5-2010

On Robustification of Some Procedures Used in Analysis of Covariance

Kuanwong Watcharotone
Western Michigan University

ON ROBUSTIFICATION OF SOME PROCEDURES USED IN ANALYSIS OF COVARIANCE

by

Kuanwong Watcharotone

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Advisor: Joseph W. McKean, Ph.D.

Western Michigan University
Kalamazoo, Michigan
May 2010

ON ROBUSTIFICATION OF SOME PROCEDURES USED IN ANALYSIS OF COVARIANCE

Kuanwong Watcharotone, Ph.D.

Western Michigan University, 2010

This study discusses robust procedures for the analysis of covariance (ANCOVA) models. These methods are based on rank-based (R) fitting procedures, which are quite analogous to the traditional ANCOVA methods based on least squares fits. Our initial empirical results show that the validity of R procedures is similar to the least squares procedures. In terms of power, there is a small loss in efficiency to least squares methods when the random errors have a normal distribution but the rank-based procedures are much more powerful for the heavy-tailed error distributions in our study.

Rank-based analogs are also developed for pick-a-point, adjusted mean, and the Johnson-Neyman procedures. Instead of regions of significance, pick-a-point procedures obtain the confidence interval for treatment differences at any selected covariate point. For the traditional adjusted means procedures, it is established that they can be derived from the underlying design by using the normal equations. This is then used to derive the rank-based adjusted means, showing that they have the desired asymptotic representation. This study compares these with their LS counterparts, the naive adjusted Hodges-Lehmann, and adjusted medians. A rank-based analog is developed for the Johnson-Neyman technique which obtains a region of significant differences of the treatments. Examples illustrate the rank-based procedures. For each of these ANCOVA procedures, Monte Carlo analysis is conducted to compare

empirically the differences of the traditional and robust methods. The results indicate that these robust, rank-based, procedures have more power than the traditional least squares for longer-tailed distributions for the situations investigated.

ACKNOWLEDGMENTS

I would like to wholeheartedly express my sincere gratitude to Dr. Joseph W. McKean, advisor and committee chair, for his patience, advice and guidance. I am also deeply grateful to Dr. Jung Chao Wang, committee member, for his suggestions and guidance including invaluable help in LaTex. The valuable advice, ideas, and guidance of Dr. Bradley E. Huitema, committee member, are gratefully acknowledged.

Special thanks to my parents, Dr. Sirichai and Suchin, for their unfailing encouragement and support throughout the long period of my study. I would also like to thank Therawat and Sukda, my husband and son, for their constant encouragement and support.

Profound appreciation and thanks are given to friends and well-wishers at Bronson Methodist Hospital for their unforgettable hospitality, wholesome friendliness, and moral support.

Kuanwong Watcharotone

TABLE OF CONTENTS

Table of Contents—continued

LIST OF TABLES

List of Tables—continued

List of Tables—continued

LIST OF FIGURES

List of Figures—continued

CHAPTER I

INTRODUCTION

The analysis of covariance (ANCOVA) is a technique that combines the features of analysis of variance (ANOVA) and regression (Snedecor and Cochran, 1980, p.365). The analysis of covariance is used to assess whether the means of two or more population groups are equal; this is similar to analysis of variance. However, the analysis of covariance has an advantage in that it reduces bias and increases power. In experimental studies involving random assignment of units to conditions, the covariate, when related to the response variable, reduces the error variance, resulting in increased statistical power and greater precision in the estimation of group effects (Keselman et al., 1998). An adjustment of the treatment effect is included as a standard part of analysis of covariance to reduce bias. Snedecor and Cochran (1980, p.380) show that the covariance adjustment removes all the bias if we have random samples and the regression of Y on X is linear with the same slope for each group. The homogeneity of within group slopes is assumed under the analysis of covariance model. If this assumption is not met, an alternative technique is required. There are several alternative techniques available; however, we are interested in the pick-a-point technique (Huitema, 1980; Rogosa, 1980) and the Johnson-Neyman technique (Johnson and Neyman, 1936).

