests 8 versions (V1-V8) of the model depending on which blocking variables are nested within squares. For instance, in the model where square, row, and column are fixed, but the greek-letter blocking variable is nested within square, we would simply align with rows and columns and treat the different combinations of square and the greek-letter levels as our new blocks.

6.2 Conclusion

We present two hypothesis test classes (based on aggregating and aligning the data) for ordered and general treatment alternatives in a replicated Latin square design, with applications to datasets from existing literature. All tests may be implemented with any software capable of transforming data and performing a Friedman and JT test. The distribution properties of the test statistics are derived and expressed based on the properties for the Friedman, Page, and JT tests once the data has been appropriately transformed to the corresponding design. For the aggregated test class, exact and asymptotic distribution properties of the tests statistics can be readily obtained. For the aligned test class, since the procedure involves alignment of the data using mean-, median-, or Hodges-Lehmann-based estimates, we support the asymptotic distribution properties by presenting a generic uniform convergence theorem based on the works by Sievers (1983) and Newey (1991). Our theorem is easier to apply by expressing the ranks in the test statistics as indicator functions, and is universal enough to use in other areas where estimates of nuisance parameters are needed. For instance, consider the use of Mack-Skillings’ Test for a general alternative on a V1 design where observations are aligned with row and column effect estimates. Similar to Proof 7.4 for Friedman’s test statistic, Mack-Skillings’ statistic can be expressed as indicator functions with
the conjecture that some $S_n(\sqrt{n}(\beta - \hat{\beta})) = S_n(0) = o_p(1)$. The theorem also provides the groundwork for proving consistency and asymptotic normality of tests based on residuals. This is similar to the idea of testing normality in the residual analysis of a experiment design, where the Shapiro-Wilk test is performed on the residuals instead of the observations.

A simulation study compares our aggregated and aligned tests to the least-squares counterparts. In general, our proposed tests maintain their significance levels and outperform the parametric tests when the data is generated from a sufficiently contaminated normal distribution. Moreover, the aligned tests tend to be more powerful than the aggregated tests under our simulation settings. Based on our findings from the simulation studies, we can gather the following comparisons for the aligned and aggregated methods:

- An exact distribution of the aggregated test statistic is available for a design with a small number of square replicates, making the aggregated method preferable for this case.

- Under our simulation settings for the aligned tests, the similarity in the simulated power and Type I error rates between different estimates for alignment suggest that the choice of estimates is not critical.

- The aggregated method averages observations to transform the design from a replicated Latin square to a RCBD, so the reduced number of observations causes a loss of power compared to the aligned test under our simulation settings.
Chapter 7

Appendix

7.1 Proof that the MSE is a Consistent Estimator for $\sigma^2$

Take the ANOVA table for the model of a V2 design, the MSE is

$$\hat{\sigma}^2 = \frac{1}{(np - 2)(p - 1)} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{n} \bar{\varepsilon}_{ijkl}^2$$

$$= \frac{np^2}{(np - 2)(p - 1)} \cdot \frac{1}{np^2} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{j=1}^{p} \bar{\varepsilon}_{ijkl}^2$$

let

$$= \frac{np^2}{(np - 2)(p - 1)} \cdot \frac{1}{n} \sum_{k=1}^{n} \omega_k$$

where $\omega_k = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \bar{\varepsilon}_{ijkl}^2$.

The first part, $\frac{np^2}{(np - 2)(p - 1)}$, converges to $\frac{p}{p-1}$ by L’Hopital’s rule, so the goal of this proof is to show that $\frac{1}{n} \sum_{k=1}^{n} \omega_k \overset{p}{\rightarrow} \frac{p-1}{p} \sigma^2$.

Taking apart $\omega_k$,
\[ \omega_k = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (Y_{ijkl} - \mu - \hat{R}_{i(k)} - \hat{C}_j - \hat{S}_k - \hat{T}_l)^2 \]

\[ = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\epsilon_{ijkl} + (\mu - \hat{\mu}) + (R_{i(k)} - \hat{R}_{i(k)}) + (C_j - \hat{C}_j) + (S_k - \hat{S}_k) + (T_l - \hat{T}_l))^2 \]

\[ = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\epsilon_{ijkl} + \mu - \hat{Y}_{i.k} + R_{i(k)} + S_k + (C_j - \hat{C}_j) + (T_l - \hat{T}_l))^2. \]

From Proof 7.2, it is understood that for \( \beta = (C_1, C_2, \ldots, C_p) \) or \((T_1, T_2, \ldots, T_p)\), \( \sqrt{n}(\beta - \hat{\beta}) = O_p(1) \). Since \( \frac{1}{\sqrt{n}} \) is \( o_p(1) \), \( \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\beta - \hat{\beta}) = o_p(1) \).

Note that

\[ \hat{Y}_{i.k} = \frac{1}{p} \sum_{j=1}^{p} Y_{ijkl} \]

\[ = \frac{1}{p} \sum_{j=1}^{p} (\mu + R_{i(k)} + C_j + S_k + T_l + \epsilon_{ijkl}) \]

\[ = \mu + R_{i(k)} + S_k + \epsilon_{i.k}. \]

So,

\[ \omega_k = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\epsilon_{ijkl} - \epsilon_{i.k} + (C_j - \hat{C}_j) + (T_l - \hat{T}_l))^2. \]

Ignoring the terms that are \( o_p(1) \), i.e. \( \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\epsilon_{ijkl} - \epsilon_{i.k})^2 \), we note that this sum depends on the \( \epsilon_{ijkl}'s \) in the \( i \)th row nested within the \( k \)th square. Furthermore,

\[ \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\epsilon_{ijkl} - \epsilon_{i.k})^2 = \frac{p-1}{p^2} \sum_{i=1}^{p} \frac{1}{p-1} \sum_{j=1}^{p} (\epsilon_{ijkl} - \epsilon_{i.k})^2 \leq \frac{p-1}{p^2} \sum_{i=1}^{p} s_{\epsilon_{ik}}^2 \text{ where } s_{\epsilon_{ik}} \]
is the sample variance of \(\{\varepsilon_{1kl}, \varepsilon_{2kl}, \ldots, \varepsilon_{pkl}\}\). Since \(E(s_{\varepsilon_{ik}}^2) = \sigma^2\),

\[
E\left(\frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i,k})^2\right) = \frac{p - 1}{p} \sigma^2
\]

which suggests that \(E(\omega_k) \approx \frac{p - 1}{p} \sigma^2\).

Define \(\omega_k(X, Y) = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i,k} + \lambda_j^T X + \kappa_l^T Y)\), where \(X = (X_1, \ldots, X_p)^T\), \(Y = (Y_1, \ldots, Y_p)^T\), and \(\lambda_j\) and \(\kappa_l\) are \(p\)-dimensional vectors with 1s in the \(j\)th or \(l\)th spot respectively, and 0s otherwise. We may rewrite \(\omega_k(X, Y) = \omega_k(v) = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i,k} + d_{jl}^T v)\), where \(v = (X^T, Y^T)^T\) and \(d_{jl} = (\lambda_j^T, \kappa_l^T)^T\). Then, \(\hat{\sigma}^2 = \frac{np^2}{(np-2)(p-1)} \cdot \frac{1}{n} \sum_{k=1}^{n} \omega_k(C - \hat{C}, T - \hat{T})\). Note that \(\omega_k(0, 0) = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i,k})^2\) and \(E[\omega_k(0, 0)] = \frac{p - 1}{p} \sigma^2\).

Furthermore, \(\frac{1}{n} \sum_{k=1}^{n} \omega_k(0, 0) \overset{p}{\longrightarrow} E[\omega_k(0, 0)]\) since \(\omega_1(0, 0), \ldots, \omega_n(0, 0)\) are iid and \(E[\omega_k(0, 0)] = \sigma^2 \left(\frac{p - 1}{p}\right)\) is finite. Next, define \(f_n(v) = \frac{1}{n} \sum_{k=1}^{n} \omega_k(v)\) and consider the first-order Taylor expansion,

\[
f_n(v) = f_n(0) + \frac{1}{1!} \nabla f_n^T(g_n)(v - 0)
= f_n(0) + \nabla f_n^T(g_n)v
\]

where \(g_n\) falls on the line segment connecting 0 and \(v\). That is, \(g_n = (1 - \lambda_n)0 + \lambda_n v\) for some \(\lambda_n \in (0, 1)\). Recall that \(f_n(0) = \frac{1}{n} \sum_{k=1}^{n} \left[ \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i,k})^2 \right] \overset{p}{\rightarrow} \frac{p - 1}{p} \sigma^2\).

