12-2015

Rank Based Procedures for Ordered Alternative Models

Yuanyuan Shao
Western Michigan University, shaoyy_mei@yahoo.com

RANK BASED PROCEDURES FOR ORDERED ALTERNATIVE MODELS

by
Yuanyuan Shao

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Statistics
Western Michigan University
December 2015

Doctoral Committee:

Joseph W. McKean, Ph.D., Chair
Jeffery T. Terpstra, Ph.D.
Jung Chao Wang, Ph.D.
Bradley E. Huitema, Ph.D.

RANK BASED PROCEDURES FOR ORDERED ALTERNATIVE MODELS

Yuanyuan Shao, Ph.D.

Western Michigan University, 2015

The ordered alternatives in a one-way layout with k ordered treatment levels are appropriate for many applications, especially in psychology and medicine. There is extensive literature in this area, and many parametric and nonparametric approaches have been introduced. Abelson-Tukey (AT) test is a frequently used parametric method. Its coefficients provide an ideal way of combining means for the purpose of detecting a monotonic relationship between the independent and dependent variables. The AT method, though, is not robust. Furthermore, our initial empirical studies show that it is not more powerful than the Jonckheere-Terpstra (JT) and the Hettmansperger-Norton (HN) nonparametric tests at normal errors for moderate sample sizes. Theses nonparametric tests, unlike the AT test, are not easily extended to general linear and mixed models.

We have developed a rank-based procedure which has the same optimal efficacy properties as the HN procedure for the ordered and umbrella alternative problems, including the unknown peak problem. It is a rank-based procedure and is easily extended to linear, mixed and covariance models. The procedure can utilize general score functions.

ACKNOWLEDGMENTS

I would like to give my deepest thanks to my advisor Dr. Joseph W. McKean for his guidance and patience. I would also express my gratitude to the committee members, Dr. Jung-Chao Wang, Dr. Jeff Terpstra, and Dr. Bradley E. Huitema.

I would also like to thank my husband, Qiang. This dissertation could not be done without his encourgement and support.

Yuanyuan Shao

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This research is on nonparametric methods of testing for k-sample location problem with ordered alternatives. The null hypothesis of interest is that there are no differences in locations (or no treatment effects), under which all k samples can be treated as a sample from one population. The alternatives considered here correspond to a trend in the locations. For example, consider a randomized group experiment with five levels of the independent variable, where the treatments are doses of a drug (say, 10, 20, 30, 40 and 50 mg). With increasing dose level, or concentration the performance of the treatment tends to improve. This is the alternative of interest.

In some cases, the researchers form the research hypothesis as either an increasing or decreasing trend based on the preliminary information they have obtained, and such a hypothesis is known as the ordered alternative hypothesis. To test such a hypothesis, several test procedures are available. When the alternative is that at least two of the k underlying distributions have different centers, the well-known test procedure is the Kruskal and Wallis (Kruskal and Wallis, 1952). When the alternative is that at least one strict inequality follows for all centers, Terpstra (1952) and Jonckheere (1954) were the first, and independently suggesting the same test. Abelson and Tukey (Abelson and Tukey, 1963), Puri (Puri, 1965), Odeh (Odeh, 1971), Archambault et al. (Archambault et al., 1977), Hettmansperger and Norton (Hettmansperger and Norton, 1987), Büning (Büning, 1999), and McKean et al.(McKean et al., 2001) have also developed methods for these hypotheses.

When the treatment effect has a monotonic increasing trend and changes in direction after reaching a peak, it is called an umbrella trend test. Mack and Wolfe (Mack and Wolfe, 1981), Hettmansperger and Norton (Hettmansperger and Norton, 1987), Shi (Shi, 1988), Chen and Wolfe (Chen and Wolfe, 1990), Chen (Chen, 1991), Pan (Pan, 1996), Terpstra and Magel (Terpstra and Magel, 2003), Kössler (Kössler, 2006) and Alvo (Alvo, 2008) discuss nonparametric tests for umbrella alternatives.

In chapter 2, Monte Carlo simulation results of an initial study of the α levels and powers of some of these existing methods are discussed as motivation for this study. In chapter 3, two new distribution-free tests and a parametric-bsed test are proposed, and their properties are developed. A simulation study involving these new procedures, comparing them with those mentioned in chapter 2, is then discussed. In chapter 4, new methods of unknown peak problems are proposed. In chapter 5, new methods for mixed models with random block effects are proposed and simulation results are discussed.

1.1 Overview of k-sample Location Problem

1.1.1 Basic One-way Notation

Let $Y_{1j},\ Y_{2j}, \dots,\ Y_{n_j j},\ j = 1, \dots,\ k$ represent a random sample of the response of the jth group with sample size n_j having common cumulative distribution function (cdf) $F_j(y)$, where $F_j(y) = F(y - \theta_j)$, where F is an absolutely continuous distribution function. Assume that θ_j ($j = 1, \dots,\ k$) is a location parameter for the population. In the following, we assume that F is twice continuously differentiable on $(-\infty,\ \infty)$. There are $n = \sum_{j=1}^{k} n_j$ observations in all. All responses are assumed to be independent of one another. Hence our full model design is the same as the one-way analysis of variance design, that is, the response Y_{ij} follows the linear model,

$$Y_{ij} = \theta_j + \epsilon_{ij},\ j = 1, \dots,\ k,\ i = 1, \dots,\ n_j, \tag{1.1}$$

Y_{ij} is the ith observation on response variable for the jth treatment level; θ_j is the jth location parameter; ϵ_{ij} are independent and identically distributed (iid) with continuous distribution function $F(x)$ and probability density function $f(x)$. The null hypothesis of interest is that there is no differences in locations among k populations, or there is no difference in treatment effects, and we write it as

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k. \tag{1.2}$$

Under the null hypothesis all k samples can be treated as one sample from a single population.

1.1.2 Hypotheses

In this thesis we are interested in several alternatives which are given by the following.

Ordered Alternatives

According to the natural labeling of the treatments, the appropriate alternatives could be set as increasing (or decreasing) treatment effects. The monotone alternative hypotheses are of the form:

$$H_1 : \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k, \text{ (with at least one strict inequality)}, \tag{1.3}$$

or,

$$H_2 : \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k, \text{ (with at least one strict inequality)}. \tag{1.4}$$

Both sets of these alternatives are called ordered alternatives.

Umbrella Alternatives

In some experiments, the performance of the treatment might tend to increase with increasing levels, but after some point higher levels tend to diminish the performance. This is called an umbrella pattern and can be written as:

$$H_3 : \theta_1 \leq \theta_2 \leq \cdots \leq \theta_p \geq \theta_{p+1} \geq \cdots \geq \theta_k. \tag{1.5}$$

This is an umbrella alternative with peak p, where $p \in \{1, 2, \ldots, k\}$. Thus the ordered alternatives 1.3 are special umbrella alternatives with peak at k.

1.2 Details on Some Existing Methods

Several methods have been proposed for ordered alternatives, including the Terpstra-Jonckheere test which was proposed by Terpstra (1952) and Jonckheere (1954); the Abelson-Tukey test which was proposed by Abelson and Tukey (1963); the Hettmansperger and Norton procedure (1987); and the bootstrap based Spearman approach which was proposed by McKean, Naranjo and Huitema (2001).

These tests are classified into nonparametric (distribution free) and parametric tests.

We next describe these tests.

1.2.1 Distribution-free Nonparametric Tests

Jonckheere-Terpstra(JT) Test

Jonckheere (Jonckheere, 1954) and Terpstra (Terpstra, 1952) were the first to consider ordered alternative rank tests. They independently proposed the following procedure.

The Jonckheere-Terpstra statistic, J is

$$J = \sum_{u=1}^{v-1} \sum_{v=2}^{k} U_{uv},$$

where

$$U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \phi(Y_{iu}, Y_{jv}), 1 \leq u < v \leq k,$$

and $\phi(a, b) = 1$ if $a < b$, 0 otherwise. Note that U_{uv} is the Mann-Whitney statistic (Mann and Whitney, 1947) for testing the difference between populations u and v. It can be shown that the JT statistics can be described asymptotically as a linear combination of Chernoff and Savage (Chernoff and Savage, 1958) statistics.

For a specified level of significance α, the decision rule is,

Reject H_0 (1.2) in favor of H_A (1.3)
if $J \geq j_\alpha$; otherwise do not reject,

where the constant j_α is chosen to make the type I error probability equal to α.

Under H_0, J is distribution-free. Although some tables for the null distribution exist, usually the asymptotic test is used. Under H_0 the expected value and variance of J are

$$\mathrm{E}_0(J) = \frac{n^2 - \sum_{j=1}^{k} n_j^2}{4}$$

$$\mathrm{var}_0(J) = \frac{n^2(2n+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)}{72}.$$

The standardized version of J is

$$J^* = \frac{J - \mathrm{E}_0(J)}{\sqrt{\mathrm{var}_0(J)}} = \frac{J - \left[\frac{n^2 - \sum_{j=1}^{k} n_j^2}{4}\right]}{\left[\frac{n^2(2n+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)}{72}\right]^{1/2}}.$$

When H_0 is true and as $\min(n_1,\ n_2, \ldots,\ n_k)$ tends to infinity, J^* has an asymptotic N(0, 1) distribution (Hollander and Wolfe, 1999).

Thus, the asymptotic test is,

Reject H_0 (1.2) in favor of H_A (1.3)
if $J^* \geq z_\alpha$; otherwise do not reject

The Spearman (SP) Test

The Spearman correlation coefficient (Spearman, 1904) (r_s) is well known in psychology. For the ordered alternative problems, the Spearman rho (ρ_s) statistic is used to test the correlation between the ordered levels of the treatment and the dependent variable (Y_{ij}s). If the null hypothesis (1.2) is true, then the population Spearman correlation $\rho_s = 0$; if the ordered alternative hypothesis (1.3 or 1.4) is true, then $\rho_s \neq 0$

(McKean et al., 2001). The Spearman rank correlation coefficient is defined by

$$\begin{aligned} r_s &= \frac{12\sum_{i=1}^{n}\left[(R_i - \frac{n+1}{2})(S_i - \frac{n+1}{2})\right]}{n(n^2-1)} \\ &= 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}, \end{aligned}$$

where R_i denotes the rank of X_i among all X; S_i denotes the rank of Y_i among all Y; and $D_i = S_i - R_i$, for $i = 1, \ldots, n$.

Under H_0, the expected value and variance of r_s are

$$\mathrm{E}_0(r_s) = 0,$$

and

$$\mathrm{var}_0(r_s) = \frac{1}{n-1},$$

respectively. The standardized version of r_s is

$$r_s^* = \frac{r_s}{\{\mathrm{var}_0(r_s)\}^{1/2}} = (n-1)^{1/2} r_s.$$

When H_0 is true, r_s^* has an asymptotic N(0, 1) distribution (Hollander and Wolfe, 1999).

The asymptotic test is

Reject H_0 (1.2) in favor of H_A (1.3)

if $r_s^* \geq z_\alpha$; otherwise do not reject.

The Bootstrap Spearman Test (BSP)

This procedure uses bootstrap methodology to independently sample n observations with replacement from the observed data Y, and randomly assign these observations in k groups. The Spearman correlation coefficient is then computed for this bootstrap sample. After sampling B (# of bootstraps, which should be large, 10,000 or more) times, the proportion of the sampled Spearman correlation coefficient that are equal to or greater than the observed Spearman correlation coefficient is the observed one-tailed p value (McKean et al., 2001). The statistic used in the bootstrap Spearman test (BPS) is

$$\bar{R}_n = \sum_{i=1}^{k} i\bar{R}_{in}.$$

This is Spearman's rho statistic (Tyron and Hettmansperger, 1973; Kendall and Stuart, 1961), where $\bar{R}_{in}$ is the average of the ranks of the items in the ith sample. And $\bar{R}_n$ can be written as

$$\bar{R}_n = \frac{1}{n}\sum_{u=1}^{v-1}\sum_{v=2}^{k}(v-u)U_{uv} + \frac{k(n+1)(k+1)}{4}.$$

The BPS procedure provides:

1. A test for ordered alternatives to the null hypothesis of equal population medians,
2. A point estimate of the magnitude of monotonic association between the ordered treatment levels and the dependent variable,
3. A confidence interval for the measure of the magnitude of association.

Note that the confidence interval is obtained from a second bootstrap which is resampled within each of the original samples.

Hettmansperger and Norton (HN) Test

The Hettmansperger and Norton (HN) test is based on a linear combination of two-sample Chernoff-Savage type statistics computed from $k(k-1)/2$ pairs of samples. The optimal weights are obtained so as to maximize the Pitman efficacy such that this procedure regarded the tests based on linear contrasts and derived the most efficient rank test for a given pattern, as discussed below (Hettmansperger and Norton, 1987). The test statistic is

$$V^* = \left(\frac{12}{n+1}\right)^{1/2}\frac{V}{\left(\sum a_j^2/\lambda_j\right)^{1/2}},$$

where $V = \sum_j a_j\bar{R}_j$, $\bar{R}_j = \frac{1}{n_j}\sum_j R_{ij}$, R_{ij} is the rank of Y_{ij} in the combined samples, $\lambda_j = \frac{n_j}{n}$.

The Pitman efficacy of the test is maximized by $a_j = \lambda_j(c_j - \bar{c}_w)$, where c_j is a set of constants, which specifies the pattern of the alternative, $\bar{c}_w = \sum \lambda_j c_j$, and

$\sum_j a_j = 0$. Often $c_j = j$ is used, which is optimal under the alternative of an increasing trend and equally spaced means.

If $H_0 : \theta_1 = \cdots = \theta_k$ is true,

$$\mathrm{E}_0(V) = 0$$

$$\mathrm{Var}_0(V) = [(n+1)/12] \sum \lambda_i (c_j - \bar{c_w})^2,$$

(Hettmansperger and Norton, 1987). The asymptotic test is

Reject H_0 (1.2) in favor of H_A (1.3)
if $V^* \geq z_\alpha$; otherwise do not reject.

They also provided statistics that are based on pairwise ranking and are equivalent in a Pitman efficacy sense to V^*. The statistic is

$$U^* = \sum \sum_{i<j} (d_j - d_i) W_{ij},$$

where $d_j = \lambda_j (c_j - \bar{c}_w)$, $W_{ij} = n_i^{-1} n_j^{-1} U_{ij}$. Under the null hypothesis $H_0 : \theta_1 = \cdots = \theta_k$, $\mathrm{E}_0(U^*) = 1/2 \sum \sum_{i<j} (d_j - d_i)$ and

$$\mathrm{var}_0(U^*) = \frac{1}{12} \left\{ \sum \sum_{i<j} \frac{(d_j - d_i)^2}{n_i n_j} + k \sum \sum_{i<j} \left(\frac{d_j}{n_j} - \frac{d_i}{n_i}\right)(d_j - d_i) \right\}.$$

The asymptotic test is

Reject H_0 (1.2) in favor of H_A (1.3)
if $(U^* - \mathrm{E}_0(U^*))\mathrm{var}^{-1/2}(U^*) \geq z_\alpha$; otherwise do not reject.

1.2.2 Parametric-based Test

Abelson-Tukey (AT) Test

This is a frequently used parametric method. The contrast coefficients provide an ideal way of combing means for the purpose of detecting a monotonic relationship between the independent and dependent variables (Abelson and Tukey, 1963). The h_j were selected in an attempt to maximize the minimum correlation between h_j and the unknown X_j (dose for the jth group), considering all possibilities under the monotonicity assumption.

The test statistic is defined as

$$t = \frac{\sum_{j=1}^{k} h_j \bar{Y}_j}{\sqrt{MSE\left[\sum_{j=1}^{k} \frac{h_j^2}{n_j}\right]}},$$

where

1. h_j is the Abelson-Tukey maximin contrast coefficient associated with treatment j, $h_j = \sqrt{(j-1)[1-(j-1)/k]} - \sqrt{j(1-j/k)}$;
2. $\bar{Y}_j$ is the jth sample mean;
3. MSE is the within-group mean square from the ANOVA (one-way least squares (LS) fit) on Y;
4. n_j is the sample size associated with the jth treatment; and
5. t is the test statistic.

The AT test selects the h_js so that the minimum value of the correlation coefficient is compatible with the restrictions on X_js are maximized, where $\sum h_j = 0$, and provides a standardized effect size measure for the maxmin contrast. If we write $\boldsymbol{\theta} = \boldsymbol{\alpha} + \boldsymbol{X\beta}$, then the problem is the same as to detect $\boldsymbol{\beta}$. Thus, we can use the linear model to test any contrasts of the θs. It's easy to extend the test to a single covariate, or multiple covariates and mixed effects cases, and the test will stay same.

However, the AT test is based on a least squares fit, thus, its result is not robust when the underlying distribution of the errors is far from normal distribution (either skewed or heavy tailed). This motivates us to propose a new method, using a rank-based fit and having all the advantages of the AT test.

1.3 Examples

Hundal (1969) described a study on the effects of knowledge of performance on workers whose job was a repetitive task. It was thought that with increased degree of knowledge of performance, the output would be increased. Eighteen male workers were selected and randomly divided into three groups: A (control group), B (treatment group) and C (treatment group). Workers in group A received no information about their performance. Workers in group B received some information about their performance. Workers in group C received detailed information about their performance. The responses indicated were the number of parts each worker processed in the experimental period. Note that to show the robustness of nonparametric tests, the third value in group B has been changed from 54 to 74.

Table 1: Example: Number of parts produced

group A	group B	group C
40	38	48
35	40	40
38	74	45
43	44	43
44	40	46
41	42	44

Let μ_i denote the mean output of group i, then the alternative hypothesis is $H_A : \mu_A \leq \mu_B \leq \mu_C$ with at least one strict inequality. Figure 1 shows the boxplot of three groups of data. There is an outlier in group B. The test statistics and p values of all mentioned methods are provided in Table 2. The AT test result is affected by the outlier. All methods reject the null hypothesis except the AT test with a p-value 0.1989.

Table 2: Test results

methond	test statistic	p-value
JT	78	0.0251
SP	2.2046	0.0212
BSP	0.4827	0.0220
HN	1.9737	0.0242
AT	0.8703	0.1989

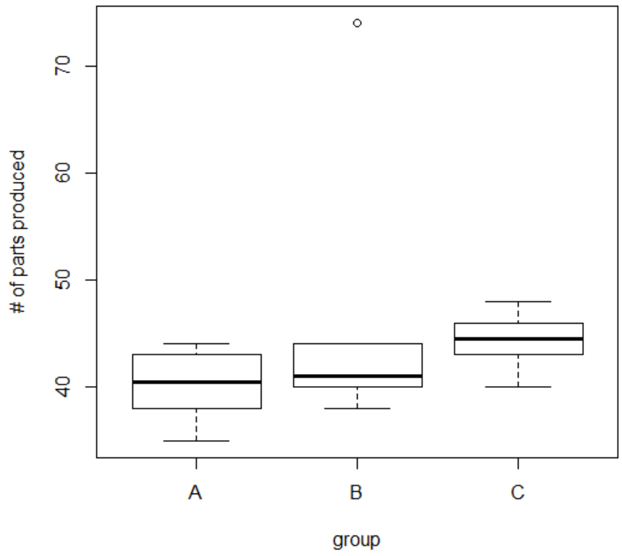

Figure 1: Boxplot of example data

CHAPTER 2

MOTIVATION

A Monte Carlo simulation method was conducted to compare the performance of the JT, SP, HN and AT procedures. The intent of the study is to compare these methods over a wide range of tail weights for the random error distributions. The normal distribution and Student t-distribution with degrees of freedom 10, 8, 5, 3, 2, and 1 (Cauchy) were selected. This is a range from light to very heavy tailed distributions. A skewed contaminated normal distribution was also used which is a mixing of a standard normal distribution with a normal distribution with mean 1 and variance 100, and with the contamination rate at 0.25. The procedures are location and scale equivariant, so standard (convenient) forms of the distributions are used. For this initial investigation we set $k = 5$ and used group sample sizes of 5, so there were 25 subjects.

A situation consists of a design, a hypothesis and an error distribution. For each situation we ran 10,000 simulations.

Initial study results are shown in section one. A new method, Robust Abelson-Tukey (RAT) test, is proposed in section two. All the simulation results are discussed in section three.

2.1 Initial Study Results

In the first part, the empirical α levels are obtained by repeating this process 10,000 times under the null hypothesis, and computing the proportions of all the p-values, which are less than or equal to the chosen alpha level (.05 and .1) on each procedure.

In the second part, the empirical power values are obtained by repeating the process, similar as in section one, but under the specified alternative hypothesis, 10,000 times. Suppose the center of each group is equally spaced, thus the only difference is an estimate of spacing between the centers being added to the data set of each group.

2.1.1 Empirical α Levels

Table 3 shows the simulation results of the empirical αs at norminal levels 10% and 5% under different population distributions.

Since SP, JT and HN are distribution free, this null simulation is not needed. The AT, though, is not distribution free; therefore, the investigation of its validity over these situations is of interest. We also included the distribution-free procedures as a confirmation, to check the αs on the simulation and to identify the liberal and conservative tests by power.

As the underlying distribution changed from standard normal distribution to heavy tailed t-distribution with 1 degree of freedom, the empirical α for all tests are stable except the AT test. Both αs (.1217 and .0425) under $t(1)$ are out of the 95% confidence intervals (.0941, .1059) and (.0457, .0543).

Table 3: Initial simulation on empirical αs.

	$N(0,1)$		$t(10)$		$t(5)$		$t(2)$		$t(1)$	
Meth.	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
JT	.0982	.0469	.1049	.0501	.1047	.0490	.1093	.0521	.1017	.0489
SP	.0937	.0486	.1004	.0518	.0993	.0494	.1030	.0542	.0966	.0492
AT	.0961	.0473	.1001	.0501	.1004	.0507	.1095	.0534	***.1217***	***.0425***
HN	.0973	.0486	.1038	.0518	.1034	.0494	.1070	.0542	.1014	.0492

2.1.2 Empirical Powers

For each simulation, n values ($\boldsymbol{Z}$) were generated from a specific distribution, and n_j of them ($\boldsymbol{Z}_j$) were randomly assigned to level j of the treatment. Suppose the center of the adjacent group is equally spaced by d. To specify the d value, the noncentrality parameter (δ) needs to first be found.

As in the null case, we set $k = 5$, $n_i = 5$, $n = 25$ and $\alpha = 0.05$. To determine d

select $p_0 = 0.15$. By solving the power expression

$$p_0 = P\left\{F(\delta,\ k-1,\ n-k) \geq F(1-\alpha,\ k-1,\ n-k)\right\},$$

we obtain the noncentrality parameter δ which is approximately 2.11.

Let X be the one-way ANOVA cell mean design matrix,

$$\boldsymbol{X}_{(n\times k)} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_k} \end{bmatrix},$$

and A be the matrix which consider group 1 as the reference group,

$$\boldsymbol{A}_{(k-1\times k)} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix},$$

and

$$\boldsymbol{\beta} = \boldsymbol{\mu} = (\mu_1 \dots \mu_k)' = (\mu_1\ \ \mu_1 + d\ \ \dots\ \ \mu_1 + (k-1)\,d)'.$$

We know that

$$\delta = \frac{1}{\sigma^2}\,(\boldsymbol{A\mu})'\left[\boldsymbol{A}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{A}'\right]^{-1}(\boldsymbol{A\mu}).$$

When $n_i = n_j$, after simplifying

$$\begin{aligned} \delta &= \frac{n_i}{k\sigma^2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\mu_i-\mu_j)^2 \\ &= \frac{n_i}{k\sigma^2}\left[(k-1)1^2 + (k-2)2^2 + (k-3)3^2 + \dots + 1(k-1)^2\right]d^2, \end{aligned}$$

and setting $\sigma^2 = 1$ and $\mu_1 = 0$, we obtain $d = \sqrt{2.11/50}$.

Under the null hypothesis, all data are generated from the same distribution. While under the alternative hypothesis, all data are generated from the same distribution and then shifted in location. For example, the data for the jth group is

$$\boldsymbol{y}_j = \boldsymbol{Z}_j + \boldsymbol{b}_j, \text{ and } \boldsymbol{Y} = (\boldsymbol{y}_1 \ \boldsymbol{y}_2 \ \ldots \ \boldsymbol{y}_k)',$$

where the $\boldsymbol{b}_j$s represents the shifts of group j from group 1, and the values of $\boldsymbol{b}_j$s are $\mathbf{0}$, $\boldsymbol{d}$, $\mathbf{2}\boldsymbol{d}$, $\mathbf{3}\boldsymbol{d}$, $\mathbf{4}\boldsymbol{d}$ ... respectively, for $j = 1, \ \ldots, \ k$.

Table 4 and Figure 2 show the empirical power and power curves when the errors have a normal distribution. For all tests, there is no significant difference among the empirical powers. Note that even though the errors are normally distributed, the AT test does not dominate the JT and HN tests.

Table 4: Initial simulation on empirical power under the $Normal$ distribution.

Meth.	Empirical power		
	0	2.11	8.44
JT	.0982	.5488	.9313
SP	.0937	.5428	.9283
AT	.0961	.5404	.9305
HN	.0973	.5492	.9313

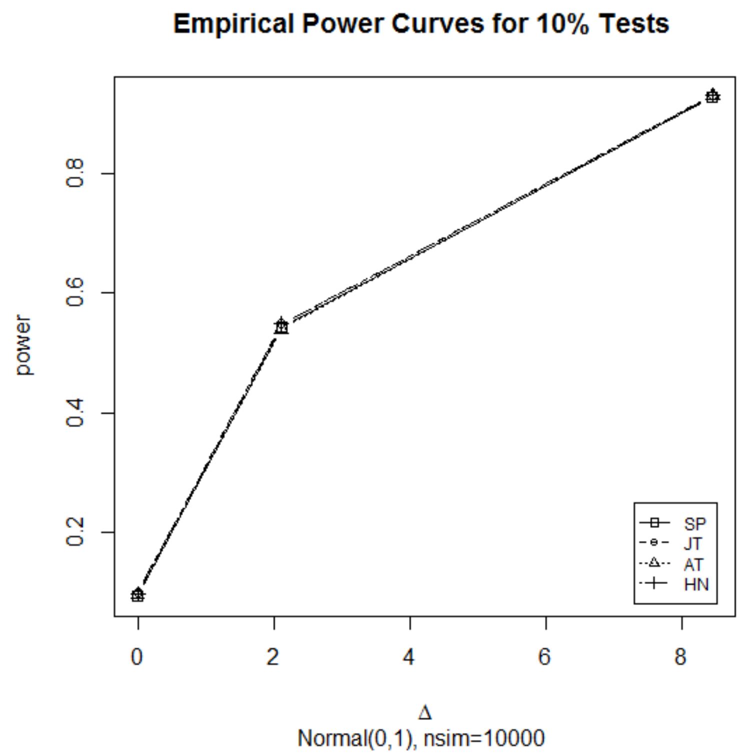

Figure 2: Plot of the initial study on power curves under a $Normal$ distribution.