The techniques we mentioned above rely on traditional least square type methods. These methods are optimal if the underlying errors have a normal distribution but they generally become less efficient tools under violations of normality. For example, one outlier can spoil the least squares fit, its associated inference, and even its diagnostic procedures (i.e., methods which should detect the outliers), (Kloke and McKean, 2010). Least squares

procedures are thus sensitive to outliers and we say that these procedures are not robust. The outliers that cause the longer tails have an effect on the least squares fit (McKean and Vidmar, 1994). Hettmansperger and McKean (1998, p.259) suggest using a rank-based analysis based on R estimation for the analysis of covariance in case of outliers. This analysis is easy to interpret because it involves substituting another norm for the Euclidean norm of least squares; see Hettmansperger and McKean (1998). The analysis is robust, being much less sensitive to outliers than the traditional analysis. The norm depends on a score function.

The R estimate we use in this study is Weighted Wilcoxon (WW), which can be obtained from wwest function for the R statistical software package (R Development Core Team, 2005) created by Terpstra and McKean (2005). The Wilcoxon weights correspond to $b_{ij} = 1$ for $i \neq j$ and 0 otherwise, and yield the well known rank-based Wilcoxon (WIL) estimate (Terpstra and McKean, 2005). We primarily use the weighted Wilcoxon in this study, but our procedures can be generalized to other weights and by other rank regression score functions. The efficiency of the Wilcoxon estimates for normal distributed data is 0.955, and is much higher for longer-tailed distribution (Hettmansperger and McKean, 1998, p.163).

In this study, we present the robust technique and are interested in comparing the results between the LS and R estimates. In Chapter 2, we compare the validity of ANCOVA between the traditional least squares and the R estimate. Furthermore, the pick-a-point method is established. We develop rank-based analogues of pick-a-point and compare the simulation results with the traditional procedure under different distributions of response variable, slopes, and sample sizes at different points on X. In Chapter 3 we illustrate two ways (simple and design matrix) to obtain the adjusted means. We also discuss the traditional adjusted means and develop R analogues for adjusted means, including the robust adjusted median, robust naive adjusted Hodges-Lehmann, robust adjusted median design,

and robust adjusted signed-rank. These LS and R adjusted means estimates are compared when outliers occur. The simulation is also conducted on standard normal and contaminated normal distributions. The R analog for the Johnson-Neyman technique is developed and presented in Chapter 4. The power of the Johnson-Neyman region of significance, as well as the power of the simultaneous region of significance based on the LS and R procedures are compared at different points of X and at different distributions of Y.

CHAPTER II

ANALYSIS OF COVARIANCE

Consider the situation where we are observing data from k groups of subjects. Assume that the sample size from Group i is n_i, $i = 1, 2, \ldots, k$ and denote the total sample size as $n = \sum_{i=1}^{k} n_i$. Along with the response variable Y, we observe a covariate variable x. Although much of what we do can be generalized to more than one covariate, a single covariate is convenient. Let Y_{ij} denote the jth response from the ith group and, correspondingly, let x_{ij} denote the value of the covariate. The suitable analysis of Y is based on the analysis of covariance (ANCOVA) model. Suppose the following linear model holds,

$$Y_{ij} = \mu_i + x_{ij}\beta_i + e_{ij}, \quad j = 1, 2, \ldots, n_i, \; i = 1, 2, \ldots, k, \tag{2.1}$$

where β_i is the slope parameter for the ith group, μ_i is the intercept parameter for the ith group, and the random errors e_{ij} are independent and identically distributed (iid) with probability density function (pdf) $f(t)$ and cumulative distribution function (cdf) $F(t)$. This is the general model in this paper and we generally call it the full model.