Let \(\hat{\nu} = \left(\hat{C} - C, \hat{T} - T\right)\) so that \(f_n(\hat{\nu}) = f_n(0) + \nabla f_n^T(g_n)\hat{\nu}\) where \(g_n = \lambda_n \hat{\nu}\) for some \(\lambda_n \in (0, 1)\). Note that \(||g_n|| = ||\lambda_n \hat{\nu}|| = ||\lambda_n|| \cdot ||\hat{\nu}|| \leq ||\hat{\nu}||\) and \(||\hat{\nu}|| = o_p(1)\). Hence, \(g_n = o_p(1)\).

Note that \(\nabla f_n(v) = (\delta f_n/\delta X_1, \ldots, \delta f_n/\delta X_p, \delta f_n/\delta Y_1, \ldots, \delta f_n/\delta Y_p)^T\) where
$f_n(v) = f_n(X, Y) = \frac{1}{n} \sum_{k=1}^{n} \left[ \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{j=p} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i \cdot k} + X_j + Y_i)^2 \right]$. 

For $\frac{\delta f_n}{\delta X_m}$:

$$f_n(v) = \frac{1}{n} \sum_{k=1}^{n} \left[ \frac{1}{p^2} \sum_{j(i,k,l)=m} (\varepsilon_{imkl} - \bar{\varepsilon}_{i \cdot k} + X_m + Y_i)^2 \right] + \frac{1}{p^2} \sum_{j(i,k,l)\neq m} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i \cdot k} + X_j + Y_i)^2 \right]$$

$$\therefore \frac{\delta f_n(v)}{\delta X_m} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{p^2} \sum_{j(i,k,l)=m} 2(\varepsilon_{imkl} - \bar{\varepsilon}_{i \cdot k} + X_m + Y_i)(1)$$

$$= \frac{2}{p^2} \left[ \frac{1}{n} \sum_{k=1}^{n} (\varepsilon_{m \cdot k} - \bar{\varepsilon}_{\cdot \cdot k}) + X_m + \bar{Y} \right].$$

For $\frac{\delta f_n}{\delta Y_m}$:

$$f_n(v) = \frac{1}{n} \sum_{k=1}^{n} \left[ \frac{1}{p^2} \sum_{l(i,j,k)=m} (\varepsilon_{ijklm} - \bar{\varepsilon}_{i \cdot k} + X_j + Y_m)^2 \right] + \frac{1}{p^2} \sum_{l(i,j,k)\neq m} (\varepsilon_{ijkl} - \bar{\varepsilon}_{i \cdot k} + X_j + Y_i)^2 \right]$$

$$\therefore \frac{\delta f_n(v)}{\delta Y_m} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{p^2} \sum_{l(i,j,k)=m} 2(\varepsilon_{ijklm} - \bar{\varepsilon}_{i \cdot k} + X_j + Y_m)(1)$$

$$= \frac{2}{p^2} \left[ \frac{1}{n} \sum_{k=1}^{n} (\varepsilon_{\cdot \cdot km} - \bar{\varepsilon}_{\cdot \cdot k}) + X + Y_m \right].$$

To show that $\nabla f_n(T) \hat{v} = o_p(1)$ where $\hat{v} = (\hat{\mathbf{C}} - \mathbf{C})^T, (\hat{T} - \mathbf{T})^T$ and $g_n = o_p(1)$, note that $\hat{v} = o_p(1)$ since (for example), $\sqrt{n}(\hat{\mathbf{C}} - \mathbf{C}) = O_p(1)$ and $\hat{\mathbf{C}} - \mathbf{C} = \frac{1}{\sqrt{n}} O_p(1) = o_p(1)$. Hence, we only need to show $\nabla f_n(g_n) = o_p(1)$. This will follow if we can show each component of $\nabla f_n(g_n) = o_p(1)$.

Consider $\frac{\delta f_n(v)}{\delta Y_m} |_{v=g_n}$ for the sake of illustration, and let $g_n = (g_n^{\mathbf{x}^T}, g_n^{\mathbf{y}^T})^T$ and recall that $g_n = o_p(1)$. Thus,
\[
\frac{\delta f_n(v)}{\partial Y_m} \Big|_{v=g_n} = \frac{2}{p} \left[ \frac{1}{n} \sum_{k=1}^{n} (\bar{\varepsilon}_{-km} - \bar{\varepsilon}_{..k}) + \frac{1}{p} \sum_{j=1}^{p} g_{nj}^X + g_{nm}^Y \right].
\]

Note that \( \frac{1}{n} \sum_{k=1}^{n} (\bar{\varepsilon}_{-km} - \bar{\varepsilon}_{..k}) \xrightarrow{p} 0 \) since \( \{\bar{\varepsilon}_{-km} - \bar{\varepsilon}_{..k}\}_{k=1}^{n} \) are iid \((0, \Sigma_\varepsilon)\) similar to the proof in 7.2.

\[
\frac{1}{p} \sum_{j=1}^{p} g_{nj}^X = o_p(1)
\]
since these sums are finite and \( g_n = o_p(1) \). Therefore, \( \frac{\delta f_n(V)}{\partial Y_m} \Big|_{v=g_n} = o_p(1) \), which shows that \( \nabla f_n^\top(g_n) = o_p(1) \). Hence,

\[
\hat{\sigma}^2 = \frac{np^2}{(np-2)(p-1)} \cdot \frac{1}{n} \sum_{k=1}^{n} \omega_k \langle C - \hat{C}, T - \hat{T} \rangle
\]

\[
= \frac{np^2}{(np-2)(p-1)} [f_n(0) + \nabla f_n^\top(g_n) \hat{\nu}]
\]

\[
\xrightarrow{p} \frac{p}{p-1} \left[ \frac{p-1}{p} \sigma^2 + 0 \right] = \sigma^2.
\]

### 7.2 Proof that \( \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \) for Aligned JT Test

For the test based on aligning observations, consider the V3 model where \( E(\varepsilon_{ijkl}) = 0, \text{Var}(\varepsilon_{ijkl}) = \sigma^2 \), and the \( \varepsilon_{ijkl} \) are iid. Take \( \beta = (R_1, R_2, \ldots, R_p)\top \) and define marginal deviations \( D_{ik} = \bar{Y}_{i.k} - \bar{Y}_{..k} \). The deviations have model \( D_{ik} = R_i + (\bar{\varepsilon}_{i.k} - \bar{\varepsilon}_{..k}) \). In matrix form, the \( D_{ik} \)'s act as a \( n \times p \) matrix where the \( n \) rows (squares) are independent, but the \( p \) columns (due to \( \bar{Y}_{..k} \)) are not independent. Hence, the \( D_{ik} \) are essentially multivariate observations corresponding to a multivariate location model. Thus, \( \hat{\beta} = (\bar{D}_1, \bar{D}_2, \ldots, \bar{D}_p)\top \). Now, consider the following calculations: \( \text{Var}(\bar{\varepsilon}_{i.k}) = \frac{\sigma^2}{p}, \bar{\varepsilon}_{..k} = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \varepsilon_{ijkl} = \frac{1}{p} \sum_{i=1}^{p} \bar{\varepsilon}_{i.k} \), and \( \text{Cov}(\bar{\varepsilon}_{i.k}, \bar{\varepsilon}_{..k}) = \)
\[ \text{Cov}(\bar{\varepsilon}_{i.k}, \frac{1}{p} \sum_{i=1}^{p} \bar{\varepsilon}_{i.k}) = \text{Cov}(\bar{\varepsilon}_{i.k}, \frac{1}{p}(\bar{\varepsilon}_{1.k} + \cdots + \bar{\varepsilon}_{p.k})) = \frac{1}{p} \text{Cov}(\bar{\varepsilon}_{i.k}, \bar{\varepsilon}_{i.k}) = \frac{\sigma^2}{p}. \]

It follows that \( \text{Var}(D_{ik}) = \frac{\sigma^2}{1 - \frac{1}{p^2}} \) and \( \text{Cov}(D_{ik}, D_{i'k}) = -\frac{\sigma^2}{p^2} \). In summary, the covariance matrix of \( \{D_{ik}\}_{i=1}^{p} \) is \( \Sigma_{\varepsilon} = \sigma^2 \begin{pmatrix} \frac{1}{p^2} & \cdots & -\frac{1}{p^2} \\ \vdots \end{pmatrix}_{p \times p} \) and \( E(D_{ik}) = 0 \).

So, \( \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N_p(0, \Sigma_{\varepsilon}) \) by the CLT, which implies that \( \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \).

### 7.3 Proof that \( S_n(\sqrt{n}(\hat{\beta} - \beta)) - S_n(0) = o_p(1) \) for the Aligned JT Test

This section discusses the proof for the following three conditions:

1. \( S_n(\Delta) - S_n(0) = o_p(1) \),

2. \( \sup_{||\Delta|| \leq C} |S_n(\Delta) - S_n(0)| = o_p(1) \), and

3. \( ||\sqrt{n}(\hat{C} - C)|| = O_p(1) \).