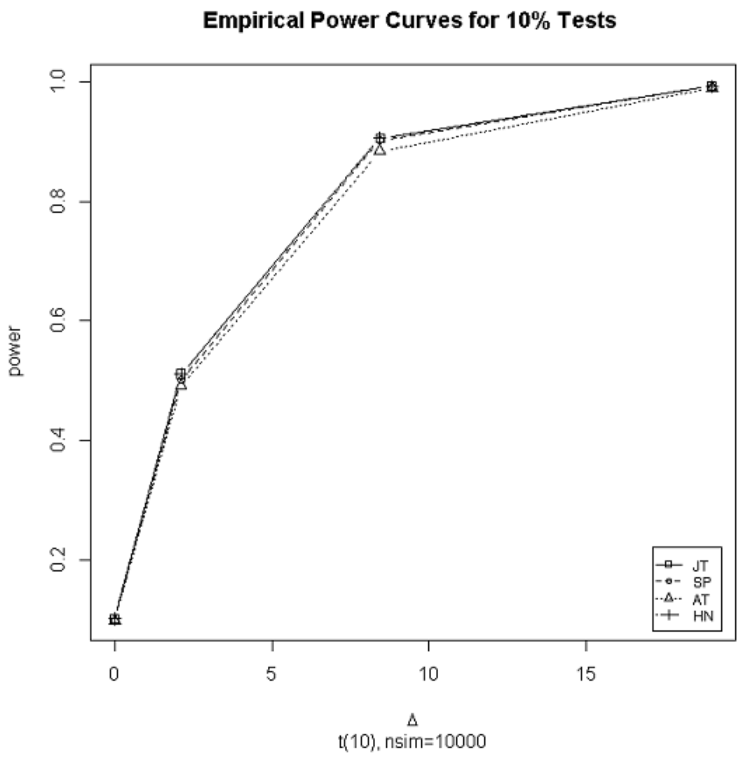

Figure 3: Plot of the initial study on power curves under $t_{(10)}$.

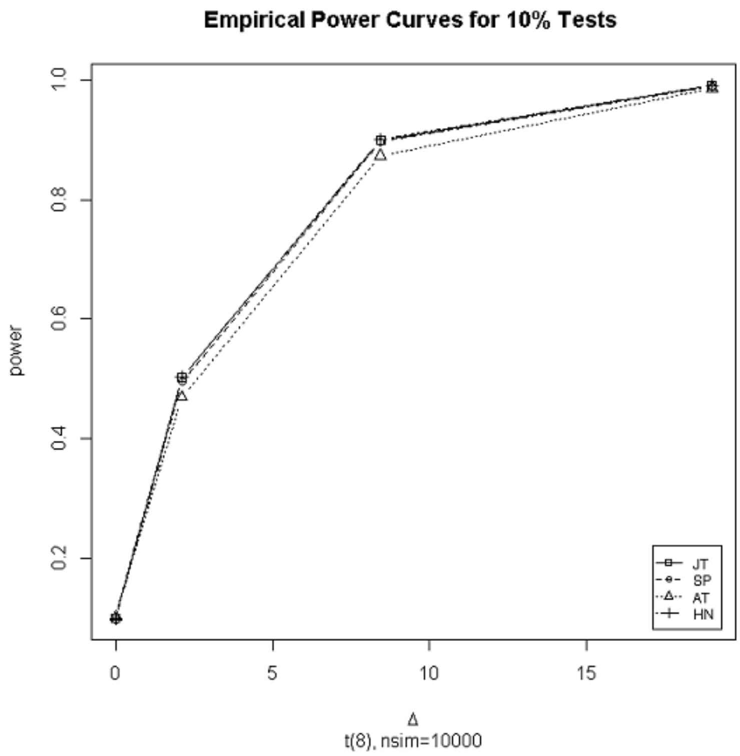

Figure 4: Plot of the initial study on power curves under $t_{(8)}$.

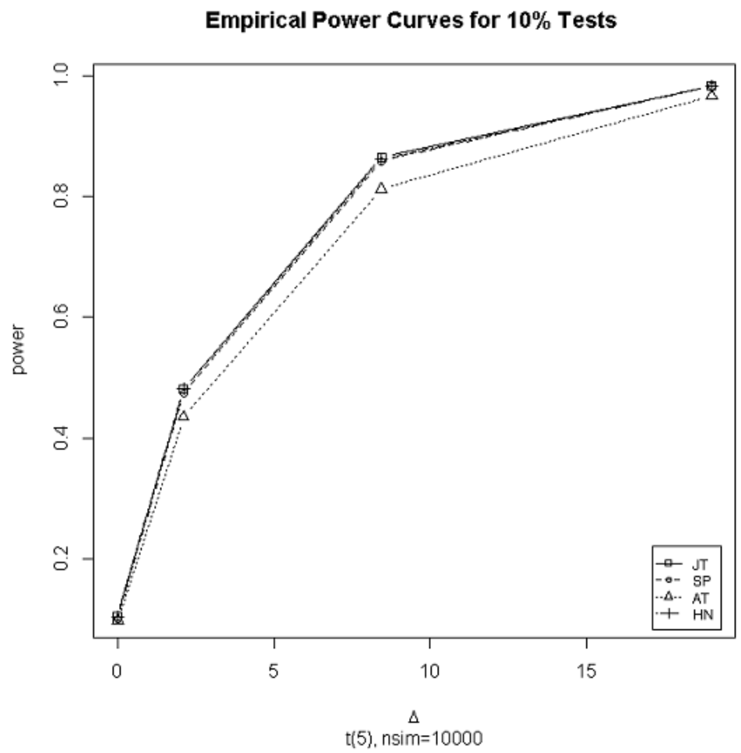

Figure 5: Plot of the initial study on power curves under $t_{(5)}$.

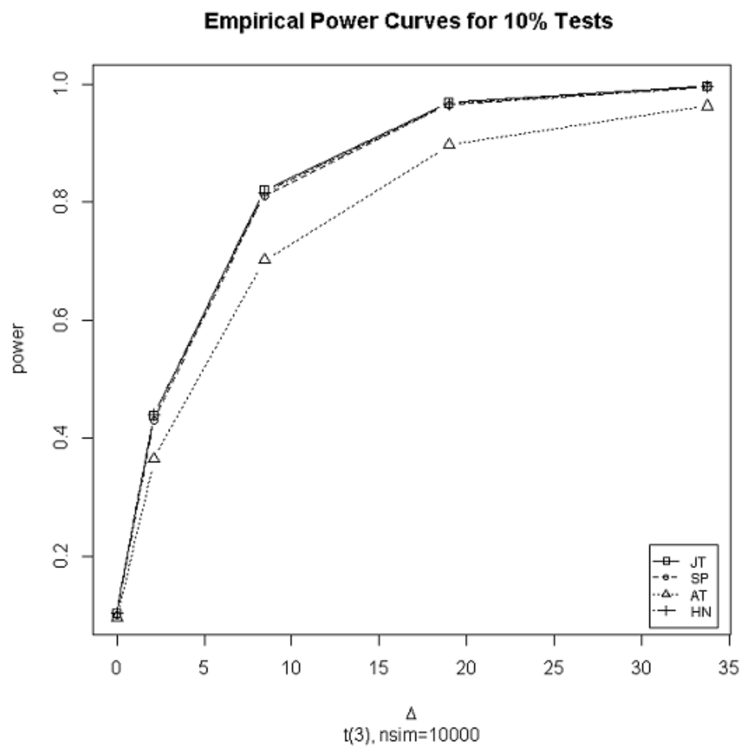

Figure 6: Plot of the initial study on power curves under $t_{(3)}$.

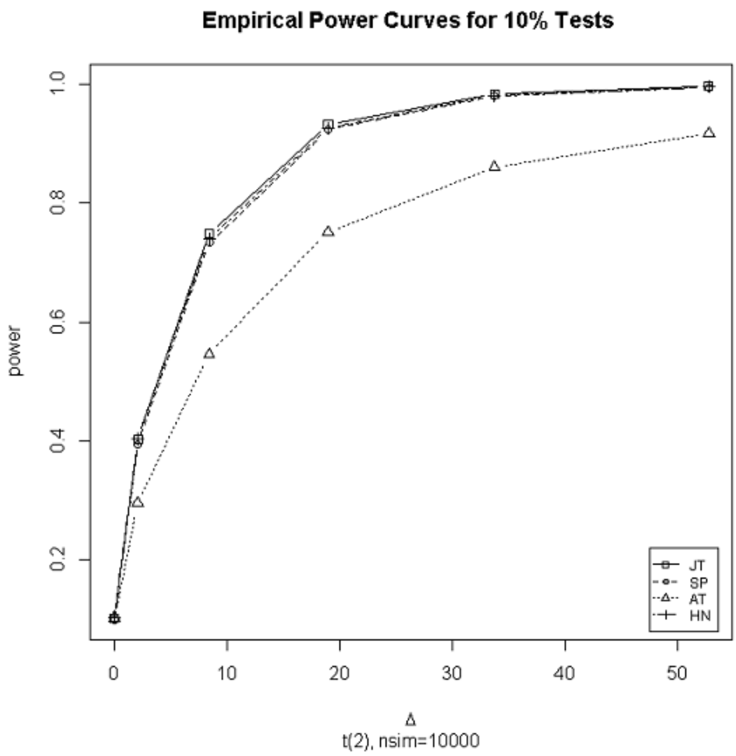

Figure 7: Plot of the initial study on power curves under $t_{(2)}$.

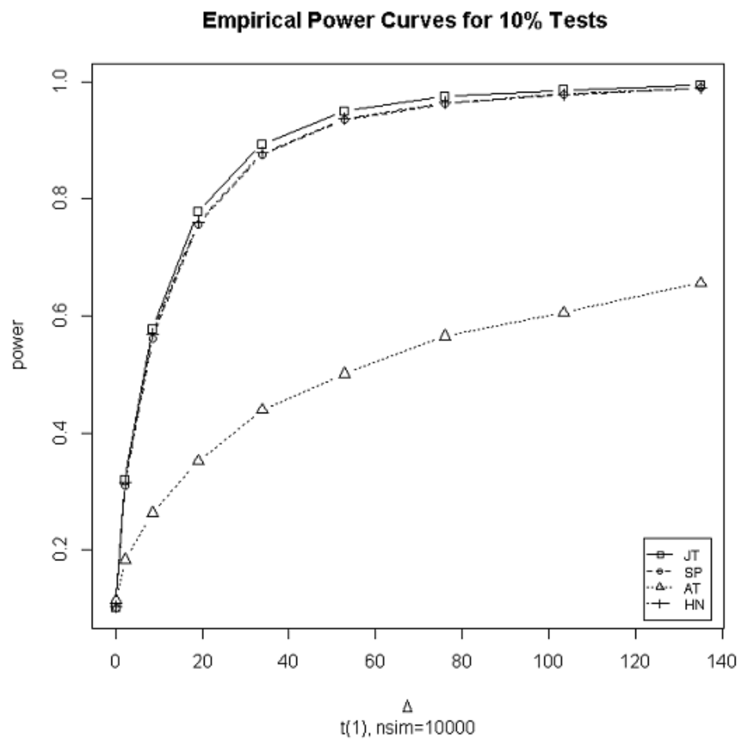

Figure 8: Plot of the initial study on power curves under $t_{(1)}$.

Figures 3 to 8 show the empirical power curves under the t distributions with degrees of freedom 10, 8, 5, 3, 2, 1, respectively. As the tails of the distribution get heavier, the AT test becomes more and more instable. All distribution free tests are stable. Figure 8 shows that with $t(1)$ error distribution, the JT test is more stable than the HN and SP tests, but there is no significant difference among them.

Table 5: Initial simulation on the empirical power under SCN distribution.

Meth.	Empirical power					
	0	2.11	8.44	18.99	33.76	52.75
SP	.0995	.3227	.5920	.7783	.8660	.9149
JT	.1051	.3324	.6107	.8034	.8959	.9424
AT	.0952	.1517	.2126	.2893	.3715	.4536
HN	.1034	.3292	.6013	.7830	.8702	.9176

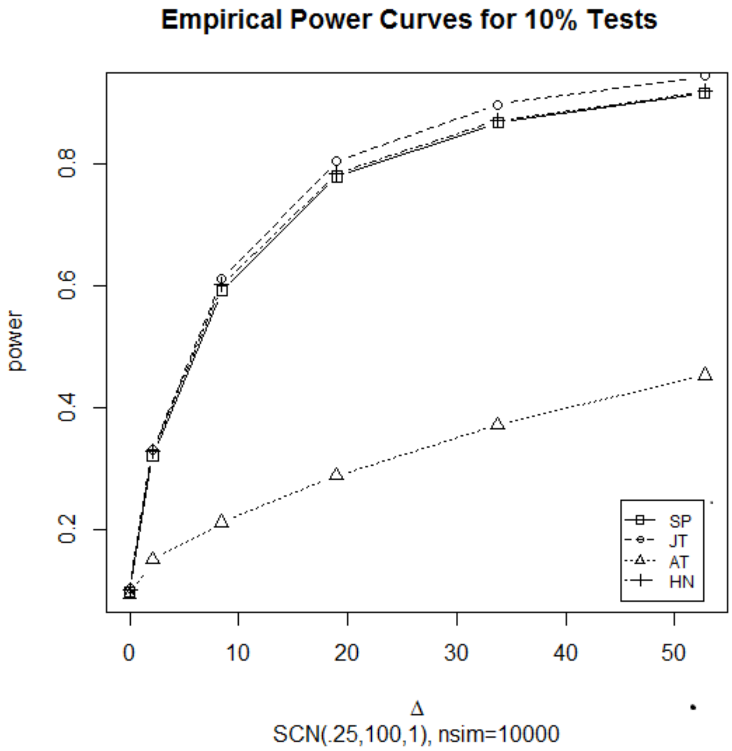

Figure 9: Plot of the initial study on power curves under SCN distribution.

Table 5 and Figure 9 show the empirical power curves under the skewed contaminated normal distribution $SCN(.25,\ 100,\ 1)$; the power of AT is quite low compared to other procedures. Statistically speaking, consider the standard two-sample, paired data, confidence interval for a difference in proportions. By using

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{[\hat{p}_1 + \hat{p}_2 - (\hat{p}_1 - \hat{p}_2)^2]/n},$$

where n is the number of simulations, 10,000. This gives a confidence interval of the proportion difference at level α, to compare the methods. For our study, we used the last column empirical powers in table 5 with the JT, HN and SP tests having significantly higher power than the AT test with 95% confidence intervals (.4677, .5099), (.4429, .4851) and (.4402, .4824), respectively. The JT test does not have significantly higher power than the HN test with a 95% confidence interval (-.0019, .0515) and slightly higher power than the SP test with (.0008, .0542). The HN and SP tests are similar to each other with a 95% confidence interval (-.0238, .0292).

Based on all simulation results, the AT test is not as robust as other nonparametric approaches when the underlying error distribution has heavy tails. The first motivation of this thesis is to robustify the AT approach.

2.2 Robust Abelson-Tukey (RAT) Test

2.2.1 Rank-based Estimator

A general linear model is of the form

$$\boldsymbol{Y} = \boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{2.1}$$

where $\boldsymbol{Y}$ is the $n \times 1$ vector of responses, $\boldsymbol{X}$ is the $n \times (k-1)$ design matrix, and $\boldsymbol{e}$ is the $n \times 1$ vector of error terms. The least squares estimator minimizes the Euclidean distance between $\boldsymbol{Y}$ and $\hat{\boldsymbol{Y}}$, the predicted value of $\boldsymbol{Y}$. For the robust Abelson-Tukey RAT procedure, a rank-based estimator, a different measure of distance based on the dispersion function of Jaeckel (Jaeckel, 1972; Jureckova, 1971) is used to substitute the LS estimator. The only assumption on the distribution of the errors required is that it is continuous.

As discussed in chapter 3 of Hettmansperger and McKean, the rank-based estimator of $\boldsymbol{\beta}$ is defined as

$$\hat{\boldsymbol{\beta}} = \text{Argmin} \left\| \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} \right\|_{\varphi},$$

where $\|\cdot\|_{\varphi}$ is a pseudo-norm defined as

$$\|u\|_{\varphi} = \sum_{i=1}^{n} a(R(u_i))u_i,$$

where $R(u_i)$ denotes the rank of u_i, $a(t)$ are scores such that $a(1) \leq \cdots \leq a(n)$ and

$$a(i) = \varphi\left(\frac{i}{n+1}\right),$$

where φ is a non-decreasing, square-integrable score function defined on the interval $(0, 1)$. Assume without loss of generality that it is standardized, so that

$$\int \varphi(u)du = 0 \text{ and } \int \varphi^2(u)du = 1.$$

Kloke and McKean (2012) have developed a R package, Rfit, for rank-based (R) estimation and inference for linear models; see also (Kloke and McKean, 2014). In Rfit the default option is to use Wilcoxon (linear) scores. Furthermore, Rfit also includes a variaty of score functions, and user-defined score functions (Kloke and McKean, 2012) are easily added. The rank-based fit depends on a selection of score functions, which we discuss next.

2.2.2 General Rank Scores

To obtain efficient statistics, the density is assumed to have finite Fisher information. The Fisher information is given by $I(f) = \int_0^1 \varphi_f^2(u)du$, where

$$\varphi_f(u) = -\frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)} \tag{2.2}$$

is the optimal score function.

When the errors come from different distributions, appropriate scores are recommended. Generally used scores include the following.

1. Wilcoxon scores are recommended when the errors come from a moderate tailed

distribution, and it is generated by the linear function $\varphi_R(u) = \sqrt{12}(u - 1/2)$. Wilcoxon scores are optimal if the error distribution is logistic. In Rfit, Wilcoxon scores can be set by `scores=wscores`.

2. Normal scores are recommended when the errors come from a light-moderated tailed distribution. Normal scores are optimal for normal error distribution. In Rfit, normal scores can be set by `scores=nscores`.

3. There are four types of scores in the family of bent (Winsorized Wilcoxons) scores. They are optimal for error distribution with a logistic center and tails of exponential order(McKean et al., 1989).

   (a) Bent1 scores are recommended when the errors come from a highly right skewed distribution, and the score function is defined in terms of 3 parameters. In Rfit, they can be set by `scores=bentscores1`.

$$\varphi(u) = \begin{cases} \frac{(s_3-s_2)u}{s_1} & \text{if } u < s_1 \\ s_3 & \text{else} \end{cases}$$

   (b) Bent2 scores are recommended when the errors come from a light tailed distribution, and the score function is defined in terms of 5 parameters. In Rfit, they can be set by `scores=bentscores2`.

$$\varphi(u) = \begin{cases} \frac{(s_5-s_3)u}{s_1} & \text{if } u < s_1 \\ \frac{(s_4-s_5)(u-s_2)}{1-s_2} + s_5 & \text{if } u > s_2 \\ s_5 & \text{else} \end{cases}$$

   (c) Bent3 scores are recommended when the errors come from a highly left skewed distribution, and the score function is defined in terms of 3 parameters. In Rfit, they can be set by `scores=bentscores3`.

$$\varphi(u) = \begin{cases} s_2 & \text{if } u < s_1 \\ \frac{(s_3-s_2)(u-s_1)}{1-s_1} + s_2 & \text{else} \end{cases}$$

(d) Bent4 scores are recommended when the errors come from a moderately heavy tailed distribution, and the score function is defined in terms of 4 parameters. In Rfit, they can be set by `scores=bentscores4`.

$$\varphi(u) = \begin{cases} s_3 & \text{if } u < s_1 \\ s_4 & \text{if } u > s_2 \\ \frac{(s_4-s_3)(u-s_1)}{s_2-s_1} + s_3 & \text{else} \end{cases}$$

Plots of all these score functions are showed in Figure 10.

An appropriate bent score function for skewed distribution with a right heavy tail is as follows (Kloke and McKean, 2012):

$$\varphi(u) = \begin{cases} 4u - 1.5 & \text{if } u \leq 0.5 \\ 0.5 & \text{if } u > 0.5 \end{cases}$$

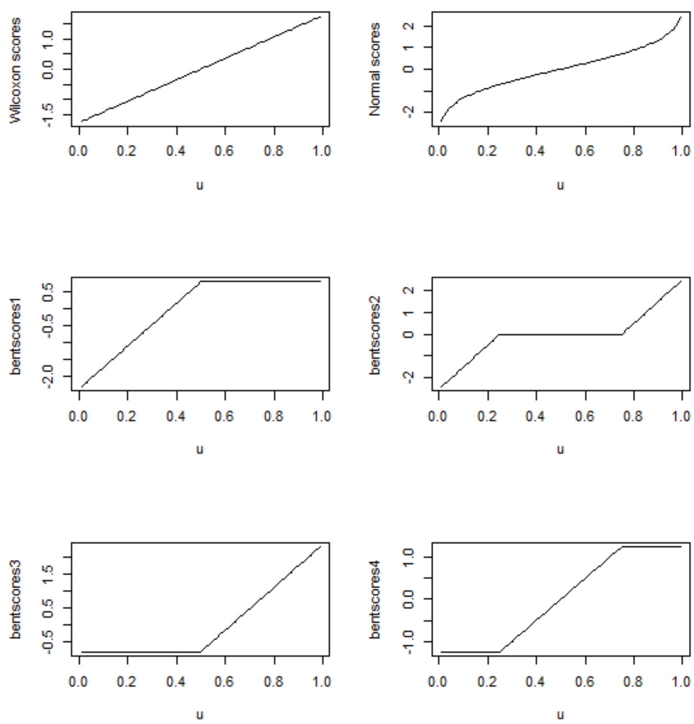

Figure 10: Plots of score functions.

As shown in HM, the influence function of the rank-based estimation is bounded in the Y-space provided the score functions are bounded. Hence, all these score functions lead to robust fits in Y-space except for the normal scores fit. The normal scores estimator, however, is technically robust, since it has an unbounded influence function, but a positive breakdown point (Huber, 1981). The Wilcoxon procedures in linear models have the same efficiency properties as the Wilcoxon-Mann-Whitney procedures in location models. In particular, if the errors have a normal distribution this efficiency is 0.95. Generally, for distributions with heavier tails than normal, this efficiency is too large. For these reasons, the Wilcoxon scores are often used in practice. In this thesis, we often use Wilcoxon scores, but we also make use of other scores, especially the bent scores. For most of our discussion, the full model design is a one-way design. So there is no need for high breakdown estimates. A weighted version of the Wilcoxon can attain 50% breakdown in the X-space at the expense of a loss in efficiency (Chang et al., 1999).

2.2.3 The RAT Test

Let

$$\Delta = \boldsymbol{h}'\boldsymbol{\theta},$$

where $\boldsymbol{\theta}$ represents a $k \times 1$ vector of the location parameters of k treatment levels; $\boldsymbol{h}$ is a $k \times 1$ vector of the AT maximin contrast coefficients. When $\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{y}}$, where $\bar{\boldsymbol{y}}$ is a $k \times 1$ vector of the average of each treatment level, then the test statistic of the AT can be written as:

$$t = \frac{\boldsymbol{h}'\bar{\boldsymbol{y}}}{\sqrt{\mathrm{Var}(\boldsymbol{h}'\bar{\boldsymbol{y}})}}.$$

Now, a rank-based estimator of Δ is obtained by using the rank-based linear model. Let $\boldsymbol{W}$ be a $n \times k$ matrix which denotes the appropriate incidence matrix of 0s and 1s, and note that the vector $\mathbf{1}_n$ is in the column space of $\boldsymbol{W}$, then

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{W}\boldsymbol{\theta} + \boldsymbol{e} \\
&= \boldsymbol{W}\boldsymbol{E}\boldsymbol{E}^{-1}\boldsymbol{\theta} + \boldsymbol{e} \\
&= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \\
&= \begin{bmatrix} \mathbf{1} & \boldsymbol{X}_{\mathbf{1}} \end{bmatrix} \begin{bmatrix} \alpha \\ \boldsymbol{\beta}_{\mathbf{1}} \end{bmatrix} + \boldsymbol{e} \\
&= \alpha\mathbf{1} + \boldsymbol{X}_{\mathbf{1}}\boldsymbol{\beta}_{\mathbf{1}} + \boldsymbol{e},
\end{aligned}$$

where,

$$\boldsymbol{W}_{(n\times k)} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_k} \end{bmatrix} \tag{2.3}$$

$$\boldsymbol{E}_{(k\times k)} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \tag{2.4}$$

$$\boldsymbol{X}_{\mathbf{1}(n\times(k-1))} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_3} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_k} \end{bmatrix} \tag{2.5}$$

and $\boldsymbol{WE} = \boldsymbol{X}$, $\boldsymbol{E}^{-1}\boldsymbol{\theta} = \boldsymbol{\beta}$. Thus, the rank-based estimator of $\boldsymbol{\Delta}$ is

$$\begin{aligned}
\hat{\Delta} &= \boldsymbol{h}'\hat{\boldsymbol{\theta}} \\
&= \boldsymbol{h}'\boldsymbol{E}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{h}'\begin{bmatrix}\mathbf{1} & \boldsymbol{E}_1\end{bmatrix}\hat{\boldsymbol{\beta}} \\
&= \begin{bmatrix}\mathbf{0} & \boldsymbol{h}'\boldsymbol{E}_1\end{bmatrix}\hat{\boldsymbol{\beta}} \\
&= \boldsymbol{h}'\boldsymbol{E}_1\hat{\boldsymbol{\beta}}_1.
\end{aligned}$$

Under the assumption that the errors are iid with continuous probability density function $f(t)$, it can be shown that the estimate $\hat{\boldsymbol{\beta}}_1$ is consistent and asymptotically normal (Hettmansperger and McKean, 2011). The result can be summarized as follows:

$$\hat{\boldsymbol{\beta}}_1 \dot{\sim} N(\boldsymbol{\beta}_1, \tau_\varphi^2(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}), \tag{2.6}$$

where $\boldsymbol{X_c}$ is centered $\boldsymbol{X}_1$, $\boldsymbol{X_c} = (\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\boldsymbol{X}_1$; where $\boldsymbol{I}$ is a $n \times n$ identity matrix and $\mathbf{1}' = (1, \ldots, 1)_{1\times n}$; τ_φ is a scale parameter which depends on f, and the score function φ and is given by

$$\tau_\varphi^{-1} = \int \varphi(u)\varphi_f(u)du,$$

$$\varphi_f(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}.$$

Thus, the asymptotic distribution of $\hat{\Delta}$ is

$$\hat{\Delta} \dot{\sim} N(\boldsymbol{h}'\boldsymbol{E}_1\boldsymbol{\beta}_1, \tau_\varphi^2\boldsymbol{h}'\boldsymbol{E}_1(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{E}_1'\boldsymbol{h}). \tag{2.7}$$

We denote this asymptotic variance by $\mathrm{Var}(\hat{\Delta})$.

Theorem 2.1. *If $H_0 : \theta_1 = \cdots = \theta_k$ is true, then $E(\hat{\Delta}) = 0$, as $n \to \infty$*

$$RAT = \frac{\hat{\Delta}}{\sqrt{Var(\hat{\Delta})}} \xrightarrow{D} Z \sim N(0, 1).$$

For a level α test based on RAT for the hypothesis (1.3), suppose H_1 is true, then $\Delta = \boldsymbol{h}'\boldsymbol{\theta} > 0$.

Since

$$\begin{aligned}\frac{1}{n}\boldsymbol{X}_c^{'}\boldsymbol{X}_c =&O(1)\\ n(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1} =&O(1),\end{aligned}$$

and

$$(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1} =\frac{1}{n}O(1),$$

therefore $\boldsymbol{h}^{'}\boldsymbol{E}_1(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1}\boldsymbol{E}_1^{'}\boldsymbol{h} \to 0$ as $n \to \infty$.

When n is sufficiently large, then we have

$$0 < z_\alpha\sqrt{\text{Var}(\hat{\Delta})} < \Delta - \epsilon.$$

Hence,

$$\begin{aligned} P_\Delta[RAT \geq z_\alpha] &= P_\Delta[\hat{\Delta} \geq z_\alpha\sqrt{\text{Var}(\hat{\Delta})}] \\ &\geq P_\Delta[\hat{\Delta} \geq \Delta - \epsilon] \\ &\geq P_\Delta[\left|\hat{\Delta} - \Delta\right| \leq \epsilon] \to 1, \end{aligned}$$

thus the RAT test is consistent.