At times, matrix notation will be helpful. Denote the vector of responses by $\mathbf{Y} = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, \ldots, Y_{k1}, \ldots, Y_{kn_k})'$. Denote the corresponding vectors of covariates and errors by $\mathbf{x}$ and $\mathbf{e}$, respectively. Denote the $n \times 1$ dummy vector for the ith group by c_i, i.e., the value of c_i is one at the coordinate corresponding to Y_{ij} for $j = 1, \ldots, n_i$ with all other values zero. Let $\mathbf{d}_i = x^* c_i$, where the * denotes coordinatewise multiplication. Let $p = 2k$ and define the $n \times p$ matrix $\mathbf{X}$ to be $[c_1 \; c_2 \; \cdots \; c_k \; d_1 \; d_2 \; \cdots \; d_k]$. Denote the vector of parameters by $\mathbf{b} = (\mu_1, \mu_2, \ldots, \mu_k, \beta_1, \beta_2, \ldots, \beta_k)'$. Then we can

write Model (2.1) as

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e} \tag{2.2}$$

An equivalent model to Model (2.2) is the incremental model. Without loss of generality, we reference the first group. Let $\mathbf{X}^* = [\mathbf{1}_n \ c_2 \ \cdots \ c_k \ x \ d_2 \ \cdots \ d_k]$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones and $x = \sum_{i=1}^{k} d_i$. Note that the column spaces of the matrices $\mathbf{X}$ and $\mathbf{X}^*$ are the same but the parameter space differs. If we let $\mu_{j1} = \mu_j - \mu_1$ and $\beta_{j1} = \beta_j - \beta_1$, $j = 2, \ldots, k$ then the vector of parameters for $\mathbf{X}^*$ is $\mathbf{b}^* = (\mu_1, \mu_{21}, \ldots, \mu_{k1}, \beta_1, \beta_{21}, \ldots, \beta_{k1})'$. Then we can write Model (2.2) as

$$\mathbf{Y} = \mathbf{X}^*\mathbf{b}^* + \mathbf{e}$$

If X and Y are closely related, we may expect this model to fit the Y_{ij} values better than the analysis of variance model (Snedecor and Cochran, 1980, p.365). This implies the random errors (e) in ANCOVA are smaller than those in ANOVA; hence, the power of the ANCOVA is generally higher. The analysis of covariance has numerous uses (Snedecor and Cochran, 1980, p.365-366): (1) to increase precision in randomized experiments, (2) to adjust for sources of bias in observational studies, (3) to throw light on the nature of treatment effects in randomized experiments, and (4) to study regressions in multiple classifications.

An important issue in the application of ANCOVA is the equality of slopes of the different treatment regression lines (Neter, Kuner, Nachtshem, and Wasserman, 1996, p.1019).

$$H_0 : \beta_1 = \ldots = \beta_k$$

It must be demonstrated that the slopes are not statistically different before conducting ANCOVA (White, 2003). If $H_0 : \beta_1 = \ldots = \beta_k$ is not true then the covariate and the levels interact (Hogg, McKean, and Craig, 2005). If we accept $H_0 : \beta_1 = \ldots = \beta_k$ then

the model can be written as

$$Y_{ij} = \mu_i + x_{ij}\beta_i + e_{ij}, \quad j = 1, 2, \ldots, n_i, \; i = 1, 2, \ldots, k.$$

In this case, the hypothesis of interest is that the treatments are the same; that is,

$$H_0 : \mu_1 = \ldots = \mu_k$$

The second hypothesis of interest is to see if the covariate is needed; that is,

$$H_0 : \beta = 0$$

Linear model procedures based on the robust R fit are discussed in general in Chapter 3 and 4 of Hettmansperger and McKean (1998). This includes a discussion of robust analysis of covariance in section 4.5 of Hettmansperger and McKean (1998). In this chapter, we want to investigate the validity of these procedures to test the following hypotheses:

(H1) The slopes are the same

(H2) The treatments have no effects (assume the slopes are the same)

(H3) The treatments have no effects (no assumption that the slopes are the same)

(H4) The covariates have no effects (assume the slopes are the same)

2.1 General R-Estimates

We describe the robust estimates for a general linear model. Let Y_i denote the ith response, i = 1, 2, ..., n, and let x_i' denote a $p \times 1$ vector of the independent variables. The

general linear model can be expressed as

$$Y_i = \alpha + x_i'\beta + e_i \tag{2.3}$$

where α is the intercept parameter, β is a vector of regression coefficients, and the errors e_i are independent and identically distributed (iid) normal distribution with mean equal to 0 and variance equal to σ^2.

Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)'$ and $\mathbf{X}$ is an $n \times p$ matrix. We can write the model (2.3) as

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}\beta + \mathbf{e} \tag{2.4}$$

where $\mathbf{1}_n$ is an $n \times 1$ vector of ones, α is the intercept parameter, β is a $p \times 1$ vector of slope parameters, and the random errors are $\mathbf{e}' = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$. The LS estimate of β is estimated by minimizing the distance between $\widehat{\mathbf{Y}}_{LS}$ and $\mathbf{Y}$

$$\widehat{\boldsymbol{\beta}}_{LS} = \text{Argmin} \left\| \mathbf{Y} - \widehat{\mathbf{Y}}_{LS} \right\| = \text{Argmin} \left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \right\|,$$

where $\|.\|_{LS}$ is the Euclidean norm.

For our generator R-estimators, another norm is used. Let $\varphi(u)$ be a nondecreasing function on $(0, 1)$. Assume that $\int_0^1 \varphi(u)du = 0$ and $\int_0^1 \varphi^2(u)du = 1$. The scores generated by φ are given by $a(i) = \varphi[i/(n+1)], \quad i = 1, 2, \ldots, n$. Consider the pseudo-norm

$$\|v\|_\varphi = \sum a(R(v_i))v_i, \quad i = 1, 2, \ldots, n$$

This is shown to be pseudo-norm on R^n; see Hettmansperger and McKean (1998). Define the R-estimator as

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

The Wilcoxon scores discussed in Chapter 2 of Hettmansperger and McKean (1998) are generated by $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$. Another example is the sign scores generated by $\varphi(u) = \text{sgn}(u - \frac{1}{2})$.

At times, dispersion notation is useful. Jaeckel's dispersion function for a general score function $\varphi(u)$ is

$$D_\varphi(\beta) = \sum_{i=1}^{n} a_\varphi[R(y_i - x_i'\beta)](y_i - x_i'\beta)$$

where x_i' denotes the ith row of $\mathbf{X}$, and $R(y_i - x_i'\beta)$ is the rank of $y_i - x_i'\beta$ among $y_i - x_i'\beta$, $\ldots$, $y_n - x_n'\beta$. Recall that the errors $e_i = y_i - x_i'\beta$. Hence we arrive at a linear combination of ordered residuals. The outlying residuals are no longer squared but instead are weighted according to their rank (Hettmansperger and McKean, 1977). This would decrease the effects of outliers (Huber, 1973). Therefore, this would be recommended for the longer tailed distributions. The function $D_\varphi(\beta)$ is a continuous and convex function of β. Hence, the R estimator of β can also be written as

$$\widehat{\boldsymbol{\beta}}_\varphi = \text{Argmin} D_\varphi(\beta).$$

As we can see, instead of using the Euclidean norm, $D_\varphi(\beta)$ utilizes the pseudo-norm: $\|w\|_\varphi = \sum_{i=1}^{n} a[R(w_i)]w_i$. The intercept parameter plays an important role in adjusted means. Because the scores sum to zero and the ranks are invariant to a constant shift, the intercept cannot be estimated using the norm (Kloke and McKean, 2010). The intercept can be estimated by using the median of the residuals: $\widehat{\alpha}_\varphi = \text{med}\left\{\mathbf{Y} - \mathbf{X}'\widehat{\boldsymbol{\beta}}_\varphi\right\}$ (Hettmansperger and McKean, 1998, p.147). Recall that the LS intercept estimator is the mean of the LS residuals, hence this R intercept estimator is analogous to the LS intercept estimator. Note that this R estimate of intercept does not require symmetrically distributed

errors (Hettmansperger and McKean, 1998, p.164).