Only conditions 1 and 2 are discussed here since Section 7.2 discusses condition 3. Starting with condition 1,
\[ E(S_n(\Delta) - S_n(0)) \]

\[
= E\left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} \left[ I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}) - I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0) \right] \right\} 
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} E[I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}) - I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0)] 
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} E\left[ E[I(\varepsilon_{ikl_1} < \varepsilon_{ikl_2} + (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}})] \right] - E[I(\varepsilon_{ikl_1} < \varepsilon_{ikl_2})] 
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} E[ E[I(\varepsilon_{ikl_1} + (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}) - P(\varepsilon_{ikl_1} < \varepsilon_{ikl_2})] 
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} \int [E[I(\varepsilon_{ikl_1} + (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}) - E[I(\varepsilon_{ikl_1}])] f(\varepsilon) d\varepsilon = E_N. 
\]

By the mean value theorem on \( F \), let \( \delta_{ikl_1 l_2} \) be some point between \( \varepsilon \) and \( \varepsilon + (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}} \) that depends on \( i, k, l_1, l_2 \). Then,

\[
E_N = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} \int f(\delta_{ikl_1 l_2}) [(\varepsilon + (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}) - \varepsilon] f(\varepsilon) d\varepsilon
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \int f(\delta_{ikl_1 l_2}) f(\varepsilon) d\varepsilon
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \int [f(\delta_{ikl_1 l_2}) - f(\varepsilon)] f(\varepsilon) d\varepsilon
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \int [f(\delta_{ikl_1 l_2}) - f(\varepsilon)] f(\varepsilon) d\varepsilon
\]

\[
+ \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \int f^2(\varepsilon) d\varepsilon
\]

\[
= E_{N_1} + E_{N_2}
\]
Starting with $E_{N_1}$, the distance between $\delta_{ikl_1l_2}$ and $\varepsilon$ is less than or equal to $|\lambda_{ikl_1} - \lambda_{ikl_2}|^T \frac{\Delta}{\sqrt{n}}$. Furthermore,

\[
|\lambda_{ikl_1} - \lambda_{ikl_2}|^T \frac{\Delta}{\sqrt{n}} \leq \frac{1}{\sqrt{n}} |\lambda_{ikl_1} - \lambda_{ikl_2}|^T \Delta | \leq \frac{1}{\sqrt{n}} \| (\lambda_{ikl_1} - \lambda_{ikl_2}) \| |\Delta| \| \text{ (by the Cauchy-Schwartz inequality)} \leq \frac{1}{\sqrt{n}} (\| \lambda_{ikl_1} \| + \| \lambda_{ikl_2} \|) |\Delta| = \frac{1}{\sqrt{n}} 2 |\Delta| \to 0.
\]

Since $f$ is assumed uniformly continuous, for $\varepsilon' > 0$, there exists $\delta(\varepsilon')$ so that

\[
|f(\delta_{ikl_1l_2}) - f(\varepsilon)| < \frac{\varepsilon'}{p(\frac{p}{2})^2 |\Delta|} \quad \text{if } |\delta_{ikl_1l_2} - \varepsilon| < \delta(\varepsilon')
\]

It is shown above that

\[
|\delta_{ikl_1l_2} - \varepsilon| \leq \frac{2 |\Delta|}{\sqrt{n}} < \delta(\varepsilon') \quad \text{if } n > \frac{4 |\Delta|^2}{\delta^2(\varepsilon')} = N(\varepsilon').
\]

Therefore, for $n > N(\varepsilon')$,

\[
|E_{N_1} - 0| = \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^T \Delta \int [f(\delta_{ikl_1l_2}) - f(\varepsilon)] f(\varepsilon) d\varepsilon - 0 \right|
\]

\[
\leq \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} 2 |\Delta| \int |f(\delta_{ikl_1l_2}) - f(\varepsilon)| f(\varepsilon) d\varepsilon
\]

\[
\leq \frac{1}{n} \sum_{k=1}^{n} p(\frac{p}{2})^2 |\Delta| \frac{\varepsilon'}{p(\frac{p}{2})^2 |\Delta|} = \varepsilon'.
\]
That is, \( \lim_{n \to \infty} E_{N_1} = 0 \).

For \( E_{N_2} \), consider \( \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \) where \( \Delta \) is a \( p \times 1 \) vector. Take, for example, a Latin square where the columns are arranged as shown in table 7.1.

Table 7.1: Latin Square with Arranged Columns

<table>
<thead>
<tr>
<th></th>
<th>T_1</th>
<th>T_2</th>
<th>T_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_1</td>
<td>( \Delta_1 )</td>
<td>( \Delta_2 )</td>
<td>( \Delta_3 )</td>
</tr>
<tr>
<td>R_2</td>
<td>( \Delta_2 )</td>
<td>( \Delta_3 )</td>
<td>( \Delta_1 )</td>
</tr>
<tr>
<td>R_3</td>
<td>( \Delta_3 )</td>
<td>( \Delta_1 )</td>
<td>( \Delta_2 )</td>
</tr>
</tbody>
</table>

For the example above,

\[
\sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta
= (\Delta_1 - \Delta_2) + (\Delta_1 - \Delta_3) + (\Delta_2 - \Delta_3)
+ (\Delta_2 - \Delta_3) + (\Delta_3 - \Delta_1) + (\Delta_2 - \Delta_1)
+ (\Delta_3 - \Delta_1) + (\Delta_1 - \Delta_2) + (\Delta_3 - \Delta_2)
= 0
\]

In general, the combinations for \( i, l_1, \) and \( l_2 \) and the arrangement of the Latin letters result in \( (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta = \Delta_{ikl_1} - \Delta_{ikl_2} \). Therefore \( \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \) simplifies to \( \binom{p}{2} \) pairs of \( \sum \Delta - \sum \Delta = 0 \). Combined with \( \int f^2(\varepsilon) d\varepsilon \) is finite since \( f \) is assumed to be bounded, \( E_{N_2} = 0 \). Hence, it has been shown that \( E_N = E_{N_1} + E_{N_2} \to 0 \).

For \( \text{Var}(S_n(\Delta) - S_n(0)) \),
\[ \text{Var}(S_n(\Delta) - S_n(0)) = E[(S_n(\Delta) - S_n(0))^2] - E^2[S_n(\Delta) - S_n(0)] \leq E[(S_n(\Delta) - S_n(0))^2]. \]

So, it is sufficient to show that \( E[(S_n(\Delta) - S_n(0))^2] \to 0 \) to prove that \( S_n(\Delta) - S_n(0) = o_p(1) \). Let \( W_{nk} = \sum_{i=1}^{p} \sum_{l_1 < l_2} (I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\Delta}{\sqrt{n}}) - I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0)) \) and note that \( |W_{nk}| \leq 2p(p) \). Then,

\[
(S_n(\Delta) - S_n(0))^2 = \frac{1}{n} \sum_{k=1}^{n} \sum_{k'=1}^{n} W_{nk} W_{nk'}
\]

\[
= \frac{1}{n} \sum_{k=k'}^{n} W_{nk}^2 + \frac{1}{n} \sum_{k \neq k'}^{n} W_{nk} W_{nk'}
\]

\[
= W_1 + W_2
\]

So, \( E[(S_n(\Delta) - S_n(0))^2] = E(W_1 + W_2) = E(W_1) + E(W_2) \). Consider \( E(W_2) \) first.

\[
E(W_2) = \frac{1}{n} \sum_{k \neq k'}^{n} E(W_{nk} W_{nk'})
\]

\[
= \frac{1}{n} \sum_{k \neq k'}^{n} E(W_{nk}) E(W_{nk'}) \quad \because \text{the } n \text{ squares are independent.}
\]

\[
E(W_2) = \frac{1}{\sqrt{n}} \sum_{k} E(W_{nk}) \frac{1}{\sqrt{n}} \sum_{k'}^{n} E(W_{nk'}) \to 0 \quad \because E_N = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} W_{nk} \to 0.
\]
Consider $E(V_{W_1})$ next.

\[
E(V_{W_1}) \leq |E(V_{W_1})| = \frac{1}{n} \sum_{k=1}^{n} E(W_{nk}^2) \\
\leq \frac{1}{n} \sum_{k=1}^{n} E(|W_{nk}| |W_{nk}|) \\
\leq \frac{1}{n} \sum_{k=1}^{n} E(2p \binom{p}{2} |W_{nk}|) = \frac{2p \binom{p}{2}}{n} \sum_{k=1}^{n} E(|W_{nk}|).
\]