For a sequence of local alternatives $H_n : \theta_j = \theta_0 + n^{-1/2}c_j\theta^*$, where $\theta^* > 0$, $j = 1, 2, \ldots, k$. Let $\boldsymbol{c}$ be a vector of c_j, a set of constants, which specifies the pattern of the alternative. Suppose H_n is true, and F has a density f with $\int f^2(x)dx < \infty$, and suppose $\hat{\boldsymbol{\theta}}_n$ is the rank based estimate of $\boldsymbol{\theta}$ under H_n, and $\hat{\boldsymbol{\theta}}_0$ is the rank based estimate of $\boldsymbol{\theta}$ under H_0 thus,

$$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_0 + n^{-1/2}\theta^*\boldsymbol{c}$$

when H_n is true,

$$\text{E}(\hat{\Delta}) = E(\boldsymbol{h}^{'}\hat{\boldsymbol{\theta}}_0 + n^{-1/2}\theta^*\boldsymbol{h}^{'}\boldsymbol{c}) = n^{-1/2}\theta^*\boldsymbol{h}^{'}\boldsymbol{c}$$

$$\text{Var}(\hat{\Delta}) = \text{Var}(\boldsymbol{h}'\hat{\boldsymbol{\theta}}_0) = \tau_\varphi^2 \boldsymbol{h}'\boldsymbol{E}_1(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{E}_1'\boldsymbol{h}.$$

The asymptotic power of the test based on RAT is given by

$$\begin{aligned} P_{\Delta_n}[RAT \geq Z_\alpha] &= P_{\Delta_n}[\hat{\Delta} \geq Z_\alpha \sqrt{\text{Var}(\hat{\Delta})}] \\ &= P_{\Delta_n}[\frac{\hat{\Delta} - n^{-1/2}\theta^* \boldsymbol{h}'\boldsymbol{c}}{\sqrt{\text{Var}(\hat{\Delta})}} \geq Z_\alpha - \frac{n^{-1/2}\theta^* \boldsymbol{h}'\boldsymbol{c}}{\sqrt{\text{Var}(\hat{\Delta})}}] \\ &= 1 - \Phi(Z_\alpha - \theta^* c_{\varphi 0}), \end{aligned}$$

where $c_{\varphi 0}$ is the Pitman efficacy of the RAT test. This can be written as

$$c_{\varphi 0} = n^{-1/2}\boldsymbol{h}'\boldsymbol{c}/\sqrt{\text{Var}(\hat{\Delta})} = n^{-1/2}\frac{\sum h_j c_j}{\tau_\varphi(\sum \frac{h_j^2}{n_j})^{1/2}}.$$

To obtain the asymptotic relative efficiency, an algebra form representing $\text{Var}(\hat{\Delta})$ is given by

$$\text{Var}(\hat{\Delta}) = \tau_\varphi^2 \sum_{j=1}^{k} \frac{h_j^2}{n_j},$$

since

$$\boldsymbol{h}'\boldsymbol{E}_1 = [h_2 \; h_3 \ldots h_k],$$

and

$$(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}_{(k-1)\times(k-1)} = \begin{bmatrix} \frac{1}{n_2} + \frac{1}{n_1} & \frac{1}{n_1} & \cdots & \frac{1}{n_1} \\ \frac{1}{n_1} & \frac{1}{n_3} + \frac{1}{n_1} & \cdots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n_1} & \frac{1}{n_1} & \cdots & \frac{1}{n_k} + \frac{1}{n_1} \end{bmatrix}$$

$$\boldsymbol{h}'\boldsymbol{E}_1(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}(\boldsymbol{h}'\boldsymbol{E}_1)' = \begin{pmatrix} \frac{h_2}{n_2} - \frac{h_1}{n_1} & \frac{h_3}{n_3} - \frac{h_1}{n_1} & \cdots & \frac{h_k}{n_k} - \frac{h_1}{n_1} \end{pmatrix} \begin{pmatrix} h_2 \\ h_3 \\ \cdots \\ h_k \end{pmatrix} = \sum_{j=1}^{k} \frac{h_j^2}{n_j}.$$

2.3 Simulation Results

In the previous sections it has been shown that the RAT test can use different score functions for different underlying distributions. The 10,000 simulation results in Table 6 show that RAT with Wilcoxon scores performs well, except that under t(1) it appears to be conservative. Because the t(1) distribution is symmetrical with heavy tails, the bent4 score would be more appropriate. We set $s_1 = 0.15$, $s_2 = 0.85$, $s_3 = -1$, $s_4 = 1$. For the bent4 score, the empirical levels of the corresponding RAT test, 0.0927 and 0.0487, were at nominal 10% and 5% levels respectively. Note that the RAT with the bent4 scores have empirical levels close to the nominal α under the t(1) distribution. Again, note the distribution freeness of the JT, SP and HN tests.

Table 6: Empirical αs at nominal levels.

	$N(0,1)$		$t(10)$		$t(5)$		$t(2)$		$t(1)$	
Meth.	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
SP	.0989	.0517	.1001	.0457	.1039	.0520	.1012	.0518	.0989	.0480
JT	.1036	.0509	.1016	.0461	.1079	.0503	.1046	.0522	.1024	.0477
AT	.1013	.0495	.0994	.0479	.1038	.0512	.1065	.0504	.1191	.0414
RAT	.1026	.0512	.0982	.0504	.1010	.0493	.0921	.0469	.0729	.0355
RATb4	.1168	.0638	.1128	.0613	.1124	.0625	.1078	.0591	.0927	.0487
HN	.1038	.0517	.1034	.0457	.1078	.0520	.1050	.0518	.1023	.0480

Table 7 shows the empirical powers of all five methods with different underlying distributions. The JT and HN tests have the highest empirical power at each simulation. Their empirical powers are not significantly different.

Table 7: Empirical power at nominal levels.

	$N(0,1)$		$t(10)$		$t(5)$		$t(2)$		$t(1)$	
Meth.	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
SP	.5433	.3966	.5108	.3629	.4638	.3344	.3948	.2598	.3094	.1896
JT	.5519	.3909	.5218	.3583	.4728	.3287	.4048	.2525	.3173	.1926
AT	.5401	.3943	.4957	.3453	.4356	.3000	.2925	.1770	.1841	.0832
RAT	.5168	.3640	.4844	.3332	.4384	.3020	.3521	.2174	.2196	.1205
RATb4	.5248	.3803	.4963	.3538	.4525	.3204	.3726	.2450	.2444	.1448
HN	.5516	.3966	.5184	.3629	.4715	.3344	.4027	.2598	.3163	.1896

Table 8 and Figure 11 show the empirical power of five methods under normal distribution. There is no significant difference.

Table 9 and Figure 12 show the empirical power of all five methods under the skewed contaminated normal distribution ($SCN(.25, 100, 1)$) . If we use the last column of the empirical power of table 9 to compute a 95% confidence interval of the proportion difference, we find the RAT method has higher power than AT methods under the skewed contaminated normal distribution with (.3135, .3603), and especially in the RAT with bent1 score method, with (.3402, .3872).

Table 8: Empirical power levels under the *Normal* distribution.

Meth.	Empirical power			
	0	2.11	8.44	18.99
SP	.1006	.5433	.9341	.9980
JT	.1030	.5519	.9356	.9978
AT	.1012	.5401	.9354	.9978
RAT	.1050	.5168	.9143	.9964
RATb1	.1174	.5248	.9060	.9951
HN	.1035	.5516	.9374	.9980

In figure 12, the power curve of the RAT method is below the JT, HN, and SP methods, but the power curve of the RAT with bent1 score method crosses the curves of the HN and SP methods, and falls right below the JT curve. There are significant differences between RAT and AT, and between RATb1 and AT. The RAT and RATb1 tests are more powerful than the AT test. The 95% confidence interval of RATb1 and JT shows there is no significant difference between them with (-.0397, .0149).

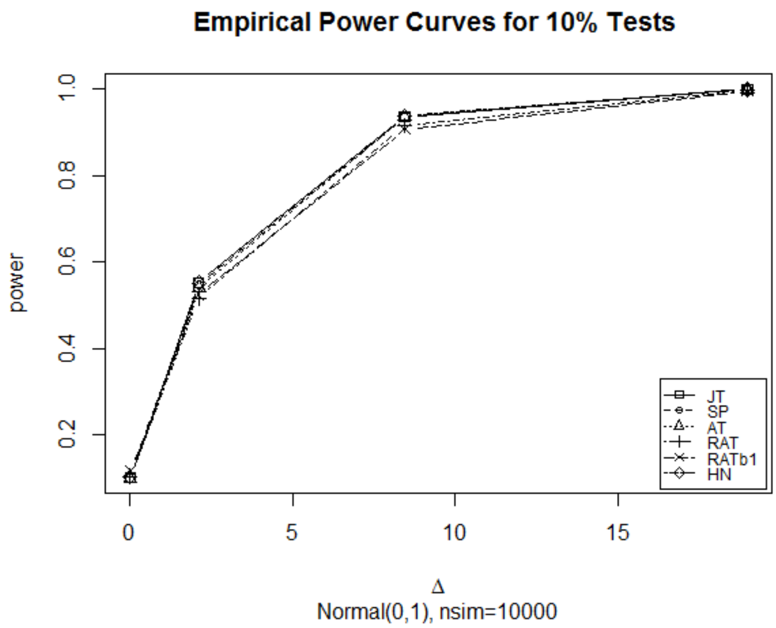

Figure 11: Plot of power curves under a *Normal* distribution.

Table 9: Empirical power levels under the SCN distribution.

Meth.	Empirical power							
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39
SP	.1079	.3233	.5953	.7763	.8657	.9150	.9395	.9548
JT	.1109	.3317	.6151	.8011	.8994	.9420	.9660	.9791
AT	.1045	.1531	.2123	.2886	.3717	.4540	.5299	.6030
RAT	.0714	.1780	.3671	.5867	.7566	.8603	.9132	.9399
RATb1	.0910	.2708	.5394	.7575	.8729	.9285	.9584	.9667
HN	.1112	.3314	.6035	.7818	.8719	.9179	.9416	.9563

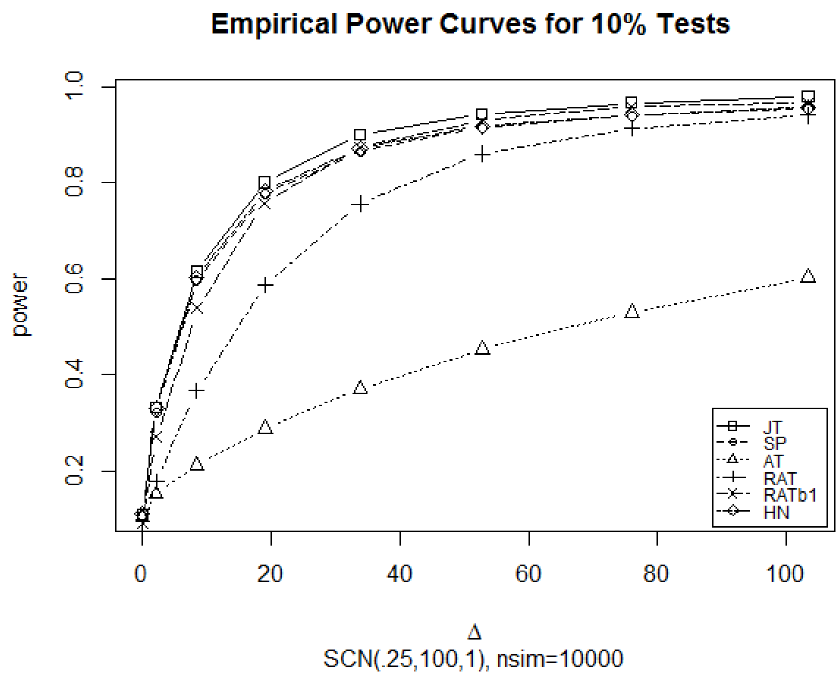

Figure 12: Plot of power curves under the SCN distribution

Based on all the simulation results, we can conclude that the RAT test is robust, it allows any scores, and it did better than the AT test. Under the skewed contaminated normal distribution ($SCN(.25, 1, 100)$) the RAT with bent1 scores test does not outperform the JT test.

Terpstra (1952) and Jonckheere (1954) proposed the test based on pairwise Mann-Whitney-Wilcoxon statistics, while RAT only uses partial shifts based on the number of groups (k) as even or odd. Such as when $k = 5$, there are 10 U_{ij}s in the JT statistic, while only 2 Δ_{ij}s exist in the RAT statistic.

In the JT statistic:

$$\begin{aligned} J &= \sum\sum_{i<j} U_{ij} \\ &= U_{12} + U_{13} + U_{14} + U_{15} + U_{23} + U_{24} + U_{25} + U_{34} + U_{35} + U_{45}. \end{aligned}$$

In the RAT statistic:

$$\begin{aligned} \Delta &= \boldsymbol{c}^{'}\boldsymbol{\theta} \\ &= -0.89\theta_1 - 0.2\theta_2 + 0.2\theta_4 + 0.89\theta_5 \\ &= 0.89(\theta_5 - \theta_1) + 0.2(\theta_4 - \theta_2) \\ &= 0.89\Delta_{51} + 0.2\Delta_{42}. \end{aligned}$$

Thus, JT uses more information than RAT, and the pairwise statistics provide more direct information on the sources of statistical significance. In light of this, we consider the Hettmansperger-Norton optimal weights to be a better choice.

CHAPTER 3

NEW METHODS

Recall that in this k-sample location problem, the response Y_{ij} follows the linear model,

$$Y_{ij} = \theta_j + \epsilon_{ij},\ j = 1, \ldots,\ k,\ i = 1, \ldots,\ n_j.$$

Y_{ij} is the ith observation on response variable for the jth treatment level; θ_j is the jth location parameter; ϵ_{ij} are independent and identically distributed (iid) with cdf $F(x)$ and pdf $f(x)$. The null hypothesis of interest is that there is no differences in locations among k populations, or, there is no difference in treatment effects (i.e. 1.2), and the alternative hypotheses of interest follow a pattern (i.e. (1.3), (1.4) or (1.5)).

In this chapter, three new methods, the robust Abelson-Tukey with HN weights (RAThn) test, the Robust Shao-McKean (RSM) test and the RSM with least square estimates (RSM$_{LS}$) test, are presented. All three methods have the same weights, the HN weights, and the same form of a test statistic, but have different estimates of the shifts.

The test statistics used in this chapter have the form:

$$T = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (a_j - a_i)\hat{\Delta}_{ji}, \tag{3.1}$$

where $a_j = \lambda_j(c_j - \bar{c}_w)$, c_j specifics the pattern of the alternative, and $\bar{c}_w = \sum \lambda_j c_j$, $\lambda_j = \frac{n_j}{n}$ and $\sum_{j=1}^{k} a_j = 0$. $\hat{\Delta}_{ji}$ is an estimate which measures the shift from group j to group i.

The robust Abelson-Tukey with HN weights (RAThn) test uses the rank-based regression estimate; the Robust Shao-McKean (RSM) test uses a pseudo-median estimate, the Hodges-Lehmann estimate; the RSM with least square estimates (RSM_{LS}) test uses the least square estimate to estimate the shift $\hat{\Delta}_{ji}$.

3.1 Robust Abelson-Tukey with HN Weights (RAThn) Test

Recall in section 2.2 that a rank-based estimate of $\boldsymbol{\beta}$ (the shifts of other groups from the first group) is defined as

$$\hat{\boldsymbol{\beta}} = \text{Argmin} \left\| \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} \right\|_{\varphi},$$

and under the assumption that the errors are iid with continuous probability density function, the estimate $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal, which is shown as 2.6. Even though it does not give the rank-based estimates of the pairwise shifts directly, but the following lemma shows that the test statistic can be represented as a linear combination of the shifts of all groups from the first group, $\hat{\Delta}_{i1}$, $i = 1,\ 2,\ \ldots\ k$.

Lemma 1. *Suppose* $\Delta_{ji} = \Delta_{j1} - \Delta_{i1}$, *where* $\sum_{j=1}^{k} a_j = 0$ *and* $\Delta_{11} = 0$. *Then*

$$\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (a_j - a_i)\Delta_{ji} = k \sum_{i=1}^{k} a_i \Delta_{i1}.$$

Proof:

$$\begin{aligned}
\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (a_j - a_i)\Delta_{ji} &= \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (a_j - a_i)(\Delta_{j1} - \Delta_{i1}) \\
&= \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_j \Delta_{j1} - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_i \Delta_{j1} \\
&\quad - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_j \Delta_{i1} + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} a_i \Delta_{i1}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{k}(i-1)a_i\Delta_{i1} + \sum_{i=1}^{k-1}(k-i)a_i\Delta_{i1} \\
&\quad - \sum_{i=2}^{k}\sum_{j<i} a_j\Delta_{i1} - \sum_{i=1}^{k-1}\sum_{j>i} a_j\Delta_{i1} \\
&= \sum_{i=1}^{k-1}(i-1)a_i\Delta_{i1} + (k-1)a_k\Delta_{k1} + \sum_{i=1}^{k-1}(k-i)a_i\Delta_{i1} - \sum_{i=1}^{k}\sum_{j\neq i} a_j\Delta_{i1} \\
&= \sum_{i=1}^{k-1}(k-i+i-1)a_i\Delta_{i1} + (k-1)a_k\Delta_{k1} - \sum_{i=1}^{k}\Delta_{i1}\sum_{j\neq i} a_j \\
&= \sum_{i=1}^{k-1}(k-1)a_i\Delta_{i1} + (k-1)a_k\Delta_{k1} + \sum_{i=1}^{k}\Delta_{i1}a_i \\
&= (k-1)\sum_{i=1}^{k} a_i\Delta_{i1} + \sum_{i=1}^{k}\Delta_{i1}a_i \\
&= k\sum_{i=1}^{k} a_i\Delta_{i1} \\
&= k\sum_{i=2}^{k} a_i\Delta_{i1}.
\end{aligned}$$

$(\sum_{j=1}^{k} a_j = 0 \Rightarrow \sum_{j\neq i} a_j = -a_i)$.

3.1.1 The RAThn Test Statistic

The RAThn test statistic is

$$T = k\boldsymbol{a}'\hat{\boldsymbol{\beta}}_1, \tag{3.2}$$

where $\hat{\boldsymbol{\beta}}_1$ is a vector of the estimates of the shifts from group i to group 1, $\underline{\boldsymbol{a}} = (a_2,\ a_3,\ \ldots\ a_k)'$, where $\sum_{i=1}^{k} a_i = 0$. Based on (2.6) the asymptotically distribution of T is

$$T \dot{\sim} N(k\boldsymbol{a}'\boldsymbol{\beta}_1, k^2\tau_\varphi^2\boldsymbol{a}'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{a}). \tag{3.3}$$

Theorem 3.1. *If $H_0 : \theta_1 = \cdots = \theta_k$ is true, then $E(T) = 0$. and $Var(T) =$*

$k^2\hat{\tau}_\varphi^2 \boldsymbol{a}^{'}(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1}\boldsymbol{a}$. *Then as* $n \to \infty$, *where* $n = \sum_{i=1}^{k} n_i$,

$$T^* = \frac{k\boldsymbol{a}^{'}\hat{\boldsymbol{\beta}}_1}{\sqrt{Var(T)}} \xrightarrow{D} Z \sim N(0,1).$$

The asymptotic test is

$$\text{Reject } H_0 \text{ (1.2) in favor of } H_A \text{ (1.3)}$$
$$\text{if } T^* \geq z_\alpha\text{; otherwise do not reject.}$$

3.1.2 Consistency and Asymptotic Power

Suppose H_1 is true, then $T = ka^{'}\beta_1 > 0$.

Since

$$\begin{aligned} \frac{1}{n}\boldsymbol{X}_c^{'}\boldsymbol{X}_c =& O(1) \\ n(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1} =& O(1), \end{aligned}$$

and

$$(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1} = \frac{1}{n}O(1),$$

thus, $\boldsymbol{a}^{'}(\boldsymbol{X}_c^{'}\boldsymbol{X}_c)^{-1}\boldsymbol{a} \to 0$ as $n \to \infty$.

If n is sufficiently large, then we have

$$0 < z_\alpha\sqrt{\text{Var}(T)} < ka^{'}\beta_1 - \epsilon.$$

Hence,

$$\begin{aligned} P\left[T^* \geq z_\alpha\right] &= P\left[T \geq z_\alpha\sqrt{\text{Var}(T)}\right] \\ &\geq P\left[ka^{'}\hat{\beta}_1 \geq ka^{'}\beta_1 - \epsilon\right] \\ &\geq P\left[\left|\hat{\beta}_1 - \beta_1\right| \leq \epsilon\right] \to 1. \end{aligned}$$

Thus, RAThn is a consistent test.

Under a sequence of alternatives $H_n : \theta_j = \theta_0 + \theta c_j n^{-1/2}$, where $\theta > 0$, $j =$

$1, \ldots, k$, and F has a density f with $\int f^2(x)dx < \infty$,

$$\Delta_{i1} = (c_i - c_1)\theta n^{-1/2}.$$

Let $\boldsymbol{D}$ denote the vector of Δ_{i1}s, and $\boldsymbol{C}$ denote the vector of $c_i - c_1$, so

$$\boldsymbol{D}_n = \boldsymbol{C}\theta n^{-1/2}.$$

When H_n is true,

$$\mathrm{E}(T) = k\boldsymbol{a}'\boldsymbol{C}\theta n^{-1/2}$$

$$\mathrm{Var}(T) = k^2\tau_\varphi^2\boldsymbol{a}'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{a}.$$

The asymptotic power of the test based on T is given by

$$\begin{aligned} P_{\boldsymbol{D}_n}[T^* \geq Z_\alpha] &= P_{\boldsymbol{D}_n}[T \geq Z_\alpha\sqrt{\mathrm{Var}(T)}] \\ &= P_{\boldsymbol{D}_n}[\frac{T - k\boldsymbol{a}'\boldsymbol{D}_n}{\sqrt{\mathrm{Var}(T)}} \geq Z_\alpha - \frac{k\boldsymbol{a}'\boldsymbol{D}_n}{\sqrt{Var(T)}}] \\ &= 1 - \Phi(Z_\alpha - \theta c_{\varphi 1}), \end{aligned}$$

where $c_{\varphi 1}$ is the Pitman efficacy of the RAThn test, and it can be written as

$$\begin{aligned} c_{\varphi 1} &= \boldsymbol{a}'\boldsymbol{C}\frac{kn^{-1/2}}{\sqrt{k^2\tau_\varphi^2}}(\boldsymbol{a}'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{a})^{-1/2} \\ &= \frac{1}{\sqrt{n}\tau_\varphi}\boldsymbol{a}'\boldsymbol{C}(\boldsymbol{a}'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{a})^{-1/2} \\ &= \frac{\sum_{i=1}^k a_i c_i}{\tau_\varphi(\sum_{i=1}^k \frac{a_i^2}{\lambda_i})^{1/2}}. \end{aligned}$$

3.2 Robust Shao-McKean (RSM) Test

Hettmansperger-Norton (1987) used a linear combination of pairwise comparisons based on the MWW statistic $U_{ij} = \sum_{m=1}^{n_i}\sum_{l=1}^{n_j}\phi(Y_{mi},\ Y_{lj})$, $(1 \leq i < j \leq k)$, where $\phi(a,\ b) = 1$, if $a < b$, 0 otherwise. Our effect-size test is based on a linear combination of the estimates $\hat{\Delta}_{ij}$.

Suppose that Y_{ij} is distributed with continuous cdf $F(y-\theta_j)$ for $j=1, \ \ldots, \ k$ and $i=1, \ \ldots, \ n_j$ and the observations are mutually independent. We want to test $H_0: \theta_1 = \cdots = \theta_k$ versus $H_A: \theta_j = \theta_0 + \theta c_j, (\theta > 0; \ j=1, \ \ldots, \ k)$, where $c_1, \ \ldots, \ c_k$ is a given set of constants that specifies the pattern of the alternative. The experiment may specify the pattern and spacings based on previous studies and/or theoretical formulations. Equally spaced constants are recommended unless there are indications to the contrary.

3.2.1 The RSM Test Statistic

For any pattern with a known peak, the RSM test statistic is

$$RSM = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (a_j - a_i)\hat{\Delta}_{ij} = \sum \sum_{i<j} a_{ij}\hat{\Delta}_{ij}. \tag{3.4}$$

The Hodges-Lehmann estimate of shift for Δ_{ij} is

$$\hat{\Delta}_{ij} = med\left\{Y_{lj} - Y_{l'i}\right\}, \ 1 \le l \le n_j, \ 1 \le l^{'} \le n_i. \tag{3.5}$$

Under the null hypothesis $E(RSM)=0$, and

$$Var(RSM) = \sum_{i<j} \sum_{i'<j'} a_{ij} a_{i'j'} Cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}). \tag{3.6}$$

From $(HM2.5.30)$, $(HM2.5.20)$ and $(HM2.5.5)$ we have the asymptotic representation

$$\sqrt{n_{ij}}\hat{\Delta}_{ij} = \frac{\tau_\varphi}{\sqrt{n_{ij}}\lambda_i^{ij}\lambda_j^{ij}} T_\varphi(0) + o_p(1), \tag{3.7}$$

where

$$T_\varphi(0) = \sum_{i'=1}^{n_{ij}} (b_{i'} - \bar{b}_{ij})\varphi\left[F(Z_{i'})\right], \tag{3.8}$$

where $n_{ij} = n_i + n_j$, $b_{i'} = 1$ if $n_i + 1 \le i' \le n_{ij}$; $b_{i'} = 0$ otherwise, $\lambda_i^{ij} = n_i/n_{ij}$, $\bar{b}_{ij} = \sum b_i/n_{ij} = n_j/n_{ij}$, $(1 - \bar{b}_{ij}) = n_i/n_{ij}$ and

$$Z_{i'} = \Delta_{ij} b_{i'} + e_{i'},\ 1 \le i' \le n_{ij},$$

and where $e_1, \ldots e_{n_{ij}}$ are iid with distribution function $F(x)$. And

$$\int_0^1 \varphi(u) d_u = 0, \text{ and } \int_0^1 \varphi^2(u) d_u = 1. \tag{3.9}$$

Thus,

$$\begin{aligned}
Cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}) &= \frac{\tau_\varphi^2}{n_{ij} n_{i'j'} \lambda_i^{ij} \lambda_j^{ij} \lambda_{i'}^{i'j'} \lambda_{j'}^{i'j'}} \sum_{m=1}^{n_{ij}} \sum_{l=1}^{n_{i'j'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) \\
&\quad cov(\varphi(F(Z_m^{ij})), \varphi(F(Z_l^{i'j'}))) \\
&= \kappa \sum_{m=1}^{n_i} \sum_{l=1}^{n_{i'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) cov(\varphi(F(Z_m^{ij})), \varphi(F(Z_l^{i'j'}))) \\
&\quad + \kappa \sum_{m=1}^{n_i} \sum_{l=i'+1}^{n_{i'j'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) cov(\varphi(F(Z_m^{ij})), \varphi(F(Z_l^{i'j'}))) \\
&\quad + \kappa \sum_{m=n_i+1}^{n_{ij}} \sum_{l=1}^{n_{i'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) cov(\varphi(F(Z_m^{ij})), \varphi(F(Z_l^{i'j'}))) \\
&\quad + \kappa \sum_{m=n_i+1}^{n_{ij}} \sum_{l=n_{i'}+1}^{n_{i'j'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) cov(\varphi(F(Z_m^{ij})), \varphi(F(Z_l^{i'j'}))),
\end{aligned}$$

where $\kappa = \frac{\tau_\varphi^2}{n_{ij} n_{i'j'} \lambda_i^{ij} \lambda_j^{ij} \lambda_{i'}^{i'j'} \lambda_{j'}^{i'j'}}$.