Under regularity conditions, Theorem 3.5.11 (Hettmansperger and McKean, 1998, p.166) shows

$$\begin{pmatrix} \widehat{\boldsymbol{\alpha}}_\varphi \\ \widehat{\boldsymbol{\beta}}_\varphi \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{bmatrix} n^{-1}\tau_S^2 & \mathbf{0}' \\ \mathbf{0} & \tau_\varphi^2(\mathbf{X}'\mathbf{X})^{-1} \end{bmatrix} \right),$$

where τ_S and τ_φ are the scale parameters which are defined as:

$$\begin{aligned} \tau_S &= (2f(0))^{-1} \\ \tau_\varphi &= (\sqrt{12}\int f^2(t)dt)^{-1}. \end{aligned}$$

Note that the asymptotic relative efficiency of the R estimate in relation to LS is the ratio of σ^2/τ^2. For Wilcoxon scores, this is the familiar value $12\sigma^2(\int f^2)^2$, which for normal errors is 0.955 (McKean and Vidmar, 1994). On the other hand, if the true distribution has tails heavier than the normal, then this efficiency is usually much larger than 1 (McKean, 2004). For Wilcoxon scores,

$$D_W(\beta) = K\sum\sum|(y_i - y_j) - (x_i - x_j)'\beta|,$$

where K is constant. Hence, the Wilcoxon estimate can also be written as

$$\widehat{\beta}_W = \text{Argmin}D_W(\beta).$$

2.2 Simulation Study

We created a full and reduced design matrix function using R code for each hypothesis:

(H1) The slopes are the same:

$$\mathbf{X}_{full} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & x_{n_1} & 0_{n_1} \\ 1_{n_2} & 1_{n_2} & x_{n_2} & x_{n_2} \end{bmatrix},$$

$$\mathbf{X}_{reduced} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & x_{n_1} \\ 1_{n_2} & 1_{n_2} & x_{n_2} \end{bmatrix}$$

(H2) The treatments have no effects (assume the slopes are the same):

$$\mathbf{X}_{full} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & x_{n_1} \\ 1_{n_2} & 1_{n_2} & x_{n_2} \end{bmatrix},$$

$$\mathbf{X}_{reduced} = \begin{bmatrix} 1_{n_1} & x_{n_1} \\ 1_{n_2} & x_{n_2} \end{bmatrix}$$

(H3) The treatments have no effects (no assumption that the slopes are the same):

$$\mathbf{X}_{full} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & x_{n_1} & 0_{n_1} \\ 1_{n_2} & 1_{n_2} & x_{n_2} & x_{n_2} \end{bmatrix},$$

$$\mathbf{X}_{reduced} = \begin{bmatrix} 1_{n_1} & x_{n_1} & 0_{n_1} \\ 1_{n_2} & x_{n_2} & x_{n_2} \end{bmatrix}$$

(H4) The covariates have no effects (assume the slopes are the same):

$$\mathbf{X}_{full} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & x_{n_1} \\ 1_{n_2} & 1_{n_2} & x_{n_2} \end{bmatrix},$$

$$\mathbf{X}_{reduced} = \begin{bmatrix} 1_{n_1} & 0_{n_1} \\ 1_{n_2} & 1_{n_2} \end{bmatrix}$$

The dependent variable, covariate, and vector indicators of levels/groups are inputted. The full and reduced model residuals are obtained for each hypothesis from fitting the full and reduced models. A simulation is conducted for each assumption in the case of one covariate and two groups from a standard normal distribution with 30 observations at the 5% level of significance. We run 10000 simulations for each scenario. The validity of each scenario is shown in the Table 1.