Let us first define the sets $a_1, a_2, a_3$ of \{i, k, l_1, l_2\} such that $(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta > 0$, $(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta = 0$, and $(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta < 0$ respectively. Using these sets, let us discuss $W_{nk}$ in terms of these three sets, i.e. $W_{nk} = W_{nk_1} + W_{nk_2} + W_{nk_3}$ where

\[
W_{nk_1} = \sum_{a_1} I(0 < \varepsilon_{ikl_1} - \varepsilon_{ikl_2} < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \sqrt{n} ) \quad \text{almost everywhere},
\]

\[
W_{nk_2} = I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0) - I(\varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0) = 0,
\]

\[
W_{nk_3} = -\sum_{a_3} I((\lambda_{ikl_1} - \lambda_{ikl_2})^\top \Delta \sqrt{n} < \varepsilon_{ikl_1} - \varepsilon_{ikl_2} < 0) \quad \text{almost everywhere}.
\]

So, \[
2p \binom{p}{2} \frac{n}{n} \sum_{k=1}^{n} E(|W_{nk}|) = \frac{2p \binom{p}{2}}{n} \sum_{k=1}^{n} E(|W_{nk_1} + W_{nk_3}|) \\
\leq \frac{2p \binom{p}{2}}{n} \sum_{k=1}^{n} E(|W_{nk_1}| + |W_{nk_3}|) \\
= \frac{2p \binom{p}{2}}{n} \sum_{k=1}^{n} [E(|W_{nk_1}|) + E(|W_{nk_3}|)].
\]

Thus, the goal is to show that $\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk_1}|) \to 0$ and $\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk_3}|) \to 0$. To begin, consider the following:
\[
\frac{1}{n} \sum_{k=1}^{n} E(\left| W_{nk1} \right|) = \frac{1}{n} \sum_{k=1}^{n} E(\left( \sum_{a_1} I(0 < \varepsilon_{ikl1} - \varepsilon_{ikl2} < (\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}) \right))
\leq \frac{1}{n} \sum_{k=1}^{n} E(\sum_{a_1} I(0 < \varepsilon_{ikl1} - \varepsilon_{ikl2} < (\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}))
\leq \frac{1}{n} \sum_{k=1}^{n} \sum_{a_1} E(I(0 < \varepsilon_{ikl1} - \varepsilon_{ikl2} < (\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}))
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_1} [G_{\varepsilon_{1-\varepsilon_2}}((\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}) - G_{\varepsilon_{1-\varepsilon_2}}(0)].
\]

By the mean value theorem, \( \frac{G_{\varepsilon_{1-\varepsilon_2}}((\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}) - G_{\varepsilon_{1-\varepsilon_2}}(0)}{(\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}} = G'(c_{ikl1l2}) \) where \( 0 < c_{ikl1l2} < (\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}} \). Therefore,

\[
\frac{1}{n} \sum_{k=1}^{n} \sum_{a_1} [G_{\varepsilon_{1-\varepsilon_2}}((\lambda_{ikl1} - \lambda_{ikl2})^\top \frac{\Delta}{\sqrt{n}}) - G_{\varepsilon_{1-\varepsilon_2}}(0)]
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_1} [(\lambda_{ikl1} - \lambda_{ikl2})^\top \Delta G'_{\varepsilon_{1-\varepsilon_2}}(c_{ikl1l2})]
= \frac{1}{n\sqrt{n}} \sum_{k=1}^{n} \sum_{a_1} [(\lambda_{ikl1} - \lambda_{ikl2})^\top \Delta G'_{\varepsilon_{1-\varepsilon_2}}(c_{ikl1l2})],
\]

where \( (\lambda_{ikl1} - \lambda_{ikl2})^\top \Delta \) is bounded by \( 2||\Delta|| \).
Considering the distribution function $G$,

\[
G_{\varepsilon_1-\varepsilon_2}(c_{ikl_1l_2}) = P(\varepsilon_1 - \varepsilon_2 \leq c_{ikl_1l_2}) = P(\varepsilon_1 \leq c_{ikl_1l_2} + \varepsilon_2) \\
= E[I(\varepsilon_1 \leq c_{ikl_1l_2} + \varepsilon_2)] \\
= E[E[I(\varepsilon_1 \leq c_{ikl_1l_2} + \varepsilon_2) | \varepsilon_2]] \\
= E[F(c_{ikl_1l_2} + \varepsilon_2)] \\
= \int F(c_{ikl_1l_2} + \varepsilon_2) f(\varepsilon_2) d\varepsilon_2 \\
\therefore G'(c_{ikl_1l_2}) = \int f(c_{ikl_1l_2} + \varepsilon_2) f(\varepsilon_2) d\varepsilon_2.
\]

However, $f$ is assumed to be bounded, so $G'(c_{ikl_1l_2})$ is bounded as well. Therefore,

\[
\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk_1}|) \leq \frac{1}{\sqrt{n}} 2||\Delta||B_G B_1 \to 0
\]

where $B_1$ is a bound for the number of elements in $a_1$.

Similarly, for $\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk_3}|)$,
\[
\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk}|) = \frac{1}{n} \sum_{k=1}^{n} E\left(\left| \sum_{a_3} I(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl1} - \varepsilon_{ikl2} < 0) \right| \right)
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} E\left(\left| \sum_{a_3} I(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl1} - \varepsilon_{ikl2} < 0) \right| \right)
\]
\[
\leq \frac{1}{n} \sum_{k=1}^{n} E\left(\sum_{a_3} I(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl1} - \varepsilon_{ikl2} < 0) \right)
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} E\left(\left| \sum_{a_3} I(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl1} - \varepsilon_{ikl2} < 0) \right| \right)
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} [G_{\varepsilon_1-\varepsilon_2}(0) - G_{\varepsilon_1-\varepsilon_2}(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}})]
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} \left[ -\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} G'(d_{ikl12}) \right]
\]

where \(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} < d_{ikl12} < 0\) and \(\frac{G_{\varepsilon_1-\varepsilon_2}(0) - G_{\varepsilon_1-\varepsilon_2}(\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}})}{\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}}} = G'_{\varepsilon_1-\varepsilon_2}(d_{ikl12})\) by the mean value theorem. Therefore,

\[
\frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} \left[ -\langle \lambda_{ikl1} - \lambda_{ikl2} \rangle^T \frac{\Delta}{\sqrt{n}} G'(d_{ikl12}) \right] \leq \frac{1}{\sqrt{n}} 2||\Delta||B_GB_2 \to 0.
\]

From the above results we now have shown that \(E(V_W) \leq \frac{2p}{n} \sum_{k=1}^{n} E(|W_{nk}|) = \frac{1}{n} \sum_{k=1}^{n} E(|W_{nk1}| + |W_{nk3}|) \to 0\). Therefore, \(Var(S_n(\Delta) - S_n(0)) \to 0\), completing the proof for condition 1. For condition 2, we refer to the uniform convergence theorem in Section 7.5 with \(\hat{Q}_n(\Delta) = S_n(\Delta) - S_n(0)\). Assumption 1 is satisfied with \(\hat{Q}_n(\Delta) = 0\) since the proof for condition 1 shows that \(S_n(\Delta) - S_n(0) = o_p(1)\). Thus, it immediately follows that assumption 2 is satisfied.
For assumption 3, we will start by defining $U_n(\Delta, r)$.

$$
|\hat{Q}_n(\Delta) - \hat{Q}_n(\tilde{\Delta})| = \left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}] \right|

- \left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\tilde{\Delta}}{\sqrt{n}}] \right|

\leq \left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} |I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}}]|

- I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\tilde{\Delta}}{\sqrt{n}}]| \right|

Define sets $a_1$, $a_2$, $a_3$ of $\{i, k, l_1, l_2\}$ such that

$$(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}} > (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\tilde{\Delta}}{\sqrt{n}},$$

$$(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}} = (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\tilde{\Delta}}{\sqrt{n}},$$

and $$(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\Delta}{\sqrt{n}} < (\lambda_{ikl_1} - \lambda_{ikl_2})^\top \frac{\tilde{\Delta}}{\sqrt{n}}$$ respectively.

Then,
\[
\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\Delta}{\sqrt{n}}] \\
- I[(\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}}]] \\
= \frac{1}{\sqrt{n}} \sum_{a_1} I[(\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}} < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\Delta}{\sqrt{n}}] \\
+ \frac{1}{\sqrt{n}} \sum_{a_3} I[(\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\Delta}{\sqrt{n}} < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}}] \\
\leq \frac{1}{\sqrt{n}} \sum_{a_1} I(0 < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) - (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) \\
+ \frac{1}{\sqrt{n}} \sum_{a_3} I(0 < (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}} - (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) < \frac{1}{\sqrt{n}} B\lambda r) \\
\leq \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I(-\frac{1}{\sqrt{n}} B\lambda r < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) - (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) \\
\overset{\text{let}}{=} U_n(\tilde{\Delta}, r)
\]
Next, we set out to verify assumptions 3.1 and 3.2.