- $cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}) = 0$, if there are no ties among i, j, i', j';
- if $i = i'$, $cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}) = \kappa \sum_{m=1}^{n_i} \sum_{l=1}^{n_{i'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) = \frac{\tau_\varphi^2}{n_i}$;
- if $j = j'$, $cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}) = \kappa \sum_{m=n_i+1}^{n_{ij}} \sum_{l=i'+1}^{n_{i'j'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) = \frac{\tau_\varphi^2}{n_j}$;
- if $i = j'$, $cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}) = \kappa \sum_{m=1}^{n_i} \sum_{l=i'+1}^{n_{i'j'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) = -\frac{\tau_\varphi^2}{n_i}$;
- if $j = i'$, $cov(\hat{\Delta}_{ij}, \hat{\Delta}_{i'j'}) = \kappa \sum_{m=n_i+1}^{n_{ij}} \sum_{l=1}^{n_{i'}} (b_m^{ij} - \bar{b}_{ij})(b_l^{i'j'} - \bar{b}_{i'j'}) = -\frac{\tau_\varphi^2}{n_j}$;

Also, the asymptotic variance of $\hat{\Delta}_{ij}$ is

$$\begin{aligned}
AV(\hat{\Delta}_{ij}) &= \frac{\tau_\varphi^2}{n_{ij}^2\lambda_i^2\lambda_j^2}\sum_{i'=1}^{n_{ij}}(b_{i'}-\bar{b}_{ij})^2 Var\left[\varphi(F(Z_{i'}))\right] \\
&= \frac{\tau_\varphi^2}{n_{ij}^2\lambda_i^2\lambda_j^2}\sum_{i'=1}^{n_{ij}}(b_{i'}-\bar{b}_{ij})^2 \\
&= \frac{\tau_\varphi^2 n_{ij}^2}{n_i^2 n_j^2}\left[n_i(0-\bar{b}_{ij})^2+n_j(1-\bar{b}_{ij})^2\right] \\
&= \frac{\tau_\varphi^2 n_{ij}^2}{n_i^2 n_j^2}\left[\frac{n_i n_j^2}{n_{ij}^2}+\frac{n_i^2 n_j}{n_{ij}^2}\right] \\
&= \frac{\tau_\varphi^2}{n_i^2 n_j^2}n_i n_j(n_i+n_j) \\
&= \frac{\tau_\varphi^2}{n_i n_j}n_{ij} \\
&= \tau_\varphi^2(\frac{1}{n_i}+\frac{1}{n_j}).
\end{aligned}$$

From (3.4) we have

$$RSM = \sum\sum_{i<j}(a_i-a_j)\hat{\Delta}_{ij}.$$

If we let

$$RSM = \sum\sum_{i<j} v_{ij}$$

Therefore,

$$\begin{aligned}
Var(RSM) &= \sum\sum_{i<j} Var(v_{ij}) + \sum\sum_{i<j}[\sum_{i'=1,i'\neq i}^{j-1} Cov(v_{ij}, v_{i'j}) \\
&+ \sum_{j'=i+1,j'\neq j}^{k} Cov(v_{ij}, v_{ij'}) + \sum_{i'=j+1,i'\neq i}^{k} Cov(v_{ij}, v_{ji'}) \\
&+ \sum_{j'=1,j'\neq j}^{i-1} Cov(v_{ij}, v_{j'i})]
\end{aligned}$$

continuously,

$$
\begin{aligned}
Var(RSM) &= \sum\sum_{i<j}(a_j-a_i)^2\tau_\varphi^2(\frac{1}{n_i}+\frac{1}{n_j}) \\
&+\sum\sum_{i<j}[\sum_{i'=1,i'\neq i}^{j-1}(a_j-a_i)(a_j-a_{i'})\frac{\tau_\varphi^2}{n_j} \\
&+\sum_{j'=i+1,j'\neq j}^{k}(a_j-a_i)(a_{j'}-a_i)\frac{\tau_\varphi^2}{n_i} \\
&-\sum_{i'=j+1,i'\neq i}^{k}(a_j-a_i)(a_{i'}-a_j)\frac{\tau_\varphi^2}{n_j} \\
&-\sum_{j'=1,j'\neq j}^{i-1}(a_j-a_i)(a_i-a_{j'})\frac{\tau_\varphi^2}{n_i}] \\
&= \tau_\varphi^2\sum\sum_{i<j}(a_j-a_i)^2(\frac{1}{n_i}+\frac{1}{n_j}) \\
&+\tau_\varphi^2\sum\sum_{i<j}(a_j-a_i)[\sum_{i'=1,i'\neq i,j}^{k}(a_j-a_{i'})\frac{1}{n_j} \\
&+\sum_{j'=1,j'\neq i,j}^{k}(a_{j'}-a_i)\frac{1}{n_i}].
\end{aligned}
$$

Since $\sum_{i'=1,i'\neq i,j}^{k} a_j = (k-2)a_j$, and $\sum_{i'=1,i'\neq i,j}^{k} a_{i'} = -(a_i+a_j)$, $(\sum a_i = 0)$, $Var(RSM)$ simplifies to

$$\begin{aligned}
Var(RSM) &= \tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)^2 (\frac{1}{n_i} + \frac{1}{n_j}) \\
&\quad + \tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)[\frac{1}{n_j}[(k-2)a_j + a_i + a_j] - \frac{1}{n_i}[(k-2)a_i + a_i + a_j]] \\
&= \tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)^2 (\frac{1}{n_i} + \frac{1}{n_j}) \\
&\quad + \tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)[k(\frac{a_j}{n_j} - \frac{a_i}{n_i}) - (a_j - a_i)(\frac{1}{n_i} + \frac{1}{n_j})] \\
&= \tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)^2 (\frac{1}{n_i} + \frac{1}{n_j}) \\
&\quad + k\tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)(\frac{a_j}{n_j} - \frac{a_i}{n_i}) - \tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)^2 (\frac{1}{n_i} + \frac{1}{n_j}) \\
&= k\tau_\varphi^2 \sum\sum_{i<j} (a_j - a_i)(\frac{a_j}{n_j} - \frac{a_i}{n_i}) \\
&= k\tau_\varphi^2 \sum_{i=1}^{j-1}\sum_{j=2}^{k} (\frac{a_j^2}{n_j} - \frac{a_j}{n_j}a_i - a_j\frac{a_i}{n_i} + \frac{a_i^2}{n_i}) \\
&= k\tau_\varphi^2 \sum_{j=2}^{k} \left[(j-1)\frac{a_j^2}{n_j} - \frac{a_j}{n_j}\sum_{i=1}^{j-1} a_i - a_j \sum_{i=1}^{j-1} \frac{a_i}{n_i} + \sum_{i=1}^{j-1} \frac{a_i^2}{n_i}\right]
\end{aligned}$$

$$\begin{aligned}
\sum_{j=2}^{k} \frac{a_j}{n_j} \sum_{i=1}^{j-1} a_i &= \sum_{j=2}^{k-1} \frac{a_j}{n_j} \sum_{i=1}^{j-1} a_i + \frac{a_k}{n_k} \sum_{i=1}^{k-1} a_i \\
&= \sum_{j=2}^{k-1} \frac{a_j}{n_j} \sum_{i=1}^{j-1} a_i - \frac{a_k^2}{n_k}
\end{aligned}$$

$$\begin{aligned}
\sum_{j=2}^{k} a_j \sum_{i=1}^{j-1} \frac{a_i}{n_i} &= \sum_{j=1}^{k-1} \frac{a_j}{n_j} \sum_{i=j+1}^{k} a_i \\
&= \frac{a_1}{n_1} \sum_{i=2}^{k} a_i + \sum_{j=2}^{k-1} \frac{a_j}{n_j} \sum_{i=j+1}^{k} a_i \\
&= -\frac{a_1^2}{n_1} + \sum_{j=2}^{k-1} \frac{a_j}{n_j} \sum_{i=j+1}^{k} a_i
\end{aligned}$$

$$
\begin{aligned}
\sum_{j=2}^{k-1}\frac{a_j}{n_j}\sum_{i=1}^{j-1}a_i-\frac{a_k^2}{n_k}-\frac{a_1^2}{n_1}+\sum_{j=2}^{k-1}\frac{a_j}{n_j}\sum_{i=j+1}^{k}a_i &= -\frac{a_k^2}{n_k}-\frac{a_1^2}{n_1}+\sum_{j=2}^{k-1}\frac{a_j}{n_j}\left[\sum_{i=1}^{j-1}a_i+\sum_{i=j+1}^{k}a_i\right]\\
&= -\frac{a_k^2}{n_k}-\frac{a_1^2}{n_1}-\sum_{j=2}^{k-1}\frac{a_j^2}{n_j}\\
&= -\sum_{j=1}^{k}\frac{a_j^2}{n_j}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
Var(RSM) &= k\tau_\varphi^2\left[\sum_{j=2}^{k}(j-1)\frac{a_j^2}{n_j}+\sum_{j=1}^{k}\frac{a_j^2}{n_j}+\sum_{j=2}^{k}\sum_{i=1}^{j-1}\frac{a_i^2}{n_i}\right]\\
&= k\tau_\varphi^2\left[\sum_{j=1}^{k}(j-1)\frac{a_j^2}{n_j}+\sum_{j=1}^{k}\frac{a_j^2}{n_j}+\sum_{j=1}^{k-1}(k-j)\frac{a_j^2}{n_j}\right]\\
&= k\tau_\varphi^2\left[\sum_{j=1}^{k}(j-1)\frac{a_j^2}{n_j}+\sum_{j=1}^{k}\frac{a_j^2}{n_j}+\sum_{j=1}^{k}(k-j)\frac{a_j^2}{n_j}\right]\\
&= k^2\tau_\varphi^2\sum_{j=1}^{k}\frac{a_j^2}{n_j}.
\end{aligned}
$$

For asymptotic theory, consider the asymptotic representation of $\hat{\Delta}_{ij}$; i.e.

$$
\begin{aligned}
\sqrt{n_{ij}}\hat{\Delta}_{ij} &= \frac{\tau_\varphi}{\sqrt{n_{ij}}\lambda_i^{ij}\lambda_j^{ij}}\sum_{m=1}^{n_{ij}}(c_m-\bar{c}_{ij})\varphi\left[F(Z_m)\right]+o_p(1)\\
&= \frac{\tau_\varphi}{\sqrt{n_{ij}}\lambda_i^{ij}\lambda_j^{ij}}\left\{\sum_{m=1}^{n_{ij}}c_m\varphi\left[F(Z_m)\right]-\bar{c}_{ij}\sum_{m=1}^{n_{ij}}\varphi\left[F(Z_m)\right]\right\}+o_p(1)\\
&= \frac{\tau_\varphi}{\sqrt{n_{ij}}\lambda_i^{ij}\lambda_j^{ij}}\left\{\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\frac{n_j}{n_{ij}}\sum_{m=1}^{n_{ij}}\varphi\left[F(Z_m)\right]\right\}+o_p(1)\\
&= \frac{\tau_\varphi}{\sqrt{n_{ij}}\lambda_i^{ij}\lambda_j^{ij}}\left\{\frac{n_i}{n_{ij}}\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\frac{n_j}{n_{ij}}\sum_{m=1}^{n_i}\varphi\left[F(Z_m)\right]\right\}+o_p(1)\\
&= \frac{\tau_\varphi}{\sqrt{n_{ij}}\lambda_i^{ij}\lambda_j^{ij}}\left\{\lambda_i^{ij}\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\lambda_j^{ij}\sum_{m=1}^{n_i}\varphi\left[F(Z_m)\right]\right\}+o_p(1)\\
&= \frac{\tau_\varphi}{\sqrt{n_{ij}}}\left\{\frac{1}{\lambda_j^{ij}}\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\frac{1}{\lambda_i^{ij}}\sum_{m=1}^{n_i}\varphi\left[F(Z_m)\right]\right\}+o_p(1),
\end{aligned}
$$

so,

$$\begin{aligned}\hat{\Delta}_{ij} &= \frac{\tau_\varphi}{n_{ij}}\left\{\frac{1}{\lambda_j^{ij}}\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\frac{1}{\lambda_i^{ij}}\sum_{m=1}^{n_i}\varphi\left[F(Z_m)\right]\right\}+o_p(\frac{1}{\sqrt{n_{ij}}})\\ &= \tau_\varphi\left\{\frac{1}{n_j}\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\frac{1}{n_i}\sum_{m=1}^{n_i}\varphi\left[F(Z_m)\right]\right\}+o_p(\frac{1}{\sqrt{n_{ij}}})\\ &= \tau_\varphi\left\{\frac{1}{n_j}\sum_{m=n_i+1}^{n_{ij}}\varphi\left[F(Z_m)\right]-\frac{1}{n_i}\sum_{m=1}^{n_i}\varphi\left[F(Z_m)\right]\right\}+o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

If $\boldsymbol{D}_\Delta$ is $p\times 1$ vector of $\hat{\Delta}_{ij}s$, $p=\binom{k}{2}$; $\boldsymbol{R}$ is $n\times 1$ vector of random variables $\varphi[F(Z_m^{n_i})]$, $i=1,\ldots,k$ and $m=1,\ldots,n_i$, then we can write

$$\boldsymbol{D}_\Delta=\tau_\varphi\boldsymbol{L}\boldsymbol{R}+o_p(\frac{1}{\sqrt{n}}),$$

where $\boldsymbol{L}$ is a $p\times n$ matrix given by

$$\boldsymbol{L}=\begin{bmatrix}-\boldsymbol{\frac{1}{n_1}}_{n_1} & \boldsymbol{\frac{1}{n_2}}_{n_2} & \boldsymbol{0}_{n_3} & \boldsymbol{0}_{n_4} & \cdots & \boldsymbol{0}_{n_{k-1}} & \boldsymbol{0}_{n_k}\\ -\boldsymbol{\frac{1}{n_1}}_{n_1} & \boldsymbol{0}_{n_2} & \boldsymbol{\frac{1}{n_2}}_{n_3} & \boldsymbol{0}_{n_4} & \cdots & \boldsymbol{0}_{n_{k-1}} & \boldsymbol{0}_{n_k}\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ \boldsymbol{0}_{n_1} & \boldsymbol{0}_{n_2} & \boldsymbol{0}_{n_3} & \boldsymbol{0}_{n_4} & \cdots & -\boldsymbol{\frac{1}{n_{k-1}}}_{n_{k-1}} & \boldsymbol{\frac{1}{n_k}}_{n_k}\end{bmatrix}.$$

Thus, the asymptotic var-cov matrix of $\boldsymbol{D}_\Delta$ is

$$\Sigma_{\boldsymbol{D}_\Delta}=\tau_\varphi^2\boldsymbol{L}Var(\boldsymbol{R})\boldsymbol{L}'=\tau_\varphi^2\boldsymbol{L}\boldsymbol{L}',$$

where $\Sigma_{\boldsymbol{R}}=I_n$.

Note that

$$\begin{aligned}RSM &= \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(a_j-a_i)\hat{\Delta}_{ij}=\sum\sum_{i<j}a_{ij}\hat{\Delta}_{ij}=\boldsymbol{A}'\boldsymbol{D}_\Delta\\ &= \tau_\varphi\boldsymbol{A}'\boldsymbol{L}\boldsymbol{R}+o_p(\frac{1}{\sqrt{n}}),\end{aligned}$$

where $\boldsymbol{A}$ is $p \times 1$ vector of $a_{ij}s$. Proof: Note from above derivations, we have

$$\boldsymbol{A}'\boldsymbol{L}(\boldsymbol{A}'\boldsymbol{L})' \;=\; k^2 \sum_{j=1}^{k} \frac{a_j^2}{n_j},$$

which is non-negative.

$$\boldsymbol{A}'\boldsymbol{L}_{1\times n} = k \begin{bmatrix} \frac{\boldsymbol{a_1}}{\boldsymbol{n_1}}_{n_1} & \frac{\boldsymbol{a_2}}{\boldsymbol{n_2}}_{n_2} & \dots & \frac{\boldsymbol{a_k}}{\boldsymbol{n_k}}_{n_k} \end{bmatrix}.$$

Since $\frac{n_j}{n} \to \lambda_j \in (0,1)$ as $n \to \infty$, n_j and n have the same order, i.e. $O(n_j) = O(n)$. Thus, as $n \to \infty$, every $\frac{a_i}{n_i} \to 0, i = 1, \dots, n$. Therefore,

$$max_{1\le i\le n} \left| \boldsymbol{A}'\boldsymbol{L}_i \right| \to 0, as\ n \to \infty.$$

Based on Corollary A.1.1 (Hettmansperger & McKean (2011), Page 447), under H_0

$$RSM = \hat{\tau}_\varphi \boldsymbol{A}'\boldsymbol{L}\boldsymbol{R} \xrightarrow{D} N(0,\ \sigma_{RSM}^2),$$

where $\sigma_{RSM}^2 = k^2 \hat{\tau}_\varphi^2 \sum_{j=1}^{k} \frac{a_j^2}{n_j}$.

When $H_0 : \theta_1 = \dots = \theta_k$ is true, then $E(RSM) = 0$. Hence, under H_0,

$$RSM^* = \frac{1}{k\sqrt{\sum_{j=1}^{k} \frac{a_j^2}{n_j}}} \boldsymbol{A}'\boldsymbol{L}\boldsymbol{R} \xrightarrow{D} Z \sim N(0,\ 1). \tag{3.10}$$

The asymptotic test is

Reject H_0 (1.2) in favor of H_A (1.3)

if $RSM^* \ge z_\alpha$; otherwise do not reject.

Theorem 3.2. *If $H_0 : \theta_1 = \dots = \theta_k$ is true, then $E(RSM) = 0$. and $Var(RSM) = k^2 \tau_\varphi^2 \sum_{j=1}^{k} \frac{a_j^2}{n_j}$. Then as $n \to \infty$, where $n = \sum_{i=1}^{k} n_i$,*

$$RSM^* = \frac{1}{k\sqrt{\sum_{j=1}^{k} \frac{a_j^2}{n_j}}} \boldsymbol{A}'\boldsymbol{L}\boldsymbol{R} \xrightarrow{D} Z \sim N(0,\ 1).$$

3.2.2 Consistency and Asymptotic Power

The null asymptotic distribution of RSM was established in the last section. Thus, a level α test based on RSM for the hypothesis (1.3) coulde be considered.

The RSM can be expressed as $\boldsymbol{A}^{'}\boldsymbol{D}_{\Delta}$. When H_1 is true, there is a $\boldsymbol{D}$ which is a vector of the pairwise shifts and $\Delta_{ij} > 0$. Since $O(n_j) = O(n)$, for $\epsilon > 0$ and for large n, $\sigma^2_{RSM} = k^2\tau^2_{\varphi}\sum_{j=1}^{k}\frac{a_j^2}{n_j} \to 0$, and $0 < z_\alpha\sigma_{RSM} < \boldsymbol{A}^{'}\boldsymbol{D} - \epsilon$, then the power of the RSM test at fixed alternatives,

$$P_{\boldsymbol{D}}(RSM \geq z_\alpha\sigma_{RSM}) \geq P_{\boldsymbol{D}}(\boldsymbol{A}^{'}\boldsymbol{D}_{\Delta} \geq \boldsymbol{A}^{'}\boldsymbol{D} - \epsilon) \geq P_{\boldsymbol{D}}(\left|\boldsymbol{A}^{'}\boldsymbol{D}_{\Delta} - \boldsymbol{A}^{'}\boldsymbol{D}\right| \leq \epsilon) \to 1.$$

Thus, RSM is a consistent test.

For a sequence of local alternatives $H_n : \theta_j = \theta_0 + \theta c_j n^{-1/2}$, $(\theta > 0, j = 1, \ldots, k)$. Under H_n, and F has a density f with $\int f^2(x)dx < \infty$,

$$\Delta_{ij} = (c_j - c_i)\theta n^{-1/2}$$

$$\boldsymbol{D}_n = \boldsymbol{C}\theta n^{-1/2},$$

where $\boldsymbol{C}$ is $p \times 1$ vector of $(c_j - c_i)'s$.

Suppose $\boldsymbol{D}_{\Delta 0}$ is the MWW estimate of $\boldsymbol{D}$ under H_0, and $\boldsymbol{D}_{\Delta n}$ is the MWW estimate of $\boldsymbol{D}$ under H_n. Since $\boldsymbol{D}_{\Delta}$ is translation equivariant, thus,

$$\boldsymbol{D}_{\Delta n} = \boldsymbol{D}_{\Delta 0} + \boldsymbol{D}_n,$$

and

$$\mathrm{Var}(\boldsymbol{D}_{\Delta n}) = \mathrm{Var}(\boldsymbol{D}_{\Delta 0}).$$

Thus, when H_n is true,

$$\mathrm{E}(RSM) = E(\boldsymbol{A}^{'}\boldsymbol{D}_{\Delta n}) = \boldsymbol{A}^{'}(\boldsymbol{0} + \boldsymbol{D}_n) = \boldsymbol{A}^{'}\boldsymbol{D}_n$$

$$\mathrm{Var}(RSM) = \sigma^2_{RSM}.$$

The asymptotic power of the test based on RSM is given by

$$\begin{aligned} P_{\boldsymbol{D}_n}[RSM^* \geq Z_\alpha] &= P_{\boldsymbol{D}_n}[RSM \geq Z_\alpha \sigma_{RSM}] \\ &= P_{\boldsymbol{D}_n}[\frac{RSM - \boldsymbol{A}^{'}\boldsymbol{D}_n}{\sigma_{RSM}} \geq Z_\alpha - \boldsymbol{A}^{'}\boldsymbol{D}_n/\sigma_{RSM}] \\ &= 1 - \Phi(Z_\alpha - \theta c_{\varphi 2}), \end{aligned}$$

where $c_{\varphi 2}$ is the Pitman efficacy of the test. We can write $c_{\varphi 2}$ as

$$\begin{aligned} c_{\varphi 2} &= \boldsymbol{A}^{'}\boldsymbol{C}/\sqrt{k\tau_\varphi^2 \sum\sum_{i<j}(a_j - a_i)(\frac{a_j}{\lambda_j} - \frac{a_i}{\lambda_i})} \\ &= \frac{k^{1/2}\sum_{i=1}^k a_i c_i}{\tau_\varphi[\sum\sum_{i<j}(a_j - a_i)(\frac{a_j}{\lambda_j} - \frac{a_i}{\lambda_i})]^{1/2}} \\ &= \frac{k^{1/2}\sum_{i=1}^k a_i c_i}{\tau_\varphi[k\sum_{i=1}^k \frac{a_i^2}{\lambda_i}]^{1/2}} \\ &= \frac{\sum_{i=1}^k a_i c_i}{[\sum_{i=1}^k \frac{a_i^2}{\lambda_i}]^{1/2}}(12)^{1/2}\int f^2(x)dx. \end{aligned}$$

Since

$$\begin{aligned} \sum\sum_{i<j}(a_j - a_i)(\frac{a_j}{\lambda_j} - \frac{a_i}{\lambda_i}) &= \sum\sum_{i<j}(\frac{a_i^2}{\lambda_i} + \frac{a_j^2}{\lambda_j} - a_i a_j(\frac{1}{\lambda_i} + \frac{1}{\lambda_j})) \\ &= (k-1)\sum_{i=1}^k \frac{a_i^2}{\lambda_i} - \sum_{i=1}^k \frac{a_i}{\lambda_i}(-a_i) \\ &= \sum_{i=1}^k \frac{a_i^2}{\lambda_i}. \end{aligned}$$

Theorem 3.3. *If $H_n : \theta_j = \theta_0 + \theta c_j n^{-1/2}, (\theta > 0, j = 1, \ldots, k)$, and F has a density f with $\int f^2(x)dx < \infty$, then as $n \to \infty$*

$$RSM^* \xrightarrow{D} Z \sim N(\theta c_\varphi,\ 1).$$

On the other hand, the Pitman efficacy of the Hettmansperger-Norton test is

$$
\begin{aligned}
e &= \frac{\sum_{i=1}^{k} a_i(c_i - \bar{c}_w)}{[\sum_{i=1}^{k} \frac{a_i^2}{\lambda_i}]^{1/2}}(12)^{1/2} \int f^2(x)dx \\
&= \frac{\sum_{i=1}^{k} a_i(c_i - \bar{c}_w)}{[\sum_{i=1}^{k} \frac{a_i^2}{\lambda_i}]^{1/2}}(12)^{1/2} \int f^2(x)dx \\
&= \frac{\sum_{i=1}^{k} a_i c_i - \sum_{i=1}^{k} a_i \bar{c}_w}{[\sum_{i=1}^{k} \frac{a_i^2}{\lambda_i}]^{1/2}}(12)^{1/2} \int f^2(x)dx \\
&= \frac{\sum_{i=1}^{k} a_i c_i}{[\sum_{i=1}^{k} \frac{a_i^2}{\lambda_i}]^{1/2}}(12)^{1/2} \int f^2(x)dx = c_{\varphi 2}.
\end{aligned}
$$

Therefore, the efficiency of RSM relative to Hettmansperger-Norton is $c_{\varphi 2}^2/e^2 = 1$, which means that RSM and HN have the same asymptotic local power (Noether, 1955).

3.3 The RSM with Least Square Estimates (RSM$_{LS}$) Test

The RSM$_{LS}$ test uses the least square estimates $\tilde{\Delta}_{ij} = \bar{Y}_j - \bar{Y}_i$ to estimate the shifts Δ_{ij}. This is our extension of the AT test.

3.3.1 The RSM$_{LS}$ Test Statistic

The RSM$_{LS}$ test statistic is

$$
RSM_{LS} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (a_j - a_i)\tilde{\Delta}_{ij} = \sum \sum_{i<j} a_{ij}\tilde{\Delta}_{ij} = \boldsymbol{A}'\tilde{\boldsymbol{D}},
$$

where the LS estimate of $\Delta_{ij} = \mu_{Y_j} - \mu_{Y_i}$ is $\tilde{\Delta}_{ij} = \bar{Y}_j - \bar{Y}_i$.

$$
\begin{aligned}
Cov(\tilde{\Delta}_{ij}, \tilde{\Delta}_{i'j'}) &= Cov(\bar{Y}_j - \bar{Y}_i, \bar{Y}_{j'} - \bar{Y}_{i'}) \\
&= Cov(\bar{Y}_j, \bar{Y}_{j'}) - Cov(\bar{Y}_j, \bar{Y}_{i'}) - Cov(\bar{Y}_i, \bar{Y}_{j'}) + Cov(\bar{Y}_i, \bar{Y}_{i'}) \\
&= I(j = j')\frac{\sigma_j^2}{n_j} - I(j = i')\frac{\sigma_j^2}{n_j} - I(i = j')\frac{\sigma_i^2}{n_i} + I(i = i')\frac{\sigma_i^2}{n_i}.
\end{aligned}
$$

Let σ^2 be the common variance of all Y's, then

$$Cov(\tilde{\Delta}_{ij}, \tilde{\Delta}_{i'j'}) = \sigma^2[\frac{I(j=j')}{n_j} - \frac{I(j=i')}{n_j} - \frac{I(i=j')}{n_i} + \frac{I(i=i')}{n_i}],$$

Since $n_i, i = 1, \ldots, k$ has the same order as n, thus as $n \to \infty$

$$\sqrt{n_i}(\bar{Y}_i - \mu_{Y_i}) \xrightarrow{D} N(0, \sigma^2),$$

and $\tilde{\Delta}_{ij}$ is approximately $N(\Delta, \sigma^2(1/n_i + 1/n_j))$.