Table 1: Validity Results

Test	Procedure	Validity		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
H1	LS	0.098	0.0496	0.0082
	WWIL	0.0999	0.0487	0.009
H2	LS	0.1046	0.0505	0.0079
	WWIL	0.0957	0.0457	0.0074
H3	LS	0.1046	0.0529	0.0101
	WWIL	0.1026	0.0516	0.0097
H4	LS	0.0994	0.0507	0.0107
	WWIL	0.0954	0.0489	0.009

The results show that the validity of WWIL estimate is slightly lower than that of LS estimate except the case of H1 at $\alpha = 0.10$ and at $\alpha = 0.01$. However, the validity of both estimates is very close to the level of significance.

The weighted Wilcoxon (WW) routine is written by Terpstra and McKean (2005) via the R statistical software package (R Development Core Team, 2005). Our computation is performed by using this R collection. We denote it by WWIL in this study.

2.3 Pick-A-Point

One of the important assumptions in the use of analysis of covariance is that the regression slopes are homogeneous. Heterogeneous regression slopes associated with analysis of covariance present interpretation problems because the magnitude of the treatment effect is not the same at different levels of X (Huitema, 1980, p.270). Heterogeneous slopes are shown in Figure 1. Figure 1a shows that the adjusted means are different as most of the data in group 1 are higher than that in group 2. This would mislead the conclusion because it seems to have no treatment effects at the lower levels of X, and the higher X, the larger effects. Figure 1b shows that both groups differ in slopes and in adjusted means. Group 1 is higher than group 2 at all levels of X. Figure 1c shows both groups have the different slopes where group 1 is inferior to group 2 at the lower values of X, then appears to overlap in the middle of X, and is superior to group 2 at the higher values of X.

We focus on obtaining the confidence interval at any covariate point we choose for the treatment effects. Consider the case of two groups and one covariate. Assume that the sample size of group i is n_i, $i = 1, 2$. Let Y_{ij} denote the jth response from the ith group and let X_{ij} be the covariate. Assume that the response variable Y is normally and independently distributed, the conditional distribution of Y given X is

$$E(Y_{ij}|X_{ij}) = \alpha_i + \beta_i X_{ij},$$

where α_i is the intercept and β_i is the slope parameters for the ith group. The difference at