For assumption 3.1, let

\[
I(-\frac{1}{\sqrt{n}} B\lambda r < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) - \frac{(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) = I_{ikl_1l_2},
\]

\[
|\text{Var}(U_n(\tilde{\Delta}, r))| = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{p} \text{Var}(\sum_{l_1 < l_2} I_{ikl_1l_2})
\]

\[
= \sum_{l_1 < l_2} \text{Var}(I_{ikl_1l_2}) + \sum_{l_1 < l_2} \sum_{l'_1 < l'_2} \text{Cov}(I_{ikl_1l_2}, I_{ikl'_1l'_2})
\]

\[
\leq \sum_{l_1 < l_2} \text{Var}(I_{ikl_1l_2}) + \sum_{l_1 < l_2} \sum_{l'_1 < l'_2} |s_{ikl_1l_2} s_{ikl'_1l'_2}|
\]

\[
\leq \sum_{l_1 < l_2} \text{Var}(I_{ikl_1l_2}) + \sum_{l_1 < l_2} \sum_{l'_1 < l'_2} |s_{ikl_1l_2}||s_{ikl'_1l'_2}| \text{ where } s_{ikl_1l_2} = \sqrt{\text{Var}(I_{ikl_1l_2})}.
\]

So, it is sufficient to show that \(\text{Var}(I_{ikl_1l_2}) \to 0 \quad \forall i, k, l_1, l_2\). Consider the following:

\[
\text{Var}(I_{ikl_1l_2}) = E(I_{ikl_1l_2}^2) - (E(I_{ikl_1l_2}))^2
\]

\[
= E(I_{ikl_1l_2})[1 - E(I_{ikl_1l_2})]
\]

\[
\leq E(I_{ikl_1l_2}) \cdot |1 - E(I_{ikl_1l_2})| \leq 1
\]

\[
E(I_{ikl_1l_2}) = E(I(-\frac{1}{\sqrt{n}} B\lambda r < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) - \frac{(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r))
\]

\[
= \frac{2}{\sqrt{n}} B\lambda r f_{\varepsilon_{ikl_1} - \varepsilon_{ikl_2}}(c_{nk\Delta}) \text{ by the mean value theorem}
\]

where \(\frac{(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \tilde{\Delta}}{\sqrt{n}} - \frac{1}{\sqrt{n}} B\lambda r < c_{nk\Delta} < \frac{(\lambda_{ikl_1} - \lambda_{ikl_2})^\top \tilde{\Delta}}{\sqrt{n}} + \frac{1}{\sqrt{n}} B\lambda r\).

\(\therefore\) \(\text{Var}(U_n(\tilde{\Delta}, r)) = o(1) \cdot \cdot f \text{ is bounded.}\)
For assumption 3.2,

\[
E(U_n(\tilde{\Delta}, r)) = E\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} I(-\frac{1}{\sqrt{n}} B \lambda r < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) - (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B \lambda r)\right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} E(I(-\frac{1}{\sqrt{n}} B \lambda r < (\varepsilon_{ikl_1} - \varepsilon_{ikl_2}) - (\lambda_{ikl_1} - \lambda_{ikl_2})^T \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B \lambda r))
\]

\[
= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} \left[\frac{2}{\sqrt{n}} B \lambda r f_{\varepsilon_{ikl_1} - \varepsilon_{ikl_2}}(c_{nk\tilde{\Delta}})\right]
\]

\[
= \frac{2}{n} B \lambda r \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l_1 < l_2} f_{\varepsilon_{ikl_1} - \varepsilon_{ikl_2}}(c_{nk\tilde{\Delta}}) \leq 2B \lambda B f r
\]

\[
\leq 2p \left(\frac{p}{2}\right) B \lambda B f r
\]

where \(f_{\varepsilon_{ikl_1} - \varepsilon_{ikl_2}}\) is bounded by \(B f\).

Therefore, \(E(U_n(\tilde{\Delta}, r)) = O(r)\).

### 7.4 Proof that \(S_n(\sqrt{n}(\hat{\beta} - \beta)) - S_n(0) = o_p(1)\) for Aligned Friedman Test

Similar to the proof for the ordered-alternative aligned JT statistic, the following needs to be shown:

1. \(S_n(\Delta) - S_n(0) = o_p(1)\),

2. \(\sup_{||\Delta|| \leq C} |S_n(\Delta) - S_n(0)| = o_p(1)\), and

3. \(||\sqrt{n}(\hat{C} - C)|| = O_p(1)\).

Starting with condition 1,
\[ E(S_n(\Delta) - S_n(0)) = E\left[ \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{1}{\sqrt{n}} \Delta) \right] \\
- \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} I(\varepsilon_{ikl} - \varepsilon_{ik1} < 0)] \]
\[= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} E[I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{1}{\sqrt{n}} \Delta) \\
- I(\varepsilon_{ikl} - \varepsilon_{ik1} < 0)] \]
\[= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} E\{E[I(\varepsilon_{ikl} < \varepsilon_{ik1} + (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}) | \varepsilon_{ik1}] \\
- E[I(\varepsilon_{ikl} < \varepsilon_{ik1}) | \varepsilon_{ik1}]\} \]
\[= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} E[P(\varepsilon_{ikl} < \varepsilon_{ik1} + (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} | \varepsilon_{ik1}) \\
- P(\varepsilon_{ikl} < \varepsilon_{ik1} | \varepsilon_{ik1})] \]
\[= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} E[F(\varepsilon + (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}) - F(\varepsilon)] \]
\[= \int \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} \int [F(\varepsilon + (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}) - F(\varepsilon)] f(\varepsilon) d\varepsilon \]
\[= E_N. \]

By the mean value theorem on \( F \), let \( \delta_{ikl} \) be some point between \( \varepsilon \) and \( \varepsilon + (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} \). Then,
\[ E_N = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} \int f(\delta_{ikl})[\varepsilon + (\lambda_{ik1} - \lambda_{ikl})^\top \Delta \frac{\Delta}{\sqrt{n}} - \varepsilon]f(\varepsilon) d\varepsilon \]

\[ = \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ikl})^\top \Delta \int f(\delta_{ikl})f(\varepsilon) d\varepsilon \]

\[ = \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ikl})^\top \Delta \int [f(\delta_{ikl}) - f(\varepsilon)]f(\varepsilon) d\varepsilon \]

\[ + \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ikl})^\top \Delta \int f^2(\varepsilon) d\varepsilon \]

\[ = E_{N_1} + E_{N_2}. \]

Starting with \( E_{N_1}, \) the distance between \( \delta_{ikl} \) and \( \varepsilon \) is less than or equal to \( |(\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}| \). Furthermore,

\[ |(\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}| = \frac{1}{\sqrt{n}} |(\lambda_{ik1} - \lambda_{ikl})^\top \Delta| \]

\[ \leq \frac{1}{\sqrt{n}} \| |(\lambda_{ik1} - \lambda_{ikl})|| \| \Delta \| \| (\text{by the Cauchy-Schwartz inequality}) \]

\[ \leq \frac{1}{\sqrt{n}} (\| \lambda_{ik1} \| + \| \lambda_{ikl} \|) \| \Delta \| \]

\[ = \frac{1}{\sqrt{n}} 2 \| \Delta \| \to 0. \]

Since \( f \) is assumed uniformly continuous, for \( \varepsilon' > 0 \), there exists \( \delta(\varepsilon') \) so that

\[ |f(\delta_{ikl}) - f(\varepsilon)| < \frac{\sqrt{p} \varepsilon'}{p(p-1) \cdot 2 \| \Delta \|} \text{ if } |\delta_{ikl} - \varepsilon| < \delta(\varepsilon'). \]
It is shown above that $|\delta_{ikl} - \varepsilon| \leq \frac{2||\Delta||}{\sqrt{n}} < \delta(\varepsilon')$ if $n > \frac{4||\Delta||^2}{\delta^2(\varepsilon')} = N(\varepsilon')$. Therefore,

$$|E_{N_1} - 0| = \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=1}^{p} (\lambda_{ik1} - \lambda_{ik1})^T \Delta \int [f(\delta_{ikl}) - f(\varepsilon)]f(\varepsilon)d\varepsilon - 0$$

$$\leq \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=1}^{p} 2||\Delta|| \int |f(\delta_{ikl}) - f(\varepsilon)|f(\varepsilon)d\varepsilon$$

$$\leq \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} p(p - 1)2||\Delta|| \frac{\sqrt{p\varepsilon'}}{p(p - 1)||\Delta||} = \varepsilon'.$$

That is, $\lim_{n \to \infty} E_{N_1} = 0$.

For $E_{N_2}$, it can be shown that $\sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ik1})^T \Delta = 0$ similar to Proof 7.3. Along with $\int f^2(\varepsilon)d\varepsilon$ being finite since $f$ is bounded, it is shown that $E_{N_2} = \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ik1})^T \Delta \int f^2(\varepsilon)d\varepsilon = 0$. Hence, it is shown that $E_{N_1} \to 0$ and $E_{N_2} = 0$, so $E_N = E_{N_1} + E_{N_2} \to 0$.