Let $\boldsymbol{V}$ be a $k \times 1$ vector

$$\boldsymbol{V} = \begin{bmatrix} \sqrt{n_1}((\bar{Y}_1 - \mu_{Y_1})) \\ \vdots \\ \sqrt{n_k}((\bar{Y}_k - \mu_{Y_k})) \end{bmatrix},$$

and if $\boldsymbol{L}$ is a $p \times k$ matrix,

$$\boldsymbol{L} = \begin{bmatrix} -\frac{1}{\sqrt{n_1}} & \frac{1}{\sqrt{n_2}} & 0 & 0 & \ldots & 0 & 0 \\ -\frac{1}{\sqrt{n_1}} & 0 & \frac{1}{\sqrt{n_3}} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -\frac{1}{\sqrt{n_{k-1}}} & \frac{1}{\sqrt{n_k}} \end{bmatrix},$$

and $\boldsymbol{D}$ is a $p \times 1$ vector

$$\boldsymbol{D} = \begin{bmatrix} \mu_{Y_2} - \mu_{Y_1} \\ \mu_{Y_3} - \mu_{Y_1} \\ \vdots \\ \mu_{Y_k} - \mu_{Y_{k-1}} \end{bmatrix},$$

then $\tilde{\boldsymbol{D}} = \boldsymbol{LV} + \boldsymbol{D}$ and $RSM_{LS} = \boldsymbol{A}'\boldsymbol{LV} + \boldsymbol{A}'\boldsymbol{D}$, where $\boldsymbol{A}$ is $p \times 1$ vector of $a'_{ij}s$.

Hence,

$$\boldsymbol{A}'\boldsymbol{L}(\boldsymbol{A}'\boldsymbol{L})' = k^2 \sum_{j=1}^{k} \frac{a_j^2}{n_j}$$

which is non-negative.

Since

$$\boldsymbol{A}'\boldsymbol{L}_{1\times k} = k\begin{bmatrix} \frac{a_1}{\sqrt{n_1}} & \frac{a_2}{\sqrt{n_2}} & \dots & \frac{a_k}{\sqrt{n_k}} \end{bmatrix}$$

and since n_j and n have the same order, as $n \to \infty$, every $a_i/\sqrt{n_i} \to 0, i = 1, \dots, k$. Thus,

$$max_{1\le i\le n}\left|\boldsymbol{A}'\boldsymbol{L}_i\right| \to 0, as\ n \to \infty.$$

Based on Corollary A.1.1 (HM Page 447)

$$RSM_{LS} = \boldsymbol{A}'\boldsymbol{L}\boldsymbol{V} + \boldsymbol{A}'\boldsymbol{D} \xrightarrow{D} N(\boldsymbol{A}'\boldsymbol{D},\ \sigma^2_{RSM_{LS}}),$$

where $\sigma^2_{RSM_{LS}} = k^2\sigma^2\sum_{j=1}^{k}\frac{a_j^2}{n_j}$.

When $H_0 : \theta_1 = \dots = \theta_k$ is true, then $E(RSM_{LS}) = 0$. Hence, under H_0,

$$RSM^*_{LS} = \frac{1}{k\sigma\sqrt{\sum_{j=1}^{k}\frac{a_j^2}{n_j}}}(\boldsymbol{A}'\boldsymbol{L}\boldsymbol{V} + \boldsymbol{A}'\boldsymbol{D}) \xrightarrow{D} Z \sim N(0,\ 1). \tag{3.11}$$

The asymptotic test is

Reject H_0 (1.2) in favor of H_A (1.3)

if $RSM^*_{LS} \ge z_\alpha$; otherwise do not reject.

3.3.2 Consistency and Asymptotic Power

For a level α test based on RSM_{LS} for the hypothesis (1.3), if H_1 is true, all elements in $\boldsymbol{D}$ are greater than 0. Since $O(n_j) = O(n)$, $\sigma^2_{RSM_{LS}} = k^2\sigma^2\sum_{j=1}^{k}\frac{a_j^2}{n_j} \to 0$ as $n \to \infty$. If n is sufficiently large, we have

$$0 < z_\alpha\sigma_{RSM_{LS}} < \boldsymbol{A}'\boldsymbol{D} - \epsilon.$$

Hence,

$$\begin{aligned} P_{\boldsymbol{D}}\left[RSM^*_{LS} \ge z_\alpha\right] &= P_{\boldsymbol{D}}\left[\boldsymbol{A}'\tilde{\boldsymbol{D}} \ge z_\alpha\sigma_{RSM_{LS}}\right] \\ &\ge P_{\boldsymbol{D}}\left[\boldsymbol{A}'\tilde{\boldsymbol{D}} \ge \boldsymbol{A}'\boldsymbol{D} - \epsilon\right] \\ &\ge P_{\boldsymbol{D}}\left[\left|\boldsymbol{A}'\tilde{\boldsymbol{D}} - \boldsymbol{A}'\boldsymbol{D}\right| \le \epsilon\right] \to 1, \end{aligned}$$

thus the RSM$_{LS}$ test is consistent.

Under $H_n : \theta_j = \theta_0 + \theta c_j n^{-1/2}, (\theta > 0, j = 1, \ldots, k)$, F has a density f with $\int f^2(x)dx < \infty$, so that

$$\Delta_{ij} = (c_j - c_i)\theta n^{-1/2}$$

$$\boldsymbol{D}_n = \boldsymbol{C}\theta n^{-1/2},$$

where $\boldsymbol{C}$ is $p \times 1$ vector of $(c_j - c_i)'s$. Thus under H_n, $RSM_{LS} \xrightarrow{D} N(\boldsymbol{A}'\boldsymbol{D}_n, \sigma^2_{RSM_{LS}})$. The asymptotic power of the test based on RSM_{LS} is given by

$$\begin{aligned} P_{\boldsymbol{D}_n}[RSM^*_{LS} \geq Z_\alpha] &= P_{\boldsymbol{D}_n}[RSM_{LS} \geq Z_\alpha \sigma_{RSM_{LS}}] \\ &= P_{\boldsymbol{D}_n}[\frac{RSM_{LS} - \boldsymbol{A}'\boldsymbol{D}_n}{\sigma_{RSM_{LS}}} \geq Z_\alpha - \boldsymbol{A}'\boldsymbol{D}_n/\sigma_{RSM_{LS}}] \\ &= 1 - \Phi(Z_\alpha - \theta c_{LS}), \end{aligned}$$

where c_{LS} is the Pitman efficacy of the test, and

$$\begin{aligned} c_{LS} &= A'C/\sqrt{k\sigma^2 \sum\sum_{i<j} (a_j - a_i)(\frac{a_j}{\lambda_j} - \frac{a_i}{\lambda_i})} \\ &= \frac{k^{1/2}\sum_{i=1}^k a_i c_i}{\sigma[\sum\sum_{i<j}(a_j - a_i)(\frac{a_j}{\lambda_j} - \frac{a_i}{\lambda_i})]^{1/2}} \\ &= \frac{k^{1/2}\sum_{i=1}^k a_i c_i}{\sigma[k\sum_{i=1}^k \frac{a_i^2}{\lambda_i}]^{1/2}} \\ &= \frac{\sum_{i=1}^k a_i c_i}{\sigma[\sum_{i=1}^k \frac{a_i^2}{\lambda_i}]^{1/2}}. \end{aligned}$$

3.4 Comparisons of New Methods

In previous sections we showed that for a fixed alternative, the RAT test, the RAThn test, the RSM test and the RSM$_{LS}$ test are all consistent. Thus, the power of these tests tends to one as the sample size increases. In this section, we compare all these methods by using the asymptotic relative efficiency (ARE). Recall that the

Pitman efficacies of four new methods are:

$$\begin{aligned} c_{RAT} &= \frac{\sum h_j c_j}{\sqrt{n}\tau_\varphi(\sum \frac{h_j^2}{n_j})^{1/2}} \\ c_{RAThn} &= \frac{\sum a_j c_j}{\sqrt{n}\tau_\varphi(\sum \frac{a_j^2}{n_j})^{1/2}} \\ c_{RSM} &= \frac{\sum a_j c_j}{\sqrt{n}\tau_\varphi(\sum \frac{a_j^2}{n_j})^{1/2}} \\ c_{RSMls} &= \frac{\sum a_j c_j}{\sqrt{n}\sigma(\sum \frac{a_j^2}{n_j})^{1/2}}. \end{aligned}$$

It's obvious that ARE(RAThn, RSM)=1 and ARE(RAThn, RSMls)=σ^2/τ_φ^2. Also ARE(RSM, RSMls)=ARE(RAThn, RSMls). The ARE of the RAT and RAThn tests is

$$\text{ARE(RAT,RAThn)} = \frac{(\sum h_j c_j)^2}{n\tau_\varphi^2 \sum \frac{h_j^2}{n_j}} / \frac{(\sum a_j c_j)^2}{n\tau_\varphi^2 \sum \frac{a_j^2}{n_j}} = \frac{(\sum h_j c_j)^2 \sum \frac{a_j^2}{n_j}}{\sum \frac{h_j^2}{n_j} (\sum a_j c_j)^2},$$

And for all sample sizes are 5, ARE(RAT,RAThn)=0.9423. Thus, in the following simulation results the RAT test is about 94% efficient as the RAThn test.

3.5 Simulation Results

Ten thousand simulations were run with data which was generated from a normal, right skewed contaminated normal distribution and a *log-F* distribution to get the empirical powers of all methods, respectively. Eight types of the test, JT (Jonckheere, 1954) and (Terpstra, 1952), SP (McKean et al., 2001), AT (Abelson and Tukey, 1963), HN (Hettmansperger and Norton, 1987), RAT, RAThn (RAThnT is RAThn with t test, RAThnN is RAThn with z test), RSM (RSMT is RSM with t test, RSMN is RSM with z test) and RSMls (RSMlsT is RSMls with t test and RSMlsN is RSMls with N test), were compared in this section.

Four simulations with different situations were run: when errors follow a normal distribution $N(0,\ 1)$, when errors follow a symmetric heavy tailed contaminated normal distribution $CN(0.25,\ 100)$, when errors follow a highly right skewed distribution, skewed contaminated normal distribution $SCN(0.25,\ 1,\ 100)$, and when errors follow a heavier-tailed and positively skewed $GF(1,\ 0.1)$ distribution. New methods were run with Wilcoxon scores under a normal distribution. For the remaining three distributions, new methods were run with Wilcoxon scores, and optimal scores.

3.5.1 With $N(0,\ 1)$ Underlying Distribution

Table 10 an Figure 13 show the empirical power for all methods with normal errors. The z test of RSMls (RSMlsN) dominates all other methods, and there is no significant difference among JT, SP, HN, RAThnT, RAThnN, RSMT and RSMN. The RAThn test outperforms the RAT test.

Table 10: Empirical power under a *Normal* distribution for all methods.

Method	Empirical Power			
	0	2.11	8.44	18.99
JT	.1007	.5447	.9261	.9972
SP	.0963	.5374	.9255	.9972
AT	.0985	.5365	.9264	.9968
HN	.0994	.5455	.9278	.9973
RAT	.0976	.5141	.9052	.9945
RAThnT	.1007	.5314	.9163	.9960
RAThnN	.1074	.5478	.9230	.9969
RSMT	.1003	.5286	.9180	.9963
RSMN	.1069	.5459	.9250	.9973
RSMlsT	.1007	.5536	.9365	.9982
RSMlsN	.1083	.5714	.9412	.9984

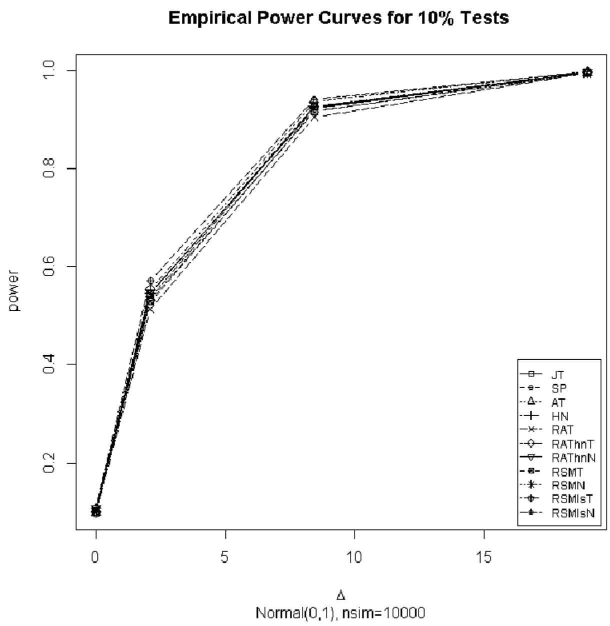

Figure 13: Plot of power curves under a *Normal* distribution for all methods.

3.5.2 With $CN(0.25,\ 100)$ Underlying Distribution

We ran a $CN(0.25,\ 100)$ distribution with both Wilcoxon scores and bent4 (.25,.75,-1,1) scores. Table 12 and Figure 15 show that the RAT test and the RAThn test both have higher power than the JT test when using bent4 scores.

Table 11: Empirical power under $CN(0.25, 100)$ with Wilcoxon scores.

Method	Empirical Power								
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39	135.04
JT	.1109	.3452	.6233	.8332	.9150	.9579	.9784	.9888	.9946
SP	.1067	.3357	.6052	.8097	.8914	.9371	.9616	.9752	.9851
AT	.1037	.1729	.2622	.3845	.4886	.5887	.6736	.7626	.8283
HN	.1073	.3149	.5753	.7796	.8586	.9149	.9393	.9620	.9722
RAT	.0730	.2176	.4594	.7091	.8426	.9233	.9588	.9741	.9849
RAThnT	.0763	.2291	.4728	.7245	.8522	.9275	.9599	.9757	.9867
RAThnN	.0840	.2437	.4927	.7383	.8615	.9275	.9619	.9777	.9875
RSMT	.0986	.2551	.4831	.7171	.8356	.9097	.9444	.9654	.9814
RSMN	.1054	.2681	.5030	.7302	.8440	.9145	.9478	.9675	.9819
RSMlsT	.1053	.1697	.2582	.3705	.4666	.5731	.6523	.7473	.8041
RSMlsN	.1135	.1801	.2709	.3859	.4826	.5897	.6644	.7603	.8144

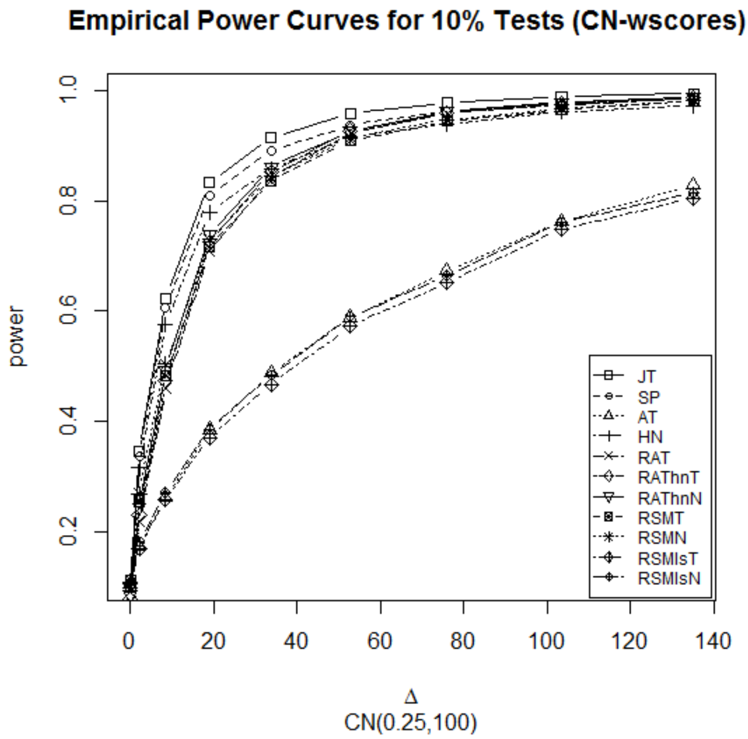

Figure 14: Plot of power curves under CN distribution with Wilcoxon scores.

Table 12: Empirical power under $CN(0.25, 100)$ with bent4 scores.

Method	Empirical Power								
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39	135.04
JT	.1014	.3421	.6341	.8283	.9225	.9638	.9801	.9882	.9916
SP	.0969	.3321	.6187	.8046	.8987	.9442	.9632	.9747	.9803
AT	.1010	.1831	.2724	.3790	.4939	.5917	.6872	.7584	.8146
HN	.0977	.3201	.5846	.7732	.8661	.9192	.9453	.9606	.9664
RAT	.1069	.3327	.6302	.8372	.9384	.9733	.9839	.9898	.9918
RAThnT	.1128	.3474	.6492	.8488	.9394	.9723	.9833	.9896	.9918
RAThnN	.1208	.3621	.6644	.8585	.9429	.9736	.9835	.9901	.9918
RSMT	.1640	.3864	.6354	.8100	.9026	.9471	.9681	.9790	.9864
RSMN	.1718	.3981	.6488	.8177	.9060	.9493	.9649	.9798	.9868
RSMlsT	.0989	.1763	.2676	.3738	.4757	.5686	.6681	.7387	.8023
RSMlsN	.1077	.1878	.2797	.3910	.4900	.5821	.6818	.7534	.8123

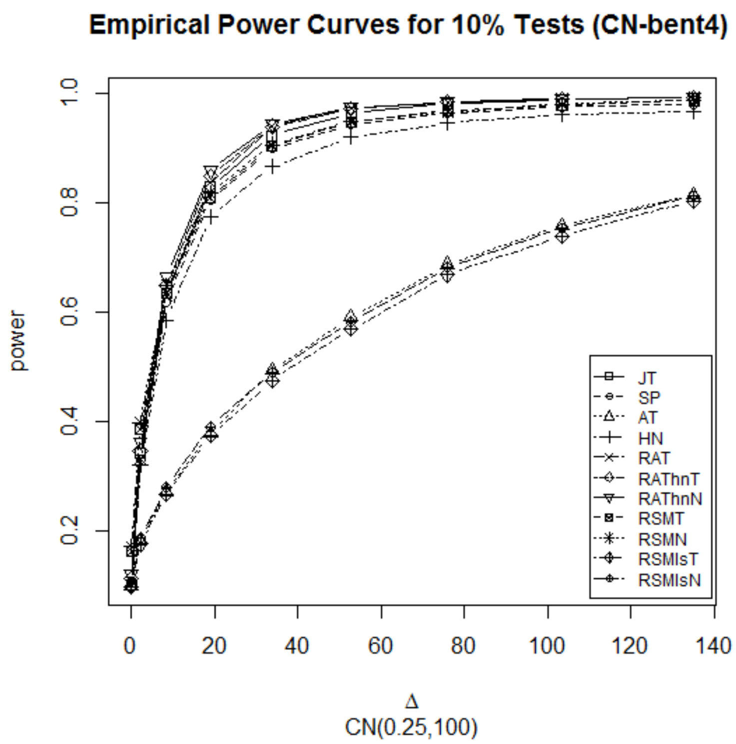

Figure 15: Plot of power curves under CN distribution with bent4 scores.

3.5.3 With $SCN(0.25,\ 1,\ 100)$ Underlying Distribution

Table 13 and Figure 16 show that under SCN(0.25,1,100) RAT, RAThn and RSM with Wilcoxon scores outperform AT and RSMls. Among RAT, RAThn and RSM, the z test RAThn (RAThnN) has the highest power.

Table 13: Empirical power under $SCN(0.25, 1, 100)$ with Wilcoxon scores.

Method	Empirical Power							
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39
JT	.1005	.3297	.6139	.8144	.8976	.9431	.9693	.9789
SP	.0968	.3195	.5938	.7887	.8697	.9145	.9409	.9555
AT	.0967	.1519	.2170	.2866	.3737	.4543	.5351	.6127
HN	.1004	.3263	.6015	.7945	.8738	.9184	.9431	.9569
RAT	.0686	.1758	.3722	.5963	.7562	.8564	.9158	.9406
RAThnT	.0768	.1967	.3942	.6161	.7666	.8591	.9140	.9322
RAThnN	.0814	.2056	.4130	.6320	.7816	.8676	.9192	.9354
RSMT	.0864	.2106	.4010	.6150	.7610	.8509	.9084	.9303
RSMN	.0916	.2233	.4187	.6304	.7742	.8592	.9134	.9332
RSMlsT	.1000	.1562	.2295	.2998	.3831	.4641	.5501	.6224
RSMlsN	.1084	.1684	.2414	.3136	.3971	.4794	.5653	.6378

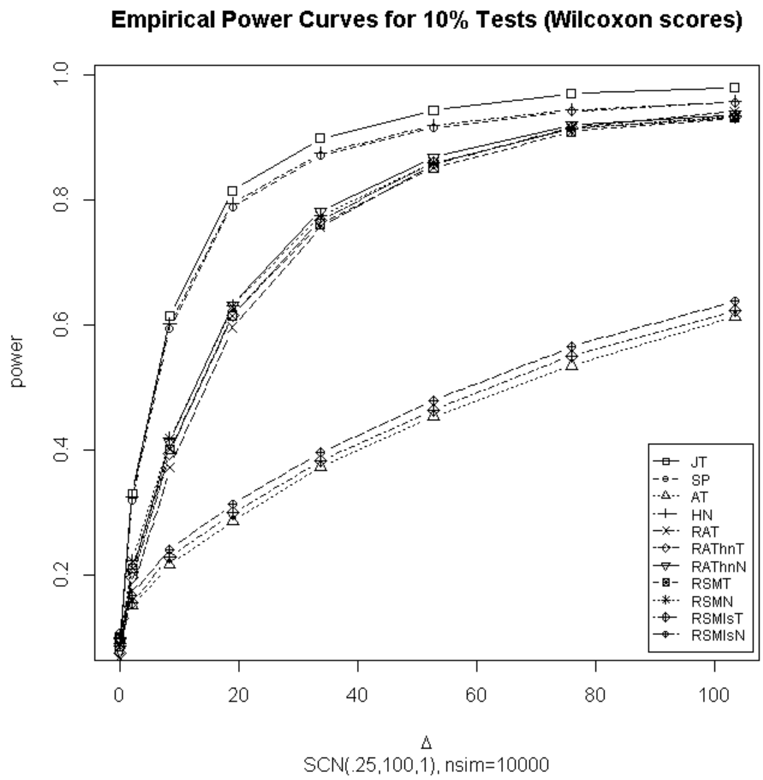

Figure 16: Plot of power curves under SCN distribution with Wilcoxon scores.

Table 14 and Figure 17 show the empirical powers of RAT, RAThn and RSM wiht bent1 scores under SCN(0.25,1,100). The RAThn has the highest power of all methods, and has no significant difference with JT.

Table 14: Empirical power under $SCN(0.25, 1, 100)$ with bent1 scores.

Method	Empirical Power									
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39	135.04	170.91
JT	.1043	.3310	.6148	.8018	.9005	.9413	.9632	.9759	.9834	.9893
SP	.1005	.3222	.5989	.7757	.8703	.9146	.9368	.9558	.9624	.9718
AT	.1002	.1555	.2167	.2848	.3656	.4539	.5283	.6078	.6773	.7346
HN	.1057	.3282	.6079	.7817	.8747	.9167	.9385	.9572	.9632	.9736
RAT	.0909	.2599	.5345	.7496	.8777	.9291	.9514	.9686	.9742	.9798
RAThnT	.0973	.2781	.5564	.7633	.8816	.9288	.9505	.9686	.9743	.9820
RAThnN	.1028	.2910	.5718	.7729	.8847	.9327	.9524	.9696	.9755	.9827
RSMT	.1431	.3151	.5614	.7412	.8470	.8997	.9279	.9488	.9571	.9681
RSMN	.1507	.3276	.5742	.7504	.8543	.9035	.9308	.9504	.9577	.9687
RSMlsT	.1016	.1563	.2275	.2962	.3742	.4662	.5441	.6224	.6857	.7527
RSMlsN	.1101	.1677	.2383	.3100	.3882	.4819	.5608	.6364	.7008	.7654

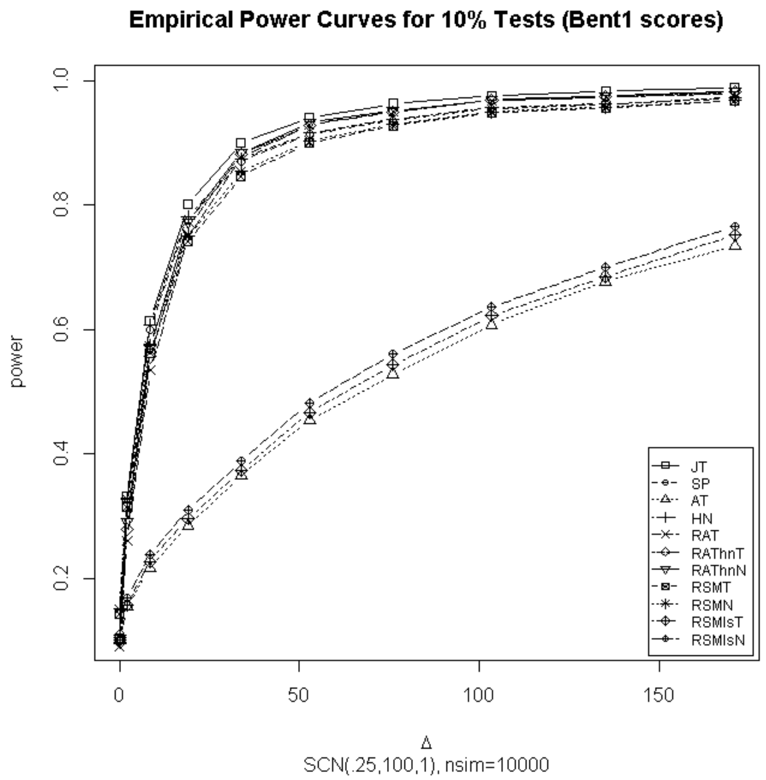

Figure 17: Plot of power curves under SCN distribution with bent1 scores.

3.5.4 With $GF(1, .1)$ Underlying Distribution

If the error distribution is contained in the log-F class, e has a $GF(2m_1, 2m_2)$ distribution. If $(m_1, m_2) = (1, 1)$, then the e has a logistic distribution. Kalbfleisch and Prentice (Kalbfleisch and Prentice, 1980) discussed the class for m_1, $m_2 \geq 1$, Hettmansperger and McKean (Hettmansperger and McKean, 2011) extended the class to m_1, $m_2 > 0$ for heavier-tailed error distributions. Errors in this section were generated from a heavy-tailed and positively skewed distribution $GF(1, .1)$. Wilcoxon scores, logGF(1,.1) scores and bent1(.25,-2,1) scores were used in the simulation. In this section, only JT, RAT and RAThn were compared.

Table 15 and Figure 18 show the level and power of JT, RAT, RAThnT and RAThnN with Wilcoxon scores. The power of JT is slightly higher than RAT, RAThnT and RAThnN.

Table 15: Empirical power under $GF(1, 0.1)$ with Wilcoxon scores.

Method	Empirical Power								
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39	135.04
JT	.1076	.1463	.2064	.2551	.3249	.3969	.4774	.5464	.6186
RAT	.1025	.1338	.1783	.2186	.2726	.3420	.4033	.4807	.5512
RAThnT	.1033	.1365	.1848	.2260	.2872	.3581	.4213	.4963	.5636
RAThnN	.1103	.1444	.1956	.2373	.3020	.3704	.4388	.5142	.5814

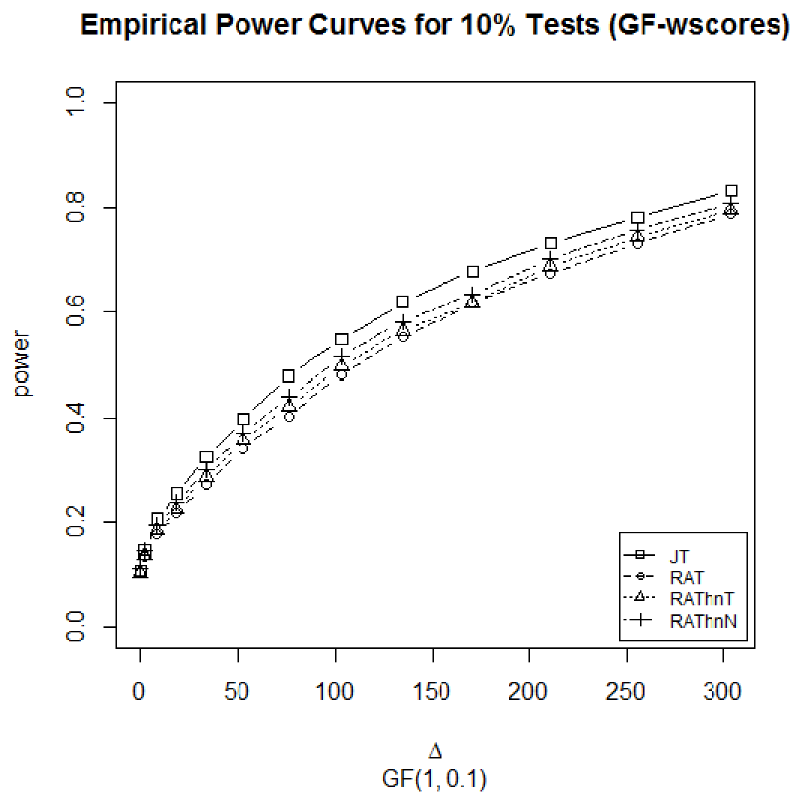

Figure 18: Plot of power curves under GF distribution with Wilcoxon scores.