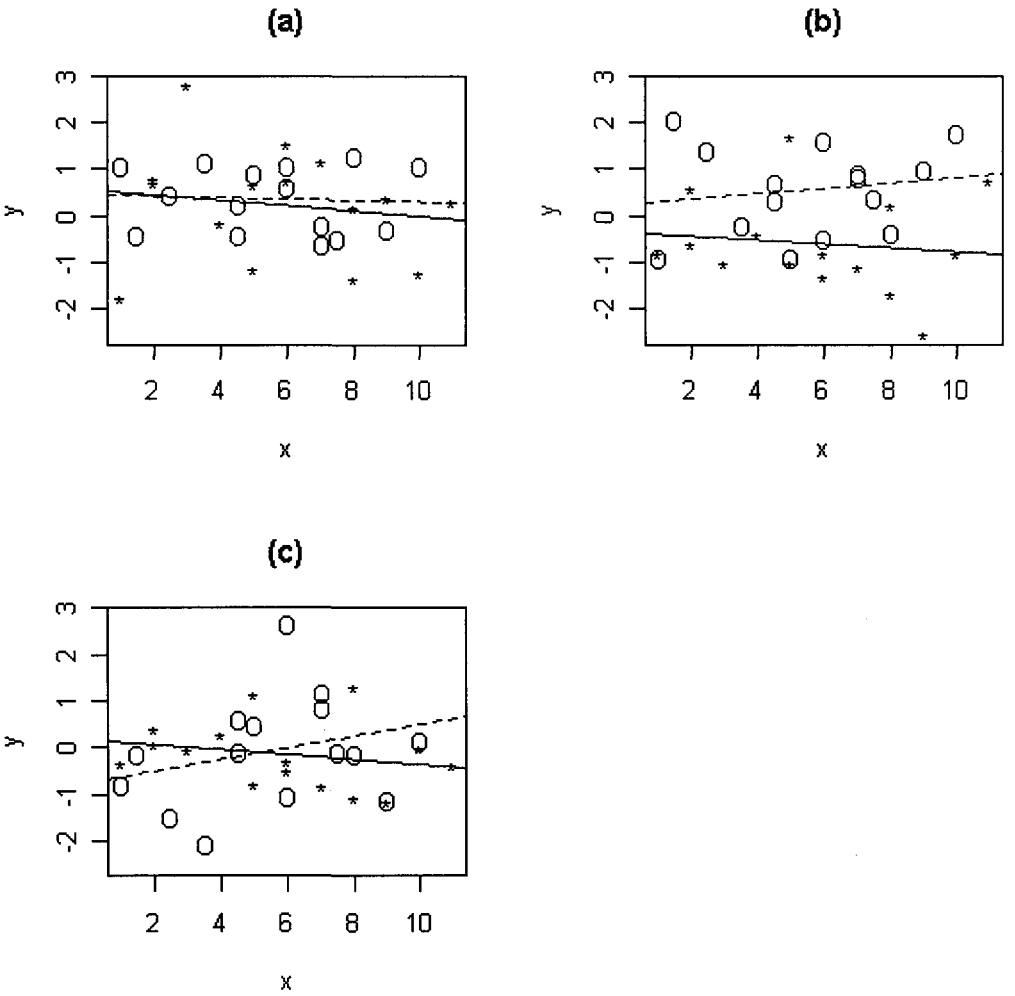

Figure 1: Three types of heterogeneous slopes

point X between both groups is

$$\begin{aligned}\Delta(X) &= E(Y_{2j}|X) - E(Y_{1j}|X) \\ &= (\alpha_2 + \beta_2 X) - (\alpha_1 + \beta_1 X) \\ &= (\alpha_2 - \alpha_1) + (\beta_2 - \beta_1)X.\end{aligned}$$

Thus, the estimator of the difference is

$$\widehat{\Delta}(X) = (\widehat{\alpha_2} - \widehat{\alpha_1}) + (\widehat{\beta}_2 - \widehat{\beta}_1)X.$$

2.3.1 The Traditional Pick-A-Point Procedure

A $100(1-\alpha)\%$ confidence interval for $\Delta(X)$ for any specified individual point X is given by

$$\widehat{\Delta}(X) \pm t_{f,1-(1/2)\alpha}[SE^2(\widehat{\Delta}(X))]^{1/2}, \tag{2.5}$$

where

$$\begin{aligned} X_{r\times 1} &= (X_1, X_2, \ldots, X_k)' \\ SE(\widehat{\Delta}(X)) &= \hat{\sigma}\sqrt{(h'\mathbf{X}'\mathbf{X})^{-1}h} \\ h &= (-1, 1, -1, 1) \\ f &= n_1 + n_2 - 4. \end{aligned}$$

2.3.2 The Rank-Based Pick-A-Point Procedure

For R estimate, a $100(1-\alpha)\%$ confidence interval for $\Delta(X)$ for any specified individual point X is the same as formulation above (2.5) except the standard error for R estimate equals to

$$SE_{\varphi}(\widehat{\Delta}(X)) = \hat{\tau}\sqrt{h'(\mathbf{X}'\mathbf{X})^{-1}h},$$

where $\hat{\tau}$ is the estimate of τ (Koul, Sievers, and McKean, 1987). This estimate is consistent under both symmetrical and asymmetrical errors (Hettmansperger, McKean, and Sheather, 2003).

2.3.3 Simulation Study

Consider the case of two groups and one covariate, the matrix X can be written as

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 \end{matrix} \\ \begin{bmatrix} 1_{n_1} & 1_{n_1} & x_{n_1} & x_{n_1} \\ 1_{n_2} & 0_{n_2} & x_{n_2} & 0_{n_2} \end{bmatrix} \end{array} \tag{2.6}$$

Suppose the point we choose is $x*$, therefore

$$\begin{aligned} E(Y_{1j}|X = x*) &= \alpha_0 + \