For $Var(S_n(\Delta) - S_n(0))$,

$$Var(S_n(\Delta) - S_n(0)) = E[(S_n(\Delta) - S_n(0))^2] - E^2[S_n(\Delta) - S_n(0)]$$

$$\leq E[(S_n(\Delta) - S_n(0))^2].$$

So, it is sufficient to show that $E[(S_n(\Delta) - S_n(0))^2] \to 0$ to prove that $S_n(\Delta) - S_n(0) = o_p(1)$.

Let $W_{nk} = \sum_{i=1}^{p} \sum_{l=2}^{p} [I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ik1})^T \Delta) - I(\varepsilon_{ikl} - \varepsilon_{ik1} < 0)]$ and note that $|W_{nk}| \leq 2p(p - 1)$. Then,

$$(S_n(\Delta) - S_n(0))^2 = \frac{1}{p_n} \sum_{k=1}^{n} \sum_{k'=1}^{n} W_{nk} W_{nk'}$$

$$= \frac{1}{p_n} \sum_{k=k'}^{n} W_{nk}^2 + \frac{1}{p_n} \sum_{k \neq k'} W_{nk} W_{nk'} = V_{W_1} + V_{W_2}$$
So, \( E[(S_n(\Delta) - S_n(0))^2] = E(V_{W_1} + V_{W_2}) = E(V_{W_1}) + E(V_{W_2}) \). Consider \( E(V_{W_2}) \) first.

\[
E(V_{W_2}) = \frac{1}{np} \sum_{k \neq k'}^n E(W_{nk}W_{nk'})
\]
\[
= \frac{1}{np} \sum_{k \neq k'}^n E(W_{nk})E(W_{nk'}) \because \text{the \( n \) squares are independent.}
\]

\[
E(V_{W_2}) = \frac{1}{\sqrt{np}} \sum_k E(W_{nk}) \frac{1}{\sqrt{np}} \sum_{k'} (W_{nk'}) \to 0
\]
\[
\because E_N = \frac{1}{\sqrt{np}} \sum_{k=1}^n E(W_{nk}) \to 0
\]

Consider \( E(V_{W_1}) \) next.

\[
E(V_{W_1}) \leq |E(V_{W_1})| = \left| \frac{1}{np} \sum_{k=1}^n E(W_{nk}^2) \right|
\]
\[
\leq \frac{1}{np} \sum_{k=1}^n E(|W_{nk}||W_{nk}|)
\]
\[
\leq \frac{1}{np} \sum_{k=1}^n E(2p(p-1)|W_{nk}|) = \frac{2(p-1)}{n} \sum_{k=1}^n E(|W_{nk}|).
\]

Let us first define the sets \( a_1, a_2, a_3 \) of \( \{i, k, l\} \) such that \((\lambda_{ik1} - \lambda_{ikl})^\top \Delta > 0\), \((\lambda_{ik1} - \lambda_{ikl})^\top \Delta = 0\), and \((\lambda_{ik1} - \lambda_{ikl})^\top \Delta < 0\) respectively. Using these sets, let us discuss \( W_{nk} \) in terms of these three sets, i.e. \( W_{nk} = W_{nk1} + W_{nk2} + W_{nk3} \) where

\[
W_{nk1} = \sum_{a_1} I(0 < \varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}) \text{ almost everywhere,}
\]
\[
W_{nk2} = I(\varepsilon_{ikl} - \varepsilon_{ik1} < 0) - I(\varepsilon_{ikl} - \varepsilon_{ik1} < 0) = 0,
\]
\[
W_{nk3} = -\sum_{a_3} I((\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < 0) \text{ almost everywhere.}
\]
Therefore,

\[
\frac{2(p-1)}{n} \sum_{k=1}^{n} E(|W_{nk}|) = \frac{2(p-1)}{n} \sum_{k=1}^{n} E(|W_{nk1} + W_{nk3}|) \\
\leq \frac{2(p-1)}{n} \sum_{k=1}^{n} E(|W_{nk1}| + |W_{nk3}|) \\
= \frac{2(p-1)}{n} \sum_{k=1}^{n} [E(|W_{nk1}|) + E(|W_{nk3}|)].
\]

Thus, the goal is to show that \(\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk1}|) \to 0 \) and \(\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk3}|) \to 0\). To begin, consider the following:

\[
\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk1}|) = \frac{1}{n} \sum_{k=1}^{n} E(\sum_{a1} I(0 < \varepsilon_{ik1} - \varepsilon_{ikl} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}})) \\
\leq \frac{1}{n} \sum_{k=1}^{n} E(\sum_{a1} I(0 < \varepsilon_{ik1} - \varepsilon_{ikl} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}})) \\
= \frac{1}{n} \sum_{k=1}^{n} E(\sum_{a1} I(0 < \varepsilon_{ik1} - \varepsilon_{ikl} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}})) \\
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a1} [G_{\varepsilon_{1-\varepsilon_{2}}}((\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}}) - G_{\varepsilon_{1-\varepsilon_{2}}}(0)].
\]

By the mean value theorem, \(\frac{G_{\varepsilon_{1-\varepsilon_{2}}}((\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}})-G_{\varepsilon_{1-\varepsilon_{2}}}(0)}{(\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}}} = G'(c_{ikl})\) where 0 < \(c_{ikl} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}}\). Therefore,

\[
\frac{1}{n} \sum_{k=1}^{n} \sum_{a1} [(\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} G'(c_{ikl})] \\
= \frac{1}{n^{1/2}} \sum_{k=1}^{n} \sum_{a1} [(\lambda_{ik1} - \lambda_{ikl})^T \Delta G'(c_{ikl})].
\]
where \((\lambda_{ik1} - \lambda_{ikl})^\top \Delta\) is bounded by \(2||\Delta||\). Considering the distribution function \(G\),

\[
G_{\varepsilon_1 - \varepsilon_2}(c_{ikl}) = P(\varepsilon_1 - \varepsilon_2 \leq c_{ikl}) = P(\varepsilon_1 \leq c_{ikl} + \varepsilon_2)
\]

\[
= E[I(\varepsilon_1 \leq c_{ikl} + \varepsilon_2)|\varepsilon_2]
\]

\[
= E[F(c_{ikl} + \varepsilon_2)]
\]

\[
= \int F(c_{ikl} + \varepsilon_2)f(\varepsilon_2)d\varepsilon_2
\]

\[
\therefore G'_{\varepsilon_1 - \varepsilon_2}(c_{ikl}) = \int f(c_{ikl} + \varepsilon_2)f(\varepsilon_2)d\varepsilon_2.
\]

However, \(f\) is assumed to be bounded, so \(G'(c_{ikl})\) is bounded as well. Therefore,

\[
\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk1}|) \leq \frac{1}{\sqrt{n}} 2||\Delta||B_{G}B_{1} \to 0
\]

where \(B_1\) is a bound for the number of elements in \(a_1\).

Similarly for \(\frac{1}{n} \sum_{k=1}^{n} E(|W_{nk3}|),\)

\[
\frac{1}{n} \sum_{k=1}^{n} E(|\sum_{a_3} I((\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < 0)|)
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} E(|\sum_{a_3} I((\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < 0)|)
\]

\[
\leq \frac{1}{n} \sum_{k=1}^{n} E(|\sum_{a_3} I((\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < 0)|)
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} E(\sum_{a_3} I((\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < 0))
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} E(I((\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < 0))
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} [G_{\varepsilon_1 - \varepsilon_2}(0) - G_{\varepsilon_1 - \varepsilon_2}(\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}}]
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} \sum_{a_3} [- (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} G'(d_{ikl})]
\]
where \((\lambda_{ik1} - \lambda_{ikl})^T \frac{A}{\sqrt{n}} < d_{ikl} < 0\) and \(G_{\varepsilon_{1-\varepsilon_{2}}(0)} - G_{\varepsilon_{1-\varepsilon_{2}}}((\lambda_{ik1} - \lambda_{ikl})^T \frac{A}{\sqrt{n}}) = - (\lambda_{ik1} - \lambda_{ikl})^T \frac{A}{\sqrt{n}} G'(d_{ikl})\) by the mean value theorem. Therefore,

\[
\frac{1}{n} \sum_{k=1}^{n} \sum_{a3} \left[ -(\lambda_{ik1} - \lambda_{ikl})^T \Delta \sqrt{n} G'(d_{ikl}) \right] \leq \frac{1}{\sqrt{n}} 2||\Delta||B_G B_2 \rightarrow 0.
\]

From the above we now have shown that \(2(p-1) \frac{\sum_{k=1}^{n}}{n} E(|W_{nk}|) \rightarrow 0\), and therefore \(E(V W_{1} + V W_{2}) \rightarrow 0\). Therefore, \(Var(S_n(\Delta) - S_n(0)) \rightarrow 0\), completing the proof for condition 1.