Table 16 and Figure 19 show the level and power of JT, RAT, RAThnT and RAThnN with loghGF scores. The powers of RAT, RAThnT and RAThnN are higher than JT, but the empirical α levels are liberial.

Table 16: Empirical power under $GF(1, 0.1)$ with logGF scores.

Method	Empirical Power								
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39	135.04
JT	.1046	.1471	.2004	.2596	.3259	.3956	.4692	.5402	.6166
RAT	.1316	.1897	.2678	.3552	.4520	.5566	.6475	.7359	.8078
RAThnT	.1365	.1947	.2775	.3687	.4703	.5731	.6627	.7501	.8189
RAThnN	.1440	.2060	.2910	.3824	.4837	.5868	.6768	.7633	.8289

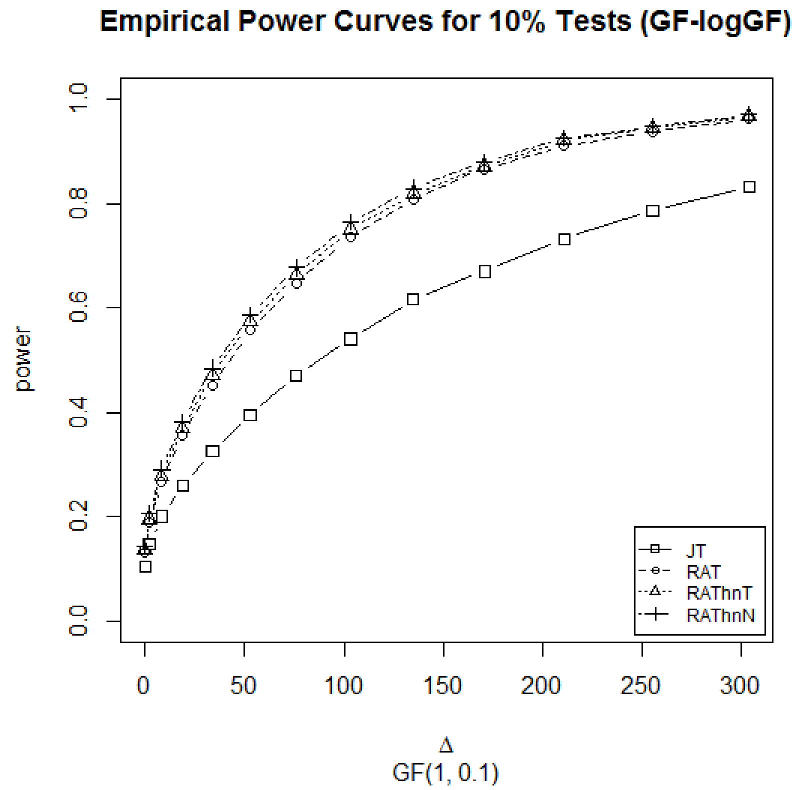

Figure 19: Plot of power curves under GF distribution with logGF scores.

Table 17 and Figure 20 show the level and power of JT, RAT, RAThnT and RAThnN with bent1 scores. The power of RAT, RAThnT and RAThnN are higher than JT. RAThnN has the highest power.

Table 17: Empirical power under $GF(1, 0.1)$ with bent1 scores.

Method	Empirical Power								
	0	2.11	8.44	18.99	33.76	52.75	75.96	103.39	135.04
JT	.1053	.1478	.2054	.2555	.3274	.4012	.4637	.5497	.6194
RAT	.0961	.1504	.2104	.2844	.3811	.4790	.5817	.6770	.7593
RAThnT	.1006	.1590	.2215	.2928	.3990	.4914	.5990	.6977	.7718
RAThnN	.1063	.1682	.2324	.3060	.4160	.5099	.6139	.7144	.7862

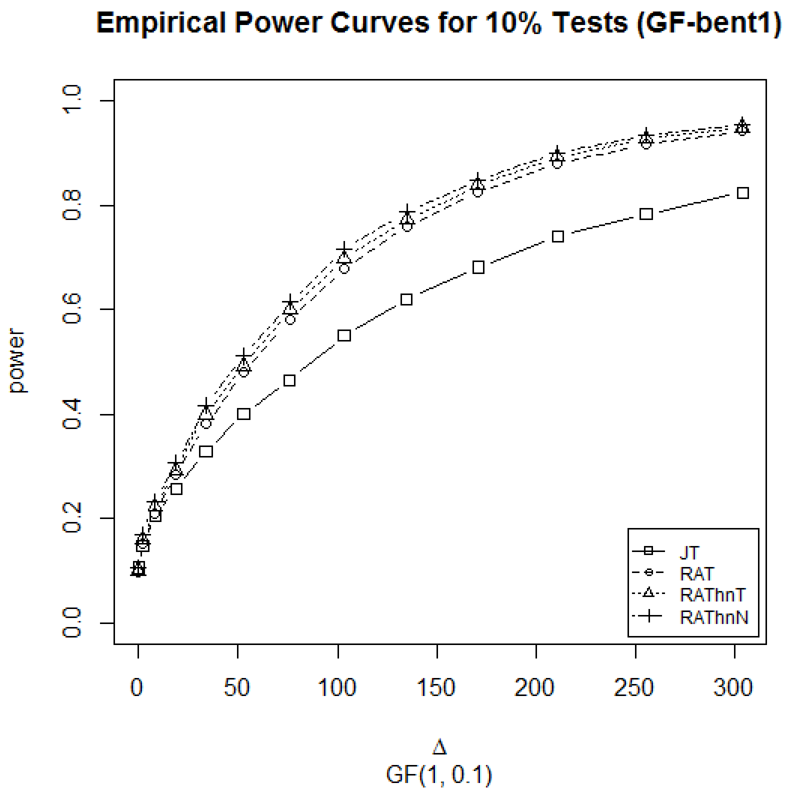

Figure 20: Plot of power curves under GF distribution with bent1 scores.

In conclusion, for the parametric-based methods, the RSM_{LS} test is better than the AT test; for the nonparametric-based methods, if the errors follow a right skewed distribution, there is no significant difference among the JT test, the RAT test, the RAThn test and the RSM test, if the errors follow a right skewed and heavy-tailed distribution, the RAT test and the RAThn test are better than the JT test.

3.6 Example

For the example in chapter 1, the new methods test results are given in the following table. RAThn and RSM return robust results, with both p-values less than 0.05. The RSMls result is affected by the outlier, and the p-value is 0.1989, thus it fails to reject the null hypothesis.

Table 18: Test results-new methods

method	test statistic	p-value
RAThn	1.8174	0.0446
RSM	1.8932	0.0389
RSMls	0.8703	0.1989

CHAPTER 4

MIXED MODELS WITH RANDOM BLOCK EFFECT

In the previous chapters we discussed the methods for an ordered alternative in a one-way linear model. In this chapter we will extend the methods for ordered alternative to a mixed model with two factors. The one factor represents the fixed effect, i.e., the groups, while the second factor is a random block effect. In this case, the observations within each block are dependent random variables. In such designs, the treatment effects have an ordering, such as a increasing intensity of drugs. All the hypothesized ordering information is lost if a Chi-square test or a Friedman test is used. Not considering the order of the hypothesis will cause the loss of efficiency. Page (1963) proposed a test which is based on the weighted rank sums where the ranks are the joint ranking of the observations within blocks. Hettmansperger (1975) proposed a multiple comparison procedure which extended the test statistic of Page and provided the point and interval estimates of the cell location parameters. For more discussion see Hollander (1967), Barlow et al. (1972) and Hollander and Wolfe (1999).

LS estimators and joint-rank estimators are used in the new methods. The method with Abelson-Tukey weights and LS estimators is called Abelson Tukey test with randomized blocks (ATb). The method with Abelson-Tukey weights and joint-rank estimators is called JrRAT. If this method contains the block effects in the design matrix, it is called JrRATb. The method with Hettmansperger-Norton weights and joint-rank estimators is called JrRAThn. If this method contains the block effects in the design matrix, it is called JrRAThnb.

Linear mixed-effects model function *lme* was used to get the LS estimates. The estimates of the fixed effects based on the joint ranks (JR) of all residuals in a linear model, with the block correlated continuous error distributions for general score functions is our rank-based fitting procedure. This is the rank-based analog of the traditional LS randomized block design fit. For computation we use the package JRFit (Kloke and McKean, 2014) and (Kloke et al., 2009).

4.1 Simple Mixed Models

Suppose there are k groups and m blocks, where block i has n_i observations; Within block i, let $\boldsymbol{Y_i}$, $\boldsymbol{X_i}$ and $\boldsymbol{e_i}$ denote the $n_i \times 1$ vector of responses, the $n_i \times p$ design matrix, and the $n_i \times 1$ vector of errors, respectively. Let $\mathbf{1}_{\boldsymbol{n_i}}$ denote the vector of n_i ones. Then the simple mixed model could be written as

$$\boldsymbol{Y_i} = \alpha \mathbf{1}_{\boldsymbol{n_i}} + \boldsymbol{X_i} \boldsymbol{\beta} + \mathbf{1}_{\boldsymbol{n_i}} b_i + \epsilon_i, i = 1, \ldots, m, \tag{4.1}$$

where $\boldsymbol{\beta}$ is the $k-1 \times 1$ vector of regression coefficients and α is the intercept parameter. The components of ϵ_i are independent and identically distributed and b_i is a continuous random variable which is independent of ϵ_i. Define $\boldsymbol{e_i} = \mathbf{1}_{\boldsymbol{n_i}} b_i + \epsilon_i$. Assume that the random effects $b_1, \ldots, b_m$ are independent and identically distributed.
Note that we can also write the model as

$$\boldsymbol{y_{ij}} = \theta_j + b_i + \epsilon_{ij},$$

where θ_j is the mean (or median) for group j. As in the previous chapters, we are interested in ordered alternatives, so we began with ordered alternatives of the form

$$H_A : \theta_1 \leq \cdots \leq \theta_k, \text{ (with at least one strict inequality).} \tag{4.2}$$

4.2 The Page Test

The usual distribution-free test for ordered alternatives is a randomized complete block design known as the Page test. To compute the Page statistic, we first order the k observations within each block. Let r_{ij} denote the rank of y_{ij} in the joint ranking of the observations in the ith block and set

$$R_j = \sum_{i=1}^{m} r_{ij}. \tag{4.3}$$

Then the Page statistic L is the weighted rank sums and given by

$$L = \sum_{j=1}^{k} jR_j. \tag{4.4}$$

Notice that this is a Spearman type test statistic.

To test H_0 (1.2) versus the ordered alternative H_1 (1.3) at the α level of significance, the decision rule of Page's test is

Reject H_0 if $L \geq l_\alpha$; otherwise do not reject.

The test statistic is distribution free under H_0. Based on distribution of L under H_0, approimate values of l_α are given in Table A.23 of Hollander & Wolfe (1999). Under H_0 the expected value and the variance of L are

$$E_0(L) = \frac{nk(k+1)^2}{4}$$

$$var_0(L) = \frac{nk^2(k+1)(k^2-1)}{144}.$$

The standardized version of L is

$$L^* = \frac{L - E_0(L)}{\sqrt{var_0(L)}}.$$

When H_0 is true, and as n tends to infinity, L^* has an asymptotic $N(0,1)$. Hence, the asymptotic test is

Reject H_0 if $L^* \geq z_\alpha$; otherwise do not reject.

4.3 New Methods

In this section two main types of methods, parametric and nonparametric methods, are introduced. The parametric method is based on the Abelson Tukey test, utilizing LS estimators. We label it the ATb test (Abelson Tukey test with randomized blocks). The robust methods are based on the joint-rank estimators of β in model 4.1; see (Kloke et al., 2009).

The general form of the test statistic of the new method is

$$T^* = \frac{\boldsymbol{w}'\hat{\boldsymbol{\theta}}}{\sqrt{\boldsymbol{w}' var(\hat{\theta})\boldsymbol{w}}}, \tag{4.5}$$

where $\boldsymbol{w}'$ is a $k \times 1$ vector of weights (the Abelson Tukey weights and the Hettmansperger-Norton weights are used in this chapter), and $\hat{\boldsymbol{\theta}}$ is the JR estimation of $\boldsymbol{\theta}$, the $k \times 1$ vector of group centers.

To estimate $\hat{\boldsymbol{\theta}}$ we need to fit the linear mixed model

$$\boldsymbol{Y} \quad = \quad \boldsymbol{X\beta} + \boldsymbol{1_n b} + \boldsymbol{\epsilon}. \tag{4.6}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_m)'$, $\boldsymbol{X} = (\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_m)'$, $\boldsymbol{b} = (b_1, \ldots, b_m)$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$. Let $\hat{\boldsymbol{\beta}}$ denote an estimation of $\boldsymbol{\theta}$ based on a fitting procedure.
Let

$$\hat{\boldsymbol{\theta}} = \boldsymbol{E}\hat{\boldsymbol{\beta}} \tag{4.7}$$

where

$$\boldsymbol{E}_{k\times k} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \end{bmatrix}. \tag{4.8}$$

Then a test statistic of the hypothesis (4.2) is given by

$$T^* = \frac{\boldsymbol{w}'\boldsymbol{E}\hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{w}'\boldsymbol{E}\Sigma_{\hat{\boldsymbol{\beta}}}\boldsymbol{E}'\boldsymbol{w}}}, \tag{4.9}$$

where $\hat{\boldsymbol{\beta}}$ is the estimate of fixed effects, and $\Sigma_{\hat{\boldsymbol{\beta}}}$ is the variance-covariance matrix of the fixed effects.

Based on different weights or different estimators of $\boldsymbol{\beta}$, the following methods are generated.

4.3.1 Abelson Tukey Test with Randomized Blocks(ATb)

ATb uses the Abelson-Tukey weights ($\boldsymbol{c}$) where the estimates of $\boldsymbol{\beta}$ are computed using the *lme* (Linear Mixed-Effects Models)function in the *nlme* (Linear and Nonlinear Mixed Effects Models) package. The manual on *lme* states that "this function fits a linear mixed-effects model in the formulation described in Laird and Ware (1982) but allowing for nested random effects. The within-group errors are allowed to be correlated and/or have unequal variances." To compute the estimates and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{\boldsymbol{LS}}$, the model was fitted by default method, restricted maximum likelihood (REML), which in effect corrects the maximum-likelihood estimator for degrees of freedom. The test statistic is

$$T^* = \frac{\boldsymbol{c}'\boldsymbol{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{LS}}}{\sqrt{\boldsymbol{c}'\boldsymbol{E}\Sigma_{\hat{\boldsymbol{\beta}}_{LS}}\boldsymbol{E}'\boldsymbol{c}}}. \tag{4.10}$$

For testing H_0 (1.2) versus the ordered alternative H_1 (1.3), large values of T^* lead to the conclusion H_1. At the α level of significance, the decision rule is

Reject H_0 if $T^* \geq t_{\alpha,n-(k+m-1)}$; otherwise do not reject.

Then two ways were used to compute the p-value

1. t test with $df = n - k - m - 1$

   - n: # of observations
   - m: # of blocks

- k: # of treatment groups
- 1: # of additional parameters estimated in variance-covariance of $\hat{\boldsymbol{\beta}}$

2. t test with $df = n - (k + m - 1)$ [only differs by 2].

4.3.2 Joint Rank Estimator

Joint rank estimators are R estimators based on the joint ranks of all the residuals for fitting linear models with independently distributed errors. Kloke, McKean, and Rashid (2009) extended those estimators to estimating the fixed effects in a linear model with cluster correlated continuous error distributions for general score functions. They called it the JR estimator for joint ranking. For model 4.1, they showed that the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}_\varphi$ is

$$\tau_\varphi^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_{\varphi,i}\boldsymbol{X}_i(\boldsymbol{X}'\boldsymbol{X})^{-1}, \boldsymbol{\Sigma}_{\varphi,i} = (1-\rho_\varphi)\boldsymbol{I}_{n_i} + \rho_\varphi\boldsymbol{J}_{n_i}, \tag{4.11}$$

where $\rho_\varphi = cov\left\{\varphi[F(e_{11})], \varphi[F(e_{12})]\right\} = E\left\{\varphi[F(e_{11})], \varphi[F(e_{12})]\right\}$.
Let $M = \sum_{i=1}^{m}\binom{n_i}{2} - (k-1)$. A simple moment estimator of ρ_φ is

$$\hat{\rho}_\varphi = M^{-1}\sum_{i=1}^{m}\sum_{l>j} a\left[R(\hat{e}_{il})\right] a\left[R(\hat{e}_{ij})\right]. \tag{4.12}$$

Thus an estimate of the asymptotic covariance matrix of the JR estimators is easy to get when plugging in the estimate of τ_φ and the estimate of ρ_φ (4.11) (Kloke et al., 2009).

Kloke and McKean (Kloke and McKean, 2014) have developed an R package jrfit, which returns the rank-based estimation and inference for mixed models. Like the Rfit package, jrfit provides variaty of score functions, and the default option is Wilcoxon scores. For our computation, we used version 0.03. This package can be found at *https://www.biostat.wisc.edu/ kloke/*.

4.3.3 Nonparametric Methods

In the following two subsections, all new methods are based on the joint-rank estimates of $\boldsymbol{\beta}$ and the variance-covariance matrix of $\boldsymbol{\beta}$. Two types of the design matrixes are performed in later simulation studies.

1. The design matrix is only composed of the group effect.

$$\boldsymbol{X}_{1,n\times(k-1)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_3} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_k} \end{bmatrix}. \tag{4.13}$$

The centered design matrix based on $\boldsymbol{X}_1$, we denote as $\boldsymbol{D}_1$.

2. The random blocks effects are also contained in the $(k+m-1)\times(k+m-1)$ design matrix given by

$$\boldsymbol{E_2} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{bmatrix}. \tag{4.14}$$

Let

$$\boldsymbol{X}_{2,n\times(k+m-1)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{B} \\ \mathbf{1}_{n_2} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{B} \\ \mathbf{0} & \mathbf{1}_{n_3} & \dots & \mathbf{0} & \boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_k} & \boldsymbol{B} \end{bmatrix} \tag{4.15}$$

where $\boldsymbol{B}$ is

$$\boldsymbol{B}_{n_j \times (m-1)} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \tag{4.16}$$

with $j = 1, \dots, k$. Denote by $\boldsymbol{D}_2$, the centered design matrix obtained from $\boldsymbol{X}_2$.

In the following methods, two procedures are used to estimate the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$; these are sandwich estimation and compound symmetry structure. The sandwich estimation procedure yields asymptotically consistent covariance matrix estimates. In the compound symmetry structure, all the variances are equal and all the covariances are equal. In the $jrfit$ function the default variance-covariance matrix is the sandwich estimation.

The following two methods use the Abelson-Tukey weights $\boldsymbol{c}$ along with the joint-rank estimates of $\boldsymbol{\beta}$. The test statistic with the different design matrix will be slightly changed:

1. JrRAT

   The JrRAT method uses design matrix $\boldsymbol{D}_1$, and the test statistic is

$$T_1^* = \frac{\boldsymbol{c}' \boldsymbol{E} \hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{c}' \boldsymbol{E} \Sigma_{\hat{\boldsymbol{\beta}}} \boldsymbol{E}' \boldsymbol{c}}}. \tag{4.17}$$

2. JrRATb

   The JrRATb method uses design matrix $\boldsymbol{D}_2$, and the test statistic is

$$T_2^* = \frac{\boldsymbol{c}' \boldsymbol{E_2} \hat{\boldsymbol{\beta}}_b}{\sqrt{\boldsymbol{c}' \boldsymbol{E}_2 \Sigma_{\hat{\boldsymbol{\beta}}_b} \boldsymbol{E}_2' \boldsymbol{c}}}. \tag{4.18}$$

   The subscript b here is for fitting fixed and blocks.

   The following two methods use the Hettmansperger-Norton weights along with the joint-rank estimates of $\boldsymbol{\beta}_1$. Note that $\boldsymbol{\beta} = (\alpha, \boldsymbol{\beta}_1')'$.

3. JrRAThn

   The JrRAThn method uses design matrix $\boldsymbol{D}_1$, and the test statistic is

$$T_3^* = \frac{k\boldsymbol{a}'\hat{\boldsymbol{\beta}}_1}{k\sqrt{\boldsymbol{a}'\Sigma_{\hat{\boldsymbol{\beta}}_1}\boldsymbol{a}}}. \tag{4.19}$$

4. JrRAThnb

   The JrRAThnb method uses design matrix $\boldsymbol{D}_2$, and the test statistic is

$$T_4^* = \frac{k\boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1b}}{k\sqrt{\boldsymbol{a}'\Sigma_{\hat{\boldsymbol{\beta}}_{1b}}\boldsymbol{a}}}. \tag{4.20}$$

   Where $\Sigma_{\hat{\boldsymbol{\beta}}_1} = \Sigma_{\hat{\boldsymbol{\beta}}[2:k,2:k]}$. For both types of the design matrixes, the test statistic formulas are the same, but with different estimates for $\boldsymbol{\beta}$s and the variance covariance matrices.

Asymptotic theory for these test statistics under the null hypothesis and local alternatives follows as in chapter 3, which uses the asymptotic theory derived in Kloke et al. (2009).

For testing H_0 (1.2) versus the ordered alternative H_1 (1.3), large values of T^* lead to the conclusion H_1. For each procedure, at the α level of significance, three tests were compired in the simulation study:

1. Reject H_0 if $T^* \geq Z_\alpha$; otherwise do not reject.

2. Reject H_0 if $T^* \geq t_{\alpha,n-k-m-1}$; otherwise do not reject.

3. Reject H_0 if $T^* \geq t_{\alpha,n-(k+m-1)}$; otherwise do not reject.

4.4 Simulation Studies

In this section, 10,000 simulations were run with four different data sets. These have different block sizes (m) or treatment groups (k). The m random block effects were randomly generated from $N(0, \sigma_b^2)$; the $n = m \times k$ random errors were randomly generated from $N(0, \sigma_\epsilon^2)$. The data set is composed of both the random effects and the random errors, and within each block they have the same block effect. The variance components σ_b^2 and σ_ϵ^2 were set to 4 and 1, respectively. The first three data sets have the same block size, 10, but with different treament groups 3, 4 and 7. The fourth data set consists of 15 blocks and 7 treatment groups. Simulation studies on all methods that were mentioned in previous sections will be summarized.

For the simulation studies, the joint-rank estimate of $\boldsymbol{\beta}$ and the variance-covariance matrix of $\boldsymbol{\beta}$ with model (4.6) were computed by using the R package $jrfit$ (version 0.03).

Table 19 summarizes the empirical alphas and the power of this study with 10 blocks and 3 treatment groups. For both types of design matrix ($\boldsymbol{D}_1$ & $\boldsymbol{D}_2$), JrRAT and JrRAThn return exactly the same results for the two types of variance covariance matrix, respectively. Additionally they are not very good on compound symmetry structure but good on sandwich structure. For the first type of design matrix, the methods with a compound symmetry structure variance covariance matrix are slightly liberal. For the second type of design matrix, both methods with both variance covariance matrix work well except the z test. The ATb method works well on both degrees of freedom, and it can significantly outperform the Page method with an empirical power of .4504 and .4531. This is in contrast to .4166 at level .10. JrRATb and JrRAThnb both outperform the Page test. The ATb method has the highest power among all appropriate methods.

Table 20 summarizes the empirical alphas and the power of this study with 10 blocks and 4 treatment groups. For the first type design matrix ($\boldsymbol{D}_1$), the methods with both compound symmetry structure and sandwich variance covariance matrix are liberal. For the second type design matrix ($\boldsymbol{D}_2$), z test works well with compound symmetry structure variance covariance matrix; additionally the t tests work well with the sandwich variance covariance matrix. The ATb method works well on both degrees of freedom, and it outperforms the Page method with a power of .4817 and .4828, as opposed to .4417 at level .10. JrRATb and JrRAThnb both outperform the Page test. The ATb method has the highest power among all appropriate methods.

Table 21 summarizes the empirical alphas and the power of this study with 10 blocks and 7 treatment groups. For the first type of design matrix ($\boldsymbol{D}_1$), the methods with both compound symmetry structure and sandwich variance covariance matrix are liberal. For the second type of design matrix ($\boldsymbol{D}_2$), the t test works well with the compound symmetry structure variance covariance matrix; all tests with sandwich variance covariance matrix are liberal. The ATb method works well on both degrees of freedom, and it outperforms the Page method with a power of .5650 and .5653, as contrasted with a power of .5382 at level .10. JrRATb and JrRAThnb both outperform the Page test. The JrRAThnb method has the highest power among all appropriate methods.

Table 22 summarizes the empirical alphas and the power of this study with 15 blocks and 7 treatment groups. For the first type of design matrix ($\boldsymbol{D}_1$), the methods with both compound symmetry structure and sandwich variance covariance matrix are liberal. For the second type of design matrix ($\boldsymbol{D}_2$), all tests work well with sandwich variance covariance matrix; all tests with compound symmetry structure variance covariance matrix are liberal. The ATb method works well on both degrees of freedom, and it does better than the Page method with a power of .5502 and .5503, as contrasted with a power of .5318 at level .10. The JrRAThnb test outperforms the Page test, and the JrRAThnb method has the highest power among all appropriate methods.

In conclusion, the JrRAThnb test outperforms the Page test among all cases. Among all nonparametric methods the JrRAThnb is the best. The ATb test is a good least square based test.

Table 19: Empirical αs and power(10x3).

Method	test	Var-Cov	Df	Empirical α		Power	
				10%	5%	10%	5%
	Z			.2035	.1524	.5696	.4784
JrRAT	T	cs	16	.1957	.1413	.5555	.4570
	T		18	.1969	.1425	.5573	.4592
	Z			.0963	.0533	.3981	.2740
JrRAT	T	sand	16	.0904	.0468	.3778	.2462
	T		18	.0908	.0476	.3799	.2489
Page				.1143	.0502	.4166	.2475
	Z			.1018	.0569	.4176	.2876
JrRATb	T	cs	16	.0933	.0485	.3999	.2570
	T		18	.0943	.0493	.4019	.2610
	Z			.1127	.0664	.4342	.3084
JrRATb	T	sand	16	.1044	.0575	.4149	.2795
	T		18	.1050	.0585	.4180	.2830
ATb	T		16	.1017	.0513	.4504	.3102
ATb	T		18	.1031	.0523	.4531	.3147
	Z			.2035	.1524	.5696	.4784
JrRAThn	T	cs	16	.1957	.1413	.5555	.4570
	T		18	.1969	.1425	.5573	.4592
	Z			.0963	.0533	.3981	.2740
JrRAThn	T	sand	16	.0904	.0468	.3778	.2462
	T		18	.0908	.0476	.3799	.2489
	Z			.1018	.0569	.4176	.2876
JrRAThnb	T	cs	16	.0933	.0485	.3999	.2570
	T		18	.0943	.0493	.4019	.2610
	Z			.1127	.0664	.4342	.3084
JrRAThnb	T	sand	16	.1044	.0575	.4149	.2795
	T		18	.1050	.0585	.4180	.2830

Table 20: Empirical αs and power (10x4).