For condition 2, we refer to the uniform convergence theorem in Section 7.5 again with \(\tilde{Q}_n(\Delta) = S_n(\Delta) - S_n(0)\). For assumption 1, let \(\hat{Q}_n(\Delta) = E(S_n(\Delta) - S_n(0))\), then the proof for assumption 1 essentially shows this. For assumption 2,

\[
\tilde{Q}_n(\Delta) = \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ikl})^T \Delta \int [f(\delta_{ikl}) - f(\varepsilon)] f(\varepsilon) d\varepsilon
\]

\[
\tilde{Q}_n(\Delta) - \tilde{Q}_n(\tilde{\Delta}) = \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ikl})^T \Delta \int [f(\delta_{ikl}) - f(\varepsilon)] f(\varepsilon) d\varepsilon
\]

\[
- \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} (\lambda_{ik1} - \lambda_{ikl})^T \tilde{\Delta} \int [f(\delta_{ikl}) - f(\varepsilon)] f(\varepsilon) d\varepsilon
\]

\[
= \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} \int [f(\delta_{ikl}) - f(\varepsilon)] f(\varepsilon) d\varepsilon (\lambda_{ik1} - \lambda_{ikl})^T (\Delta - \tilde{\Delta})
\]

\[
\leq \frac{1}{n} \frac{1}{\sqrt{p}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} \int [f(\delta_{ikl}) - f(\varepsilon)] f(\varepsilon) d\varepsilon ||\lambda_{ik1} - \lambda_{ikl}|| \Delta - \tilde{\Delta}||
\]

\[
< \frac{1}{n} \sum_{k=1}^{n} \varepsilon = \varepsilon \quad \text{if } |\Delta - \tilde{\Delta}| < \delta(\varepsilon) = \frac{\varepsilon}{2B_f}
\]

where \(f[\delta_{ikl}] - f(\varepsilon)] f(\varepsilon) d\varepsilon\) is bounded by \(B_f\) and \(||\lambda_{ik1} - \lambda_{ikl}||\) is bounded by 2.

For assumption 3, we first define \(U_n(\tilde{\Delta}, r)\).
\[ |\tilde{Q}_n(\Delta) - \hat{Q}_n(\tilde{\Delta})| = \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} \left( I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{1}{\sqrt{n}} \Delta) \right) \]

\[ - \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{1}{\sqrt{n}} \tilde{\Delta}) | \]

\[ \leq \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{1}{\sqrt{n}} \Delta) \]

Define the sets \( a_1, a_2, \) and \( a_3 \) of \( \{i,k,l\} \) so that

\[ (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} > (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} , \]

\[ (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} = (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} , \]

and \( (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} \) respectively.
Then,

\[
\begin{align*}
\frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} |I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}}) - I(\varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}})| \\
= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_1} I((\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}}) \\
+ \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_3} I((\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} < \varepsilon_{ikl} - \varepsilon_{ik1} < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}}) \\
= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_1} I(0 < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} < |(\lambda_{ik1} - \lambda_{ikl})^T (\Delta - \tilde{\Delta})|) \\
+ \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_3} I(0 < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}} < 0) \\
\leq \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_1} I(0 < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} < \frac{||\lambda_{ik1} - \lambda_{ikl}|| ||\Delta - \tilde{\Delta}||}{\sqrt{n}}) \\
+ \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_3} I(0 < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}} - (\varepsilon_{ikl} - \varepsilon_{ik1}) < \frac{||\lambda_{ik1} - \lambda_{ikl}|| ||\Delta - \tilde{\Delta}||}{\sqrt{n}}) \\
\leq \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_1} I(0 < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) \\
+ \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_3} I(0 < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}} - (\varepsilon_{ikl} - \varepsilon_{ik1}) < \frac{1}{\sqrt{n}} B\lambda r) \\
\leq \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_1} I(-\frac{1}{\sqrt{n}} B\lambda r < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^T \frac{\Delta}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) \\
+ \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{a_3} I(-\frac{1}{\sqrt{n}} B\lambda r < (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) \\
= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{t=2}^{p} I(-\frac{1}{\sqrt{n}} B\lambda r < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^T \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}} B\lambda r) \\
= U_n(\Delta, r)
\end{align*}
\]
Next, we set out to verify assumptions 3.1 and 3.2. For assumption 3.1, let

\[ I(-\frac{1}{\sqrt{n}}B\lambda r < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}}B\lambda r) = I_{ikl}. \]

Then, it is sufficient to show that \( Var(I_{ikl}) \to 0 \) \( \forall i, k, l. \) Consider the following:

\[ Var(I_{ikl}) = E(I_{ikl}^2) - (E(I_{ikl}))^2 \]
\[ = E(I_{ikl})[1 - E(I_{ikl})] \]
\[ \leq E(I_{ikl}) \quad \because |1 - E(I_{ikl})| \leq 1 \]
\[ E(I_{ikl}) = E(I(-\frac{1}{\sqrt{n}}B\lambda r < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} < \frac{1}{\sqrt{n}}B\lambda r)) \]
\[ = \frac{2}{\sqrt{n}}B\lambda r f_{\varepsilon_{ikl} - \varepsilon_{ik1}}(c_{nk\tilde{\Delta}}) \to 0 \quad \text{by the mean value theorem} \]
\[ \text{where } (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} - \frac{1}{\sqrt{n}}B\lambda r < c_{nk\tilde{\Delta}} < (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\tilde{\Delta}}{\sqrt{n}} + \frac{1}{\sqrt{n}}B\lambda r \]
\[ \therefore Var(U_n(\tilde{\Delta}, r)) = o(1) \]

For assumption 3.2,
\begin{align*}
E(U_n(\Delta, r)) &= E\left( \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} I(-\frac{1}{\sqrt{n}} B_{ir} < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \frac{1}{\sqrt{n}} B_{ir}) \right) \\
&= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} E(I(-\frac{1}{\sqrt{n}} B_{ir} < (\varepsilon_{ikl} - \varepsilon_{ik1}) - (\lambda_{ik1} - \lambda_{ikl})^\top \frac{\Delta}{\sqrt{n}} < \frac{1}{\sqrt{n}} B_{ir})) \\
&= \frac{1}{\sqrt{p}} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} \frac{2}{\sqrt{n}} B_{ir} f(c_{nik\Delta}) \\
&= \frac{1}{\sqrt{p}} \frac{2}{\sqrt{n}} B_{ir} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{l=2}^{p} f(c_{nik\Delta}) \leq 2p\sqrt{p}B_{r}B_{fr} \\
\therefore E(U_n(\Delta, r)) &= O(r)
\end{align*}

Therefore, with all assumptions valid, Theorem 1 from Section 7.5 applies here and thus completes the proof.

\section*{7.5 A Uniform Convergence Theorem}

In this section, we present a uniform convergence theorem that is used in Sections 7.3 and 7.4. To begin, let \( \hat{Q}_n(\Delta) \) be a random function (random variables are generic and are not denoted for the sake of notation) that depends on an argument \( \Delta \). Similarly, let \( \bar{Q}_n(\Delta) \) denote a function of \( \Delta \) that typically corresponds to some kind of expectation and/or limit of \( \hat{Q}_n(\Delta) \). Consider the following conditions:

1. \( \hat{Q}_n(\Delta) = \bar{Q}_n(\Delta) + o_p(1) \quad \forall \Delta \) (pointwise convergence),

2. \( \{\bar{Q}_n(\Delta)\}_{n=1}^{\infty} \) is uniformly equicontinuous. That is, \( \forall \Delta, \tilde{\Delta} \) and \( n, \left| \bar{Q}_n(\Delta) - \bar{Q}_n(\tilde{\Delta}) \right| < \varepsilon \) if \( ||\Delta - \tilde{\Delta}|| < \delta(\varepsilon) \), and

3. \( \sup_{||\Delta - \tilde{\Delta}|| < \delta} |\hat{Q}_n(\Delta) - \bar{Q}_n(\tilde{\Delta})| \leq U_n(\tilde{\Delta}, r) \), where

   3.1 \( U_n(\tilde{\Delta}, r) = E[U_n(\tilde{\Delta}, r)] + o_p(1) \quad \forall (\tilde{\Delta}, r) \)
3.2 \( E[U_n(\tilde{\Delta}, r)] = O(r) \) i.e. \( |E[U_n(\tilde{\Delta}, r)]| \leq B_U r \) for \( n > N_U \) where \( B_u \) is free of \( \tilde{\Delta} \) and \( r \).

Our theorem is essentially a hybrid of Theorem 2.1 from Newey (1991) and Theorem 5.1 from Sievers (1983).

**Theorem.** Under conditions 1, 2, and 3,

\[
\sup_{\Delta \in K} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)| = o_p(1)
\]

if \( K \) is a compact set.

Remark: \( K = \{ \Delta : ||\Delta|| \leq r \} \) is compact because it is both closed and bounded.

Proof. Note that \( K \subset \bigcup_{\Delta \in K} \{ \Delta' : ||\Delta' - \Delta|| < r \} \), where \( r > 0 \). Hence, by compactness of \( K \), \( \exists \Delta_1, \Delta_2, \ldots, \Delta_J \) such that \( K \subset \bigcup_{j=1}^{J} \{ \Delta : ||\Delta - \Delta_j|| < r \} \). Let \( K_j = \{ \Delta : ||\Delta - \Delta_j|| < r \} \). Next, choose \( \varepsilon > 0 \) and \( \delta > 0 \) and let \( r = r(\varepsilon) < \min(\delta(\varepsilon/6), \frac{\varepsilon}{12B_u}) \) where \( \delta(\varepsilon/6) \) corresponds to the \( \delta \), \( n \) in assumption 2 and \( \frac{\varepsilon}{12B_u} \) is from assumption 3.2.

We want to show that \( P(\sup_{\Delta \in K} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)| > \varepsilon) < \delta \) if \( n > N(\varepsilon, \delta) \). To begin, note that,

\[
\varepsilon < \sup_{\Delta \in K} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)| \leq \sup_{\Delta \in K} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)| \leq \max_{1 \leq j \leq J} \sup_{\Delta \in K_j} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)|.
\]

Hence,

\[
P(\sup_{\Delta \in K} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)| > \varepsilon) \leq P(\max_{1 \leq j \leq J} \sup_{\Delta \in K_j} |\hat{\Phi}_n(\Delta) - \bar{\Phi}_n(\Delta)| > \varepsilon).
\]
Thus, to prove the result it suffices to show that $\max_{1 \leq j \leq J} \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \tilde{Q}_n(\Delta)| = o_p(1)$. However, since $X_{n_j} = o_p(1)$ for $j = 1, 2, \ldots, J$ (finite) implies that $\max \{X_{n_1}, X_{n_2}, \ldots, X_{n_J}\} = o_p(1)$, we need to only show that $\sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \tilde{Q}_n(\Delta)| = o_p(1)$ for each $j \in \{1, 2, \ldots, J\}$.

So, consider $\sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \tilde{Q}_n(\Delta)|$. Choose $\varepsilon > 0$ and $\delta > 0$. Then,

$$\varepsilon < \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \tilde{Q}_n(\Delta)|$$

$$= \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j) + \hat{Q}_n(\Delta_j) - \tilde{Q}_n(\Delta_j) + \tilde{Q}_n(\Delta_j) - \tilde{Q}_n(\Delta)|$$

$$\leq \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j)| + \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta_j) - \tilde{Q}_n(\Delta_j)| + \sup_{\Delta \in K_j} |\tilde{Q}_n(\Delta) - \tilde{Q}_n(\Delta_j)|$$

$$= \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j)| + |\hat{Q}_n(\Delta_j) - \tilde{Q}_n(\Delta_j)| + \sup_{\Delta \in K_j} |\tilde{Q}_n(\Delta) - \tilde{Q}_n(\Delta_j)|.$$

Thus, since $\varepsilon < X_1 + X_2 + X_3$ implies that at least one $X_i > \frac{\varepsilon}{3}$, it follows that,

$$P\left( \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \tilde{Q}_n(\Delta)| > \varepsilon \right) \leq P\left( \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j)| > \frac{\varepsilon}{3} \right)$$

$$+ P\left( |\hat{Q}_n(\Delta_j) - \tilde{Q}_n(\Delta_j)| > \frac{\varepsilon}{3} \right)$$

$$+ P\left( \sup_{\Delta \in K_j} |\tilde{Q}_n(\Delta) - \tilde{Q}_n(\Delta_j)| > \frac{\varepsilon}{3} \right)$$

$$= P_{n_1} + P_{n_2} + P_{n_3}.$$

Consider $P_{n_3}$ first, and recall that

$$P\left( \sup_{\Delta \in K_j} |\tilde{Q}_n(\Delta) - \tilde{Q}_n(\Delta_j)| > \frac{\varepsilon}{3} \right) = P\left( \sup_{||\Delta - \Delta_j|| < r} |\tilde{Q}_n(\Delta) - \tilde{Q}_n(\Delta_j)| > \frac{\varepsilon}{3} \right)$$

However, by assumption 2.2, $|\tilde{Q}_n(\Delta) - \tilde{Q}_n(\Delta_j)| < \frac{\varepsilon}{6}$ (any arbitrary bound less than
\( \frac{\varepsilon}{3} \) will suffice) as long as \( \| \Delta - \Delta_j \| < \delta(\frac{\varepsilon}{6}) \). By our choice of \( r, r = r(\varepsilon) < \delta(\frac{\varepsilon}{6}) \) so it follows that

\[
\sup_{\| \Delta - \Delta_j \| < r} |\bar{Q}_n(\Delta) - \bar{Q}_n(\Delta_j)| \leq \frac{\varepsilon}{6} = \frac{1}{2}(\frac{\varepsilon}{3}) < \frac{\varepsilon}{3}
\]

Hence, \( P_{n_3} = 0 \).

Consider \( P_{n_2} \) next. By assumption 1, \( |\hat{Q}_n(\Delta_j) - \bar{Q}_n(\Delta_j)| = o_p(1) \forall j \). Thus, \( \exists N_{2j} \) such that for \( n > N_{2j} \), \( P(|\hat{Q}_n(\Delta_j) - \bar{Q}_n(\Delta_j)| > \frac{\varepsilon}{3}) = P_{n_2} < \frac{\delta}{2} \).

Lastly, consider \( P_{n_1} \).

\[
P_{n_1} = P(\sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j)| > \frac{\varepsilon}{3})
\]
\[
= P(\frac{\varepsilon}{3} < \sup_{\Delta \in K_j} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j)|)
\]
\[
= P(\frac{\varepsilon}{3} < \sup_{\| \Delta - \Delta_j \| < r} |\hat{Q}_n(\Delta) - \hat{Q}_n(\Delta_j)|).
\]

By assumption 3, \( P_{n_1} \leq P(\frac{\varepsilon}{3} < U_n(\Delta_j, r)) \)
\[
= P(\frac{\varepsilon}{3} < |U_n(\Delta_j, r)|) \because U_n(\Delta_j, r) \geq 0
\]
\[
= P(\frac{\varepsilon}{3} < |U_n(\Delta_j, r) - E[U_n(\Delta_j, r)] + E[U_n(\Delta_j, r)]|)
\]
\[
\leq P(\frac{\varepsilon}{3} < |U_n(\Delta_j, r) - E[U_n(\Delta_j, r)]| + |E[U_n(\Delta_j, r)]|)
\]
\[
\leq P(|U_n(\Delta_j, r) - E[U_n(\Delta_j, r)]| > \frac{\varepsilon}{6} \cup |E[U_n(\Delta_j, r)]| > \frac{\varepsilon}{6})
\]
\[
\leq P(|U_n(\Delta_j, r) - E[U_n(\Delta_j, r)]| > \frac{\varepsilon}{6}) + P(|E[U_n(\Delta_j, r)]| > \frac{\varepsilon}{6})
\]
\[
= P_{n_1A} + P_{n_1B}.
\]
Consider $P_{n_1 A}$ first. By assumption 3.2, $E[U_n(\Delta_j, r)] = O(r)$ so that $|E[U_n(\Delta_j, r)]| \leq B_U r$ for $n > N_U$. However, $B_U r < \frac{\varepsilon}{12}$ since $r = r(\varepsilon) < \frac{\varepsilon}{12B_U}$. Hence, for our choice of $r$,

$|E[U_n(\Delta_j, r)]| \leq B_U r \leq B_U \left(\frac{\varepsilon}{12B_U}\right) = \frac{\varepsilon}{12} < \frac{\varepsilon}{6}$. Thus, $P_{n_1 A} = 0$.

Consider $P_{n_1 A} = P(|U_n(\Delta_j, r) - E[U_n(\Delta_j, r)]| > \frac{\varepsilon}{6})$ next. By assumption 3.1

$\exists N^*_1 j$ such that for $n > N^*_1 j$, $P_{n_1 A} < \frac{\delta}{2}$. It follows that $P_{n_1 A} < \frac{\delta}{2}$ provided $n > \max(N^*_1 j, N_U) = N_{1j}$. Putting the three pieces together, for $r = r(\varepsilon) < \min(\delta(\frac{\varepsilon}{6}), \frac{\varepsilon}{12B_U})$

and $n > \max(N_{1j}, N_{2j}) = N_{1j}$, $P(\sup_{\Delta \in K_j} |Q_n(\Delta) - \bar{Q}_n(\Delta)| > \varepsilon) \leq P_{n_1} + P_{n_2} + P_{n_3} < \frac{\delta}{2} + \frac{\delta}{2} + 0 = \delta$. 


Bibliography