Method	test	Var-Cov	Df	Empirical α		Power	
				10%	5%	10%	5%
	Z			.1724	.1181	.5624	.4502
JrRAT	T	cs	25	.1665	.1114	.5510	.4335
	T		27	.1669	.1119	.5522	.4348
	Z			.0893	.0476	.3963	.2685
JrRAT	T	sand	25	.0842	.0422	.3827	.2488
	T		27	.0845	.0428	.3835	.2502
Page				.1064	.0456	.4417	.2757
	Z			.0993	.0505	.4416	.3139
JrRATb	T	cs	25	.0945	.0438	.4277	.2898
	T		27	.0949	.0441	.4283	.2915
	Z			.1089	.0614	.4570	.3326
JrRATb	T	sand	25	.1034	.0551	.4441	.3108
	T		27	.1040	.0553	.4448	.3126
ATb	T		25	.1038	.0505	.4817	.3385
ATb	T		27	.1044	.0511	.4828	.3399
	Z			.1737	.1196	.5694	.4590
JrRAThn	T	cs	25	.1678	.1123	.5590	.4390
	T		27	.1679	.1127	.5600	.4402
	Z			.0896	.0473	.3994	.2737
JrRAThn	T	sand	25	.0842	.0417	.3850	.2538
	T		27	.0847	.0422	.3858	.2552
	Z			.1001	.0515	.4493	.3225
JrRAThnb	T	cs	25	.0943	.0452	.4373	.2990
	T		27	.0947	.0458	.4376	.3008
	Z			.1098	.0621	.4623	.3396
JrRAThnb	T	sand	25	.1039	.0561	.4482	.3198
	T		27	.1045	.0565	.4488	.3207

Table 21: Empirical αs and power (10x7).

Method	test	Var-Cov	Df	Empirical α		Power	
				10%	5%	10%	5%
	Z			.1406	.0883	.5961	.4760
JrRAT	T	cs	52	.1384	.0852	.5907	.4659
	T		54	.1385	.0853	.5908	.4663
	Z			.0451	.0189	.3190	.1854
JrRAT	T	sand	52	.0435	.0176	.3107	.1784
	T		54	.0436	.0178	.3108	.1785
Page				.1038	.0542	.5382	.3988
	Z			.1040	.0523	.5426	.3965
JrRATb	T	cs	52	.1014	.0493	.5353	.3858
	T		54	.1014	.0496	.5356	.3859
	Z			.1131	.0653	.5487	.4146
JrRATb	T	sand	52	.1094	.0613	.5432	.4053
	T		54	.1094	.0615	.5433	.4059
ATb	T		52	.1026	.0519	.5650	.4151
ATb	T		54	.1026	.0519	.5653	.4153
	Z			.1455	.0921	.6263	.5075
JrRAThn	T	cs	52	.1423	.0888	.6220	.4976
	T		54	.1423	.0889	.6221	.4982
	Z			.0449	.0192	.3413	.2018
JrRAThn	T	sand	52	.0437	.0174	.3342	.1935
	T		54	.0437	.0175	.3345	.1935
	Z			.1045	.0558	.5687	.4263
JrRAThnb	T	cs	52	.1009	.0540	.5620	.4155
	T		54	.1009	.0540	.5624	.4158
	Z			.1159	.0681	.5778	.4429
JrRAThnb	T	sand	52	.1137	.0647	.5717	.4333
	T		54	.1137	.0647	.5719	.4336

Table 22: Empirical αs and power (15x7).

Method	test	Var-Cov	Df	Empirical α		Power	
				10%	5%	10%	5%
	Z			.1221	.0703	.5667	.4426
JrRAT	T	cs	82	.1206	.0684	.5620	.4357
	T		84	.1206	.0684	.5620	.4357
	Z			.0659	.0301	.4156	.2724
JrRAT	T	sand	82	.0646	.0293	.4111	.2667
	T		84	.0646	.0293	.4115	.2670
Page				.1009	.0494	.5318	.3846
	Z			.0918	.0460	.5204	.3730
JrRATb	T	cs	82	.0901	.0439	.5165	.3666
	T		84	.0902	.0439	.5165	.3668
	Z			.1020	.0575	.5262	.3929
JrRATb	T	sand	82	.0999	.0549	.5232	.3867
	T		84	.1001	.0549	.5232	.3867
ATb	T		82	.0968	.0450	.5502	.4048
ATb	T		84	.0968	.0451	.5503	.4049
	Z			.1285	.0716	.5962	.4747
JrRAThn	T	cs	82	.1267	.0694	.5920	.4681
	T		84	.1268	.0694	.5921	.4683
	Z			.0674	.0301	.4450	.2940
JrRAThn	T	sand	82	.0654	.0287	.4401	.2870
	T		84	.0655	.0287	.4402	.2872
	Z			.0924	.0475	.5481	.4096
JrRAThnb	T	cs	82	.0915	.0455	.5438	.4035
	T		84	.0915	.0455	.5440	.4035
	Z			.1021	.0550	.5558	.4180
JrRAThnb	T	sand	82	.1005	.0536	.5513	.4101
	T		84	.1005	.0536	.5513	.4101

CHAPTER 5

UNKNOWN PEAK

Consider a dose-response design. In some cases, with increasing dose level, the treatment effect tends to improve, but after some point, advancing dose level tends to mean diminishing performance. The alternative with such a particular pattern of increasing group locations followed by decreasing group locations is called an umbrella alternative. The change point (peak) may be known, but generally it is not known. In such cases the estimate of the peak is of interest. Mack and Wolfe (1981) used a standardized two-sample Mann-Whitney statistic computed between the ith group and the remaining $k-1$ groups combined to estimate the peak. Hettmansperger and Norton (1987) used the index at where the maximum test statistic occurs to estimate the peak, where the maximum is taken over for all p possible patterns of $c_1, \ldots, c_k$.

Using the methods first introduced in chapter 3, we now explore three methods of estimating the peak. We first describe the methods and then report on the results of a simulation study.

5.1 The Hettmansperger and Norton Method

When the peak is unknown, there are k possible umbrella alternatives, H_A : $\theta_1 \leq \cdots \leq \theta_t \geq \theta_{t+1} \geq \cdots \geq \theta_k$. With equally spaced $c_1, \ldots, c_k$, where $c_j = j$, for $j = 1, \ldots, t$ and $c_j = 2t - j$ for $j = t+1, \ldots, k$, there are p possible sets of coefficients $c_1, \ldots, c_k$. Recall in chapter one that the standardized Hettmansperger and Norton test statistic is

$$V^* = (\frac{12}{n+1})^{1/2} \frac{V}{(\sum a_j^2/\lambda_j)^{1/2}},$$

where $V = \sum_j a_j \bar{R}_j$, $\bar{R}_j = \frac{1}{n_j} \sum_j R_{ij}$, R_{ij} is the value of Y_{ij} in the combined samples $Y_{11}, \ldots, Y_{kn_k}$, $\lambda_j = \frac{n_j}{n}$, and $\sum_j a_j = 0$.

5.2 New Methods

We take equally spaced $c_1, \ldots, c_k$, where $c_j = j$, for $j = 1, \ldots, t$ and $c_j = 2t - j$ for $j = t+1, \ldots, k$. For a given test statistic T_t, we use the index of maxT_t^* to estimate the peak, where the maximum is taken over all possible patterns of $c_1, \ldots, c_k$ and $t = 1, 2, \ldots, p$. We next present three such tests.

5.2.1 RAThn

Recall in chapter 3 that the test statistic of the RAThn is based on the rank-based estimates of the shifts of each group from group 1. Here we use T^*, (6.1), to carry out the test. Thus the T_t^* can be writen as

$$T_t^* = \frac{k\boldsymbol{a}'\hat{\boldsymbol{\beta}}_1}{\sqrt{k^2\tau_\varphi^2 \boldsymbol{a}'(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{a}}}, \tag{5.1}$$

where $\boldsymbol{a}$ is a vector of the HN weights (a_js), and $a_j = \lambda_j(c_j - \sum \lambda_j c_j), j = 1, \ldots, k$.

5.2.2 RSM

Recall in chapter 3 the RSM test statistic is based on the MWW estimates of the pairwise shifts. We use the standardized RSM as the T^*. Then the T_t^* can be writen as

$$T_t^* = \frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(a_j - a_i)\hat{\Delta}_{ij}}{\sqrt{k\tau_\varphi^2 \sum\sum_{i<j}(a_j - a_i)(\frac{a_j}{n_j} - \frac{a_i}{n_i})}}. \tag{5.2}$$

The MWW (or HL) estimate of shift for Δ_{ij} is $\hat{\Delta}_{ij} = med\{Y_{lj} - Y_{l'i}\}$, $1 \le l \le n_j$, $1 \le l' \le n_i$.

5.2.3 RSMls

Recall in chapter 3 the RSM$_{LS}$ uses the LS estimates of the pairwise shifts. We use the standardized RSM$_{LS}$ as the T^*. Then the T_t^* can be writen as

$$T_t^* = \frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(a_j - a_i)\hat{\Delta}_{ij}}{\sqrt{k\sigma_\varphi^2 \sum\sum_{i<j}(a_j - a_i)(\frac{a_j}{n_j} - \frac{a_i}{n_i})}}, \tag{5.3}$$

where the LS estimate of $\Delta = \mu_{Y_j} - \mu_{Y_i}$ is $\hat{\Delta}_{ij} = \bar{Y}_j - \bar{Y}_i$.

5.3 Simulation Studies

We consider the situation where the number of groups k is 10 and the sample size in each group is 16. Two types of data, a set with a clear peak and a set with a flat peak, were generated both from a normal distribution, and a symmetric heavy tailed contaminated normal distribution.

Initially each group is generated from a normal distribution $N(\mu_i, \sigma)$, for $i = 1,\ 2,\ \ldots,\ 10$. We set $\sigma = 1$. The values of μ_i from the clear peak data set were chosen from $1,\ 2,\ \ldots,\ 8,\ 9,\ 10$. For the flat peak data set, the centers of the groups close to the peak should have no significant difference. We set the μ_is closed to the peak with a 0.5 unit difference from each other. Thus, the μ_is from the flat peak data set were chosen from $2,\ 3,\ \ldots,\ 8,\ 9,\ 9.5,\ 10$, and the order of the μ_i depends on where the peak is located.

Secondly, each group is generated from a contaminated normal distribution $CN(\epsilon, \mu_i, \sigma^2)$. We set $\epsilon = 0.25$ and $\sigma^2 = 100$. Thus data is generated from mixed $N(\mu_i, 1)$ and $N(\mu_i, 100)$. The values of μ_i are choosen at the same way as the normal distribution case. That is, the values of μ_i from clear peak data set were chosen from $1,\ 2,\ \ldots,\ 8,\ 9,\ 10$, and the μ_is from the flat peak data set were chosen from $2,\ 3,\ \ldots,\ 8,\ 9,\ 9.5,\ 10$.

We consider all ten situations for the peak; i.e., the peak in the ith group for i. If the peak is located at the 1st group, there is a gradual decrease in the locations of the 10 groups, and the value of μ_i for the respective groups, one through ten, are $10,\ 9,\ \ldots,\ 1$ for the clear peak data set, and are $10,\ 9.5,\ 9,\ \ldots,\ 2$ for the flat peak data set. If the peak is located at the 10th group, there is a gradual increase in the locations of the 10 groups, so the value of μ_i are 1, 2, …, 10 for the clear peak data set and are 2, 3, …, 9, 9.5, 10 for the flat peak data set.

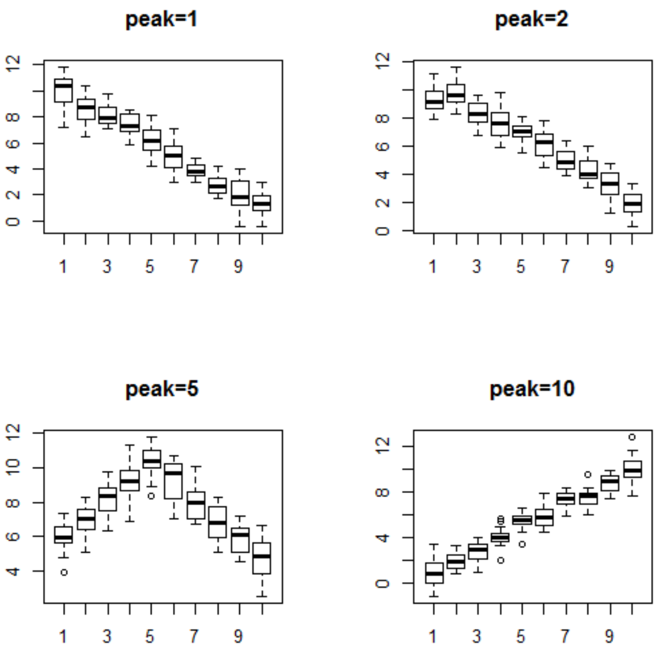

Figure 21: Plots of clear peak data with different peaks.

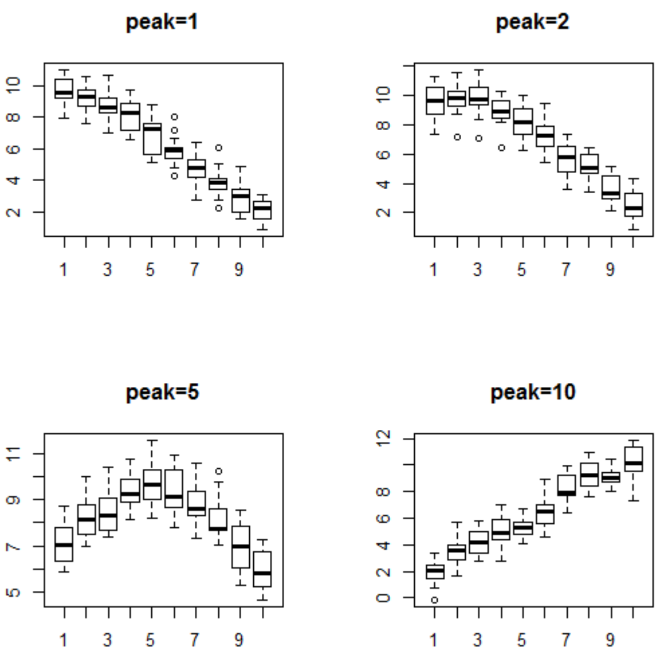

Figure 22: Plots of flat peak data with different peaks.

If the peak is located at the ith group, for $i = 2, \ldots, 9$, the value of μ_i is 10. The value of μ_{i-1} and μ_{i+1} is 9 for the clear peak data set, and is 9.5 for the flat peak data set. The value of μ_{i-2} and μ_{i+2} is 8 for the clear peak data set, and is 9 for the flat peak data set, ..., values are symmetrical on both sides of the peak, and so on. Figure 21 gives sample box plots from the clear peak data set; and figure 22 gives sample box plots from the flat peak data set. All of them are generated from $N(\mu_i, 1)$.

Table 23: Unknown peaks estimation (clear peak $N(\mu_i, 1)$).

Real	Method					# of estimates					
Peak		1	2	3	4	5	6	7	8	9	10
1	HN	9999	1								
	RAThn	9998	2								
	RSM	9998	2								
	RSM_{ls}	9998	2								
2	HN	31	9969								
	RAThn	6	9994								
	RSM	6	9994								
	RSM_{ls}	4	9996								
3	HN		7	9993							
	RAThn			10000							
	RSM		1	9999							
	RSM_{ls}			10000							
4	HN				10000						
	RAThn				10000						
	RSM				10000						
	RSM_{ls}				10000						
5	HN					10000					
	RAThn					10000					
	RSM					10000					
	RSM_{ls}					10000					
6	HN						10000				
	RAThn						10000				
	RSM						10000				
	RSM_{ls}						10000				
7	HN							10000			
	RAThn							10000			
	RSM							10000			
	RSM_{ls}							10000			
8	HN								9996	4	
	RAThn								9999	1	
	RSM								9999	1	
	RSM_{ls}								10000		
9	HN									9960	40
	RAThn									9991	9
	RSM									9990	10
	RSM_{ls}									9994	6
10	HN										10000
	RAThn									1	9999
	RSM									1	9999
	RSM_{ls}									2	9998

Ten thousand simulations were run for each situation under $N(\mu_i, 1)$. In Table 23, for clear peak data, new methods work slightly better than HN. In Table 24, the flat peak simulations, the HN test has an average success rate of 0.9736; the RAThn test has a success rate of 0.9528; the RSM test has an average success rate of 0.9529, and the RSM_{ls} test has an average success rate of 0.9548.

Table 24: Unknown peaks estimation (flat peak under $N(\mu_i, 1)$)

Real Peak	Method	# of estimates									
		1	2	3	4	5	6	7	8	9	10
1	HN	9075	925								
	RAThn	7851	2149								
	RSM	7847	2153								
	RSM_{ls}	7896	2104								
2	HN	106	9877	17							
	RAThn	9	9871	120							
	RSM	9	9870	121							
	RSM_{ls}	9	9889	102							
3	HN		255	9745							
	RAThn		40	9960							
	RSM		40	9960							
	RSM_{ls}		30	9970							
4	HN			30	9970						
	RAThn			21	9979						
	RSM			21	9979						
	RSM_{ls}			15	9985						
5	HN					10000					
	RAThn				1	9999					
	RSM					10000					
	RSM_{ls}				1	9999					
6	HN						10000				
	RAThn						9997	3			
	RSM						9997	3			
	RSM_{ls}						9998	2			
7	HN							9972	28		
	RAThn							9979	21		
	RSM							9979	21		
	RSM_{ls}							9984	16		
8	HN								9747	253	
	RAThn								9952	48	
	RSM								9954	46	
	RSM_{ls}								9962	38	
9	HN								9	9877	114
	RAThn								112	9879	9
	RSM								116	9876	8
	RSM_{ls}								100	9893	7
10	HN									900	9100
	RAThn									2191	7809
	RSM									2177	7823
	RSM_{ls}									2095	7905

For the flat peak situation, however, from the peak at the 2nd group to the peak at the 9th group, the average success rate of HN is 0.9899; of RAThn is 0.9952; of RSM is 0.9952, and the average success rate of RSM_{ls} is 0.9960. Both Table 23 and Table 24 show that the closer the peak is to the center, the more accurate the estimate will be. The closer the peak is to the two sides, the higher the error rate will be.

Table 25: Unknown peaks estimation (clear peak $CN(0.25, \mu_i, 100)$.)

Real Peak	Method	# of estimates 1	2	3	4	5	6	7	8	9	10
1	HN	8419	1557	24							
	RAThn	9794	206								
	RSM	9779	221								
	RSM_{ls}	7675	2030	292	3						
2	HN	1655	7940	405							
	RAThn	318	9631	51							
	RSM	341	9600	59							
	RSM_{ls}	2490	5779	1679	52						
3	HN	16	818	9068	98						
	RAThn	3	115	9873	9						
	RSM	4	124	9856	16						
	RSM_{ls}	501	1583	6688	1221	7					
4	HN			296	9648	56					
	RAThn		1	51	9945	3					
	RSM		1	64	9932	3					
	RSM_{ls}	97	192	1528	7170	1008	5				
5	HN				88	9877	35				
	RAThn				18	9980	2				
	RSM				23	9974	3				
	RSM_{ls}	19	15	66	1321	7532	1021	21	2	1	2
6	HN					37	9862	101			
	RAThn					4	9976	20			
	RSM					6	9971	23			
	RSM_{ls}	3	1	1	27	1051	7522	1299	56	24	16
7	HN						52	9660	288		
	RAThn						2	9951	47		
	RSM						5	9935	60		
	RSM_{ls}					8	990	7234	1512	148	108
8	HN							118	9018	847	17
	RAThn							11	9851	134	4
	RSM							15	9832	148	5
	RSM_{ls}						4	1248	6528	1676	544
9	HN								388	8009	1603
	RAThn								48	9654	298
	RSM								50	9632	318
	RSM_{ls}							45	1647	5894	2414
10	HN								24	1610	8366
	RAThn									190	9810
	RSM									200	9800
	RSM_{ls}							3	305	2083	7609

Ten thousand simulations are run for each situation under $CN(0.25, \mu_i, 100)$. For the clear peak simulations, in Table 25, the HN test has an average success rate of 0.8987; the RAThn has an average success rate of 0.9847; the RSM has an average success rate of 0.9831 and the RSM_{ls} test has an average success rate of 0.6963. Thus, both the RAThn and the RSM test work better than the HN test.

Table 26: Unknown peaks estimation (flat peak $CN(0.25, \mu_i, 100)$)

Real Peak	Method	# of estimates 1	2	3	4	5	6	7	8	9	10
1	HN	6556	3292	152							
	RAThn	6785	3208	7							
	RSM	6781	3209	10							
	RSM_{ls}	5508	3451	1009	32						
2	HN	1713	7048	1239							
	RAThn	365	8789	846							
	RSM	387	8747	866							
	RSM_{ls}	2568	4351	2836	243	2					
3	HN	113	1637	7942	308						
	RAThn	17	532	9379	72						
	RSM	22	554	9342	82						
	RSM_{ls}	1029	1943	5100	1892	36					
4	HN		7	905	8956	132					
	RAThn	3	1	487	9485	24					
	RSM	3	1	514	9453	29					
	RSM_{ls}	299	462	2131	5714	1360	32	2			
5	HN				329	9570	101				
	RAThn				184	9792	24				
	RSM				207	9764	29				
	RSM_{ls}	75	52	262	1924	6110	1437	111	14	9	6
6	HN					131	9562	307			
	RAThn					33	9794	173			
	RSM					34	9777	189			
	RSM_{ls}	10	3	5	90	1418	6163	1936	254	68	53
7	HN						120	8967	903	9	1
	RAThn						16	9514	469	1	
	RSM						21	9484	492	3	
	RSM_{ls}		1		3	26	1386	5789	2057	490	248
8	HN							262	7994	1631	113
	RAThn							77	9386	522	15
	RSM							86	9346	550	1894
	RSM_{ls}					1	37	1876	5098	1894	1094
9	HN							2	1275	7023	1700
	RAThn							1	923	8688	388
	RSM								942	8656	401
	RSM_{ls}						1	251	2954	4320	2474
10	HN								151	3211	6638
	RAThn								9	3065	6926
	RSM								10	3096	6894
	RSM_{ls}							21	1006	3391	5582

In Table 26, for the flat peak situation, the HN test has an average success rate of 0.8026; the RAThn test has an average success rate of 0.8854; the RSM test has an average success rate of 0.8824 and the RSM_{ls} test has an average success rate of 0.5374.

In conclusion, under normal distribution, there is no significant difference among all four tests. Under contaminated normal distribution, the RAThn test is the best among all four tests.

CHAPTER 6

ONE-WAY LAYOUT WITH COVARIATES

To reduce large error term variances, researchers may have some covariate measurements along with the treatment measurements. The new method RAThn is based on a rank-based full model, thus it is easy to extend to an ANCOVA model. The full ANCOVA model is fit and the same linear combination of treatments is used as in chapter (3.1). Thus the test automatically adjust for the covariates. In particular, the RAThn test remains consistant.

For the model, we also assume that the slopes for the different groups are homogeneous. If this is not true then the ordering of the treatment effects may depend on the location in the covariate space. For heteroscedastic slopes, a point in covariate space may be selected and an ordering could be tested at that point. This generalizes the rank-based picked-point-analyses developed in Watcharotone and McKean et al. (2015).

An investigation of such a procedure may be done in the future, but, for now, it is beyond the scope of this thesis.

Clearly, tests of homogeneous slopes may be conducted as an initial test. We would recommend the rank-based drop in dispersion test as discussed in Schrader and McKean (1977) and, more recently, in chapter 4 of Kloke and McKean (2014).

Based on the results of this test, one may or may not proceed with an ordering procedure. Such an adaptive procedure may be investigated in the future.

6.1 The Covariance Model

Let Y_{ij} $(i = 1, \ldots, n_j,\ j = 1, \ldots, k)$ be the ith response fot the jth treatment. Let $\boldsymbol{Y}$ be a $n \times 1$ vector of Y_{ij}, where $n = \sum n_j$, let $\boldsymbol{\theta}$ be a $k \times 1$ vector of θ_j, and let $\boldsymbol{X_{cov}}$ be a $n \times p$ matrix of covariates, where p is the number of covariates. Then the covariance model is written as

$$\boldsymbol{Y} = \boldsymbol{W\theta} + \boldsymbol{X_{cov}}\gamma + \boldsymbol{e}, \tag{6.1}$$

where $\boldsymbol{W}$ is a $n \times k$ matrix in which the ith column has ones corresponding to the ith sample and zeros otherwise, $\boldsymbol{\gamma}$ is a $p \times 1$ vector of the slop parameters and $\boldsymbol{e}$ is a $n \times 1$ vector of random errors which are assumed to be *iid* from a continuous distribution.

Then,

$$E(Y_{ij}) = \theta_j + x_{ij}^{'}\gamma.$$

For two treatment groups j and $j^{'}$, if the covariates are the same, we have

$$E(Y_{ij}) - E(Y_{ij'}) = \theta_j - \theta_{j'}.$$

6.2 The AT Test

The AT procedure test statistic for covariates case is presented as

$$T = \frac{\sum_{i=1}^{k} h_i \bar{y}_i}{\sqrt{MS(\sum_{i=1}^{k} \frac{h_i^2}{n_i} + \boldsymbol{d}^{'} \boldsymbol{W_x}^{-1} \boldsymbol{d})}},$$

where h_is are the maxmin contrast coefficients, $\bar{y}_i$s are the adjusted group means, MS is the mean square residual within groups from the multiple ANCOVA, $\boldsymbol{W_x}$ is the deviation (i.e., centered) score within group sum of products matrix for the covariates, $\boldsymbol{d}$ is the weighted sample covariate means (see (Huitema, 2011), for details of this test statistic).

6.3 The RAThn Test

As in chapters 3 and 4, the RAThn test for the covariates model is also based on the rank-based model. The estimate of $\boldsymbol{\beta}_1$ is obtained from fitting the following model

$$
\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{W\theta} + \boldsymbol{X_{cov}\gamma} + \boldsymbol{e} \\
&= \begin{bmatrix} \mathbf{1} & \boldsymbol{X_1} & \boldsymbol{X_{cov}} \end{bmatrix} \begin{bmatrix} \alpha \\ \boldsymbol{\beta_1} \\ \boldsymbol{\gamma} \end{bmatrix} + \boldsymbol{e} \\
&= \boldsymbol{X\beta} + \boldsymbol{e}.
\end{aligned}
$$

The RAThn test statistic is

$$T = k\boldsymbol{a}'\hat{\boldsymbol{\beta}}_1, \tag{6.2}$$

where $\underline{\boldsymbol{a}} = (a_2,\ a_3,\ \dots\ a_k)'$ and $\sum_{i=1}^{k} a_i = 0$.

The *var-cov* matrix of $\boldsymbol{\beta}$ is

$$\boldsymbol{\Sigma}_\beta = \begin{bmatrix} \kappa_n & -\tau_\varphi^2 \bar{\boldsymbol{X}}'(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ -\tau_\varphi^2 \bar{\boldsymbol{X}}'(\boldsymbol{X}'\boldsymbol{X})^{-1} & \tau_\varphi^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \end{bmatrix}, \tag{6.3}$$

where $\kappa_n = n^{-1}\tau_s^2 + \tau_\varphi^2 \bar{\boldsymbol{X}}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\bar{\boldsymbol{X}}$ and $\tau_s = (2f(\theta_e))^{-1}$.

Let

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{0}_{(k-1)\times 1} & \mathrm{I}_{k-1} & \mathbf{0}_{(k-1)\times p} \end{bmatrix}. \tag{6.4}$$

Thus,

$$var(T) = k^2\hat{\tau}_\varphi^2 \boldsymbol{a}'\boldsymbol{H}(\boldsymbol{X}_c^{*'}\boldsymbol{X}_c^*)^{-1}\boldsymbol{H}'a, \tag{6.5}$$

where

$$X^* = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_{cov} \end{bmatrix}.$$

Based on (2.6) the asymptotically distribution of T is

$$T \dot{\sim} N(k\boldsymbol{a}'\boldsymbol{\beta}_1, k^2\tau_\varphi^2 \boldsymbol{a}'\boldsymbol{H}(\boldsymbol{X}_c^{*'}\boldsymbol{X}_c^*)^{-1}\boldsymbol{H}'a). \tag{6.6}$$

We next state the null asymptotic distribution of T:

Theorem 6.1. *If $H_0 : \theta_1 = \cdots = \theta_k$ is true, then $E(T) = 0$, and $Var(T) = k^2\hat{\tau}_\varphi^2 \boldsymbol{a}'\boldsymbol{H}(\boldsymbol{X}_c^{*'}\boldsymbol{X}_c^*)^{-1}\boldsymbol{H}'a$. Then as $n \to \infty$, where $n = \sum_{i=1}^k n_i$,*

$$T^* = \frac{k\boldsymbol{a}'\hat{\boldsymbol{\beta}}_1}{\sqrt{k^2\hat{\tau}_\varphi^2 \boldsymbol{a}'\boldsymbol{H}(\boldsymbol{X}_c^{*'}\boldsymbol{X}_c^*)^{-1}\boldsymbol{H}'a}} \xrightarrow{D} Z \sim N(0,1).$$

The asymptotic test is

Reject H_0 (1.2) in favor of H_A (1.3)

if $T^* \geq z_\alpha$; otherwise do not reject.

6.4 Simulation Studies

Ten thousand simulations were run to compare the empirical αs and the empirical power of methods mentioned in previous sections. We set $k = 5$, $p = 3$, $n_j = 15$. Three covariates were generated from a multi-variate normal distribution $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}' = (0\ 0\ 0)$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.4 & 0.3 \\ 0.4 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix}.$$

By solving the power expression

$$p_0 = P\left\{F(\delta,\ k-1,\ n-k-p) \geq F(1-\alpha,\ k-1,\ n-k-p)\right\},$$

we obtain the noncentrality parameter δ which is approximately 1.81.
Using the same method as in chapter 2, when setting $\sigma^2 = 1$ and $\mu_1 = 0$, we obtained $d = \sqrt{1.81/150}$.

Under H_0 the response observations are generated from $Normal(0,1)$, $CN(0.25, 100)$ and $SCN(0.25, , 1, 100)$ distributions, respectively. The optimal score functions, Wilcoxon scores, bent4 (0.25,0.75,-1,1) and bent1 (0.75,-2,1) are used for the corresponding distribution.

6.4.1 With *Normal* Underlying Distribution

Table 27 and Figure 23 show the simulation results of the AT, RAThnT and RAThnN tests under normally distributed errors. There are no significant differences among all three tests.

Table 27: Empirical α and power under *Normal* distribution.

Method	Empirical Power			
	0	0.60	2.41	5.43
AT	.0971	.4960	.8973	.9939
RAThnT	.0979	.4948	.8954	.9934
RAThnN	.0997	.4996	.8971	.9935

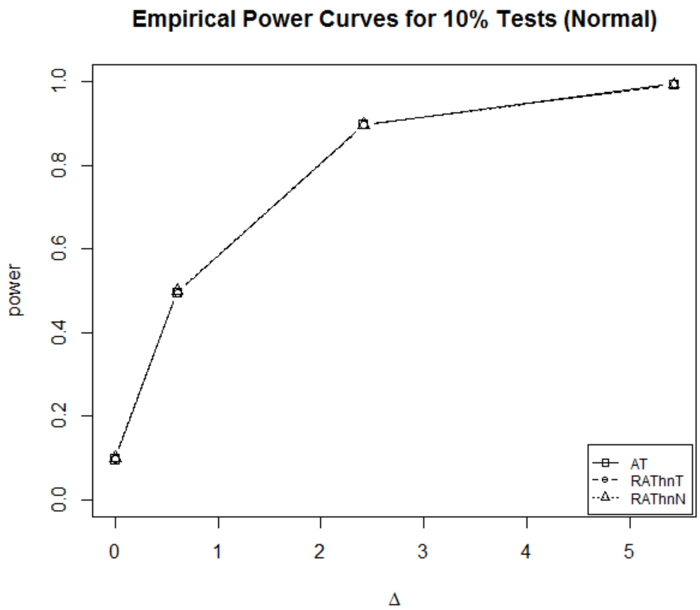

Figure 23: Plot of power curves under *Normal* distribution.

6.4.2 With $CN(0.25,\ 100)$ Underlying Distribution

Tables 28 and 29 and Figures 24 and 25 all show the simulation results of the AT, RAThnT and RAThnN tests under contaminated normally distributed errors. The scores used for the RAThn procedures are Wilcoxon scores and bent4 (.25,.75,-1,1) scores. Both RAThnT and RAThnN have higher power than the AT test.

Table 28: Empirical α and power under $CN(0.25, 100)$ distribution with Wilcoxon scores.

Method	Empirical Power					
	0	0.60	2.41	5.43	9.65	15.08
AT	.1019	.1604	.2306	.3187	.4057	.5101
RAThnT	.0854	.2730	.5405	.7905	.9254	.9762
RAThnN	.0872	.2764	.5451	.7947	.9272	.9766

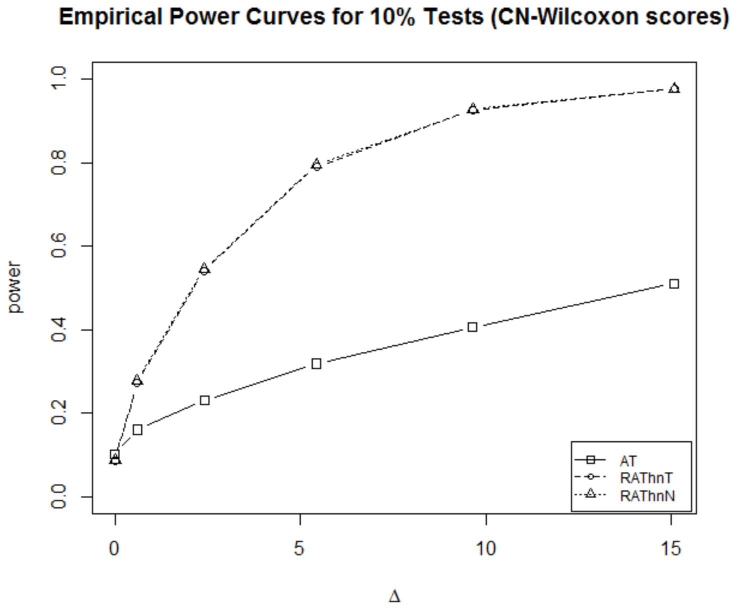

Figure 24: Plot of power curves under CN distribution (Wilcoxon scores).

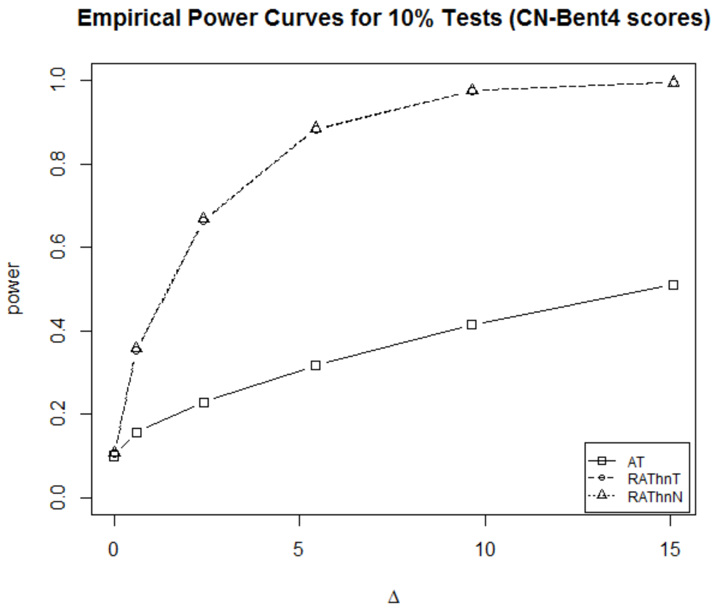

Figure 25: Plot of power curves under CN distribution (bent4 scores).

Table 29: Empirical α and power under $CN(0.25, 100)$ distribution with bent4 scores.

Method	Empirical Power					
	0	0.60	2.41	5.43	9.65	15.08
AT	.0999	.1564	.2289	.3166	.4143	.5097
RAThnT	.1041	.3519	.6635	.8828	.9753	.9945
RAThnN	.1067	.3574	.6676	.8854	.9761	.9945

6.4.3 With $SCN(0.25,\ 1,\ 100)$ Underlying Distribution

Tables 30 and 31 and Figures 26 and 27 all show the simulation results of the AT, RAThnT and RAThnN tests under skewed contaminated normal distribution. The scores used for the RAThn procedures are Wilcoxon scores and bent1 (.75,-2,1) scores. Both RAThnT and RAThnN have higher power than the AT test.

Table 30: Empirical α and power under SCN distribution with Wilcoxon scores.

Method	Empirical Power					
	0	0.60	2.41	5.43	9.65	15.08
AT	.1017	.1588	.2280	.3155	.4215	.5207
RAThnT	.0888	.2732	.5486	.7925	.9232	.9754
RAThnN	.0899	.2789	.5538	.7972	.9253	.9763

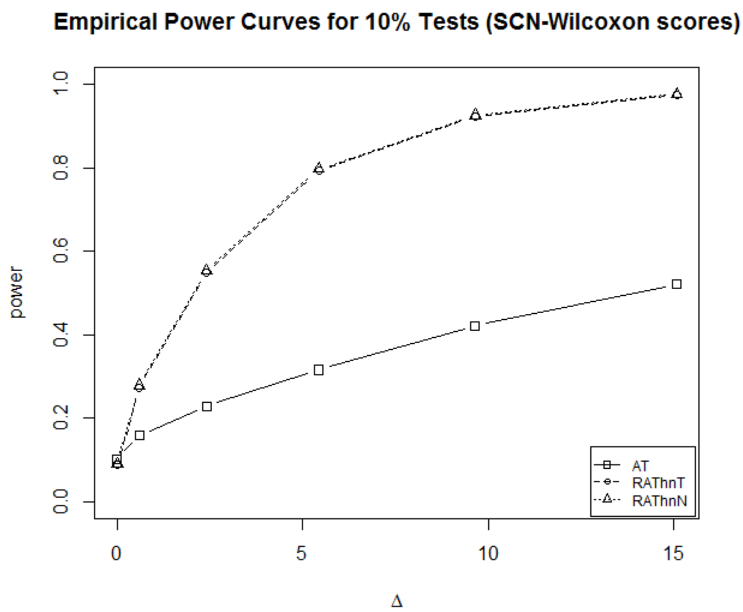

Figure 26: Plot of power curves under SCN distribution (Wilcoxon scores).

Table 31: Empirical α and power under SCN distribution with bent1 scores.

Method	Empirical Power					
	0	0.60	2.41	5.43	9.65	15.08
AT	.1030	.1396	.1907	.2499	.3084	.3862
RAThnT	.0779	.2310	.4864	.7340	.8847	.9586
RAThnN	.0799	.2359	.4908	.7383	.8875	.9594

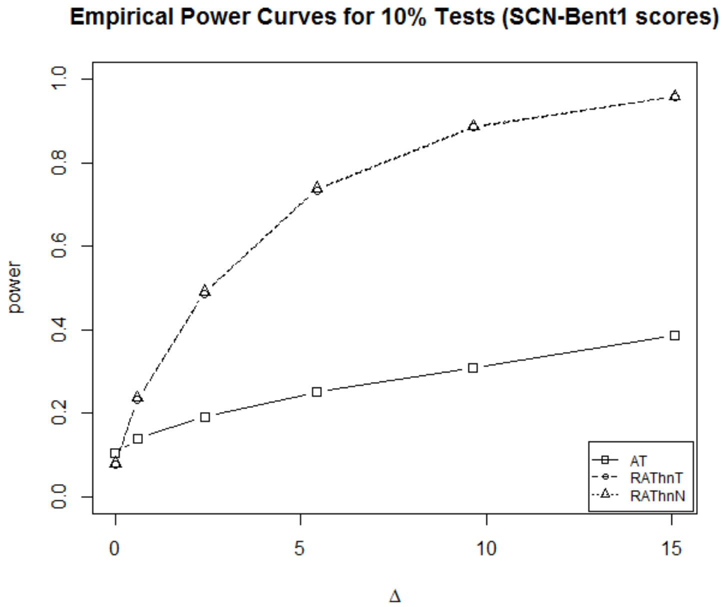

Figure 27: Plot of power curves under SCN distribution (bent1 scores).

When errors follow a normal distribution, the RAThn test works as well as the AT test. When errors follow a contaminated normal distribution or a skewed contaminated normal distribution, the RAThn test outperforms the AT test.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This chapter summarizes the thesis, and outlines the directions for future work.

7.1 Conclusions

Our investigation consisted of a large simulation study with the number of simulations 10,000, the number of groups set at 5 and with group sizes 5. Error distributions included the normal, Student t with degrees of freedom 10, 8, 5, 3, 2 and 1, and a skewed contaminated normal with mean 1, variance 100 and contaminateion rate 0.25. Besides the Abelson-Tukey (AT), for comparison purposes our study included the robust (well-known) nonparametric procedures: the Jonckheere-Terpstra (JT), the Spearman (SP), and the Hettmansperger-Norton (HN). These robust tests performed much better for the heavy-tailed error distributions. The JT test is even as efficient as AT on normal errors. The AT test, however, is easily extended to general linear and mixed models.

For a general linear model of the form

$$\boldsymbol{Y} = \boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{Y}$ is the $n \times 1$ vector of responses, $\boldsymbol{X}$ is the $n \times (k-1)$ design matrix, and $\boldsymbol{e}$ is the $n \times 1$ vector of error terms. The least squares estimator minimizes the Euclidean distance between $\boldsymbol{Y}$ and $\hat{\boldsymbol{Y}}$, the predicted value of $\boldsymbol{Y}$. A rank-based estimator is a different measure of distance which is based on the dispersion function of Jaeckel. The only assumption on the distribution of the errors required is that it is continuous.

Our initial investigation of the Abelson-Tukey (AT) procedure, showed that the AT test is not robust for error distribuiton with heavy tails. Thus, the goal of this study is to robustify the AT test. We have developed rank-based procedures for ordered alternative models. These procedures, as the AT procedure, are model based and easily extended to mixed model and ANCOVA cases. Several procedures are proposed and tested in this thesis, i.e. the RAT test, the RATηn test, the RSM test and the RSM_{ls} test. The RAT and RAThn tests are rank-based procedures, the estimates of the shifts in location are obtained from a rank-based regression model as the AT's estimates are obtained from a least squares (LS) fit. Thus, these tests are easily extended to mixed model and ANCOVA cases. The difference of these two tests are the weights. The RAT test uses the AT weights as the AT test. The RAThn test uses the HN weights as the HN test. The RSM test uses the HN weights and the Hodges-Lehmann estimates of the shifts. The RSM_{ls} test uses the HN weights and the least square estimates of the shifts.

Asymptotic variance of four test statistics were derived in the thesis. They are $\tau_\varphi^2 \sum_{j=1}^k \frac{h_j^2}{n_j}$, $k^2\tau_\varphi^2 \sum_{j=1}^k \frac{a_j^2}{n_j}$, $k^2\hat{\tau}_\varphi^2 \sum_{j=1}^k \frac{a_j^2}{n_j}$, $k^2\sigma^2 \sum_{j=1}^k \frac{a_j^2}{n_j}$ of the RAT test, the RAThn test, the RSM test and the RSM_{ls}, respectively. Asymptotic distributions of all test statistics were proposed. The asymptotic power of these tests were also proposed. All four tests are consistent. We showed that the Pitman efficacy of the RAT, RAThn, RSM and RSM_{ls} tests are $n^{-1/2}\frac{\sum h_j c_j}{\tau_\varphi(\sum h_j^2/n_j)1/2}$, $n^{-1/2}\frac{\sum a_j c_j}{\tau_\varphi(\sum a_j^2/n_j)1/2}$, $n^{-1/2}\frac{\sum a_j c_j}{\tau_\varphi(\sum a_j^2/n_j)1/2}$ and $n^{-1/2}\frac{\sum a_j c_j}{\sigma(\sum a_j^2/n_j)1/2}$, respectively. When the number of groups and group size both are 5, ARE(RAT, RAThn)= 0.9423. Thus, under this case, the RAT test is as 94.23% efficient as the RAThn test. The RAThn and the RSM tests have the same Pitman efficacy as the HN test, thus, they have the same asymptotic local power. Thus, asymptotically they are equivalent.

To investigate their small sample properties, simulation studies were obtained on general linear models, mixed models, unknown peak and covariance models, respectively. Normal distribution and some heavy tailed and/or skewed distributions, i.e. contaminated normal distribution, skewed contaminated normal distribution and GF distribution, are used as the error distributions to do the simulations. Ten thousand simulations are run for each situation.

For general linear models, simulations were run with different situations: when errors follow a normal distribution $N(0,\ 1)$, when errors follow a symmetric heavy tailed contaminated normal distribution $CN(0.25,\ 100)$, when errors follow a highly right skewed distribution, skewed contaminated normal distribution $SCN(0.25,\ 100.\ 1)$, and when errors follow a heavier tailed and positively skewed $GF(1,\ 0.1)$ distribution. We set number of groups 5 and group sizes 5. Under the normal distribution, the RAT, RAThn and RSM tests were run with Wilcoxon scores. For the remaining distributions, they were run with Wilcoxon scores and optimal scores. Under the normal distribution, the RSM$_{ls}$ test dominates all other methods. Under the $CN(0.25,\ 100)$ distribution, the RAT and RAThn tests both have higher power than the JT test when using bent4 scores. Under the $SCN(0.25,\ 100.\ 1)$ distribution, the RAThn test has higher power than all other methods except the JT test, and has no significant difference with the JT test when using bent1 scores. Under the $GF(1,\ 0.1)$ distribution, the RAT and RAThn tests have significantly higher power than the JT test with logGF and bent1 scores.

For mixed models, two new methods were generated from the RAT and RAThn tests, but using the JRFit package (Kloke and McKean, 2014) to obtain the estimates of the shifts. Thus, they were called JrRAT and JrRAThn. Two types of the design matrixes, only composed of the group effects or composed of the group effects and the block effects, were used in this chapter. They were called JrRAT, JrRAThn, JrRATb and JrRAThnb, respectively. Two types of the variance-covariance matrixes, the sandwich estimation and the compound symmetry structure, were used in the simulation. Four types of the data sets , 10 blocks with 3, 4 and 7 treatment groups and 15 blocks with 7 groups, were used in the simulation. Ten thousand simulations were run to compare the new methods with the Page test, the standard nonparametric test in this situation. The variance components σ_b^2 and σ_ϵ^2 were set to 4 and 1, respectively. The simulation results show that the JrRAThnb test outperforms the Page test in all cases.

For unknown peak problems, three new methods, the RAThn, RSM and RSM$_{ls}$ tests, were compared with the HN test. Two types of data, a set with a clear peak and a set with a flat peak, were generated both from a normal distribution, and a symmetric heavy tailed contaminated normal distribution. The number of groups and the group size were set as 10 and 16, respectively. All simulation results show that the closer the peak is to the two sides, the higher the error rate will be. And except for the flat peak, under the normal distribution case, the RAThn and RSM tests have higher average success rate than the HN test to estimate the peak.

For covariance models, ten thousand simulations under each case were run to compare the RAThn and the AT tests. We set 5 groups, 3 covariates and 15 observations in each group. Under H_0 the response observations were generated from $N(0,\ 1)$, $CN(0.25,\ 100)$ and $SCN(0.25,\ 100.\ 1)$, respectively. The simulation results show that when errors follow a normal distribution, the RAThn test works as well as the AT test; when errors follow a contaminated normal distribution or a skewed contaminated normal distribution, the RAThn test outperforms the AT test.

All simulation results show that the RAThn test works well, especially when errors follow a heavy-tailed and/or heavy-tailed skewed distributions.

7.2 Future Work

Although the research presented in this thesis have showed the effectiveness of the RAThn test, it could be further developed in the following three ways.

The RAThn test has an advantage that different optimal score functions correspond to different underlying distributions can be used. Thus, the power of the test can be optimized. The choose of the optimal score functions, however, depends on the underlying error distribution which is unknown in practice. Thus, adding an adaptive procedure will be very useful. Such as the Hogg's adaptive procedure which is discussed in the Hogg, McKean and Craig (2013). We are planning future studies of such schemes.

Another future work of this study is adding an adaptive procedure to detect the homogeneity slopes of the covariance model to decide whether or not to proceed to an ordering procedure. We would recommend the rank-based drop in dispersion test as discussed in Schrader and McKean (1977), and, more recently, in chapter 4 of Kloke and McKean (2014).

The third future work of this study is investigating a procedure for the heteroscedastic slopes. We would recommend the rank-based picked-point-analyses developed in Watcharotone and McKean etc. (2015).

Acronym Index

Abbreviation	Definition
JT	Jonckheere-Terpstra test
SP	Spearman test
AT	Abelson-Tukey test
HN	Hettmansperger-Norton test
RAT	Robust Abelson-Tukey test (with AT weights and Wilconson score)
RATb1	Robust Abelson-Tukey test (with AT weights and bent1 score)
RATb4	Robust Abelson-Tukey test (with AT weights and bent4 score)
RAThnT	Robust Abelson-Tukey test (with HN weights and t test)
RAThnN	Robust Abelson-Tukey test (with HN weights and z test)
RSMT	Robust Shao-Mckean test (with HN weights and t test)
RSMN	Robust Shao-Mckean test (with HN weights and z test)
Page	Page test
JrRAT	Jr estimates with AT weights (design matrix only contains group effects)
JrRATb	Jr estimates with AT weights (design matrix contains group effects and block effects)
ATb	Abelson-Tukey test for mixed models
JrRAThn	Jr estimates with HN weights (design matrix only contains group effects)
JrRAThnb	Jr estimates with HN weights (design matrix contains group effects and block effects)

Bibliography

Abelson, R. and Tukey, J. (1963). Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order, *The Annals of Mathematical Statistics* **34**(4): 1347–1369.

Alvo, M. (2008). Nonparametric tests of hypotheses for umbrella alternatives, *The Canadian Journal of Statistics* **36**: 143–156.

Archambault, W., Mack, G. and Wolf, D. (1977). K-sample rank tests using pair-specific scoring functions, *Canad. J. Statist.* **5**: 195–207.

Büning, H. (1999). Adaptive jonckheere-type tests for ordered alternatives, *Journal of Applies Statistics* **26**: 541–551.

Chang, W., McKean, J., Naranjo, J. and Sheather, S. (1999). High-breakdown rank regression, *Journal of the American tatistical Association* **94**: 205–219.

Chen, Y. (1991). Notes on the mack-wolfe and chen-wolfe tests for umbrella alternatives, *Biometrical Journal, Journal of Mathematical Methods in Biosciences* **33**: 281–290.

Chen, Y. and Wolfe, D. (1990). A study of distribution-free tests for umbrella alternatives, *Biometrical Journal, Journal of Mathematical Methods in Biosciences* **32**: 47–57.

Chernoff, H. and Savage, I. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics, *Annals of Mathematical Statistics* **29**: 972–994.

Hettmansperger, T. and McKean, J. (2011). *Robust Nonparametric Statistical Methods*, Taylor & Francis Group, FL:Boca Ratonl.

Hettmansperger, T. and Norton, R. (1987). Tests for patterned alternatives in k-sample problems, *Journal of the American Statistical Association* **82**(397): 292–299.

Hollander, M. and Wolfe, A. D. (1999). *Nonparametric Statistical Methods*, John Wiley & Sons, Inc., New York.

Huber, P. (1981). *Robust Statistics*, John Wiley & Sons, Inc., New York.

Huitema, B. (2011). *Analysis of Covariance and Alternatives: Statistical Methods for Experiments,Quasi-Experiments, and Single-Case Studies*, John Wiley & Sons, Inc., New York.

Jaeckel, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals, *The Annals of Mathematical Statistics* **43**(5): 1449–1458.

Jonckheere, A. (1954). A distribution-free k-sample test against ordered alternatives, *Biometrika* (41): 133–145.

Jureckova, J. (1971). Nonparametric estimate of regression coefficients, *The Annals of Mathematical Statistics* **42**(4): 1328–1338.

Kalbfleisch, J. and Prentice, R. (1980). *The statistical analysis of failure time data*, John Wiley & Sons, Inc., New York.

Kendall, M. and Stuart, A. (1961). *Advanced Theory of Statistics, 2*, Hafer Publishing Company, New York.

Kloke, J. and McKean, J. (2012). Rfit: Rank-based estimation for linear models, *The R Journal* **4**(2): 57–64.

Kloke, J. and McKean, J. (2014). *Nonparametric Statistical Methods Using R*, Boca Raton, FL: Chapman-Hall.

Kloke, J., McKean, J. and Rashid, M. (2009). Rank-based estimation and associated inferences for linear models with cluster correlated errors, *Journal of the American Statistical Association* **104**(485): 384–390.

Kössler, W. (2006). Some c-sample rank tests of homogeneity against umbrella alternatives with unknown peak, *Journal of Statistical Computation and SImulation* **76**: 57–74.

Kruskal, W. and Wallis, W. (1952). Use of ranks in one-criterion variance analysis, *Journal of American Statistical Association* **47**: 583–621.

Mack, G. and Wolfe, D. (1981). k-sample rank tests for umbrella alternatives, *Journal of the American Statistical Association* **76**: 175–181.

Mann, H. and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Statist.* **18**: 50–60.

McKean, J., Naranjo, J. and Huitema, B. (2001). A robust method for the analysis of experiments with ordered treatment levels, *Psychological Report* **89**(2): 267–273.

McKean, J., Vidmar, T. and Sievers, G. (1989). A robust two-stage multiple comparison procedure with application to a random drug screen, *Biometrics* **45**(4): 1281–1297.

Noether, G. (1955). On a theorem of pitman, *Annals of Mathematical Statistics* **26**: 64–68.

Odeh, R. (1971). On jonckheere's k-sample test against ordered alternatives, *Technometrics* **13**: 912–918.

Pan, G. (1996). Distribution-free tests for umbrella alternatives, *Communications in Statistics-Theory and Methods* **25**: 3185–3194.

Puri, M. (1965). Some distribution-free k-sample rank tests of homogeneity against ordered alternatives, *Comm. Pure Appl. Math.* **18**: 51–63.

Shi, N. (1988). Rank test statistics for umbrella alternatives, *Communications in Statistics. Theory and Methods* **17**: 2059–2073.

Spearman, C. (1904). The proof and measurement of association between two things, *American Journal of Psychology* **15**: 72–101.

Terpstra, J. and Magel, R. (2003). A new nonparametric test for the ordered alternative problem, *Nnparametric Statistics* **15**: 289–301.

Terpstra, T. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present, *Indagationes Mathematicae* (14): 327–333.

Tyron, P. and Hettmansperger, T. (1973). A class of non-parametric tests for homogeneity against ordered alternatives, *Annals of Statistics* **6**: 1061–1070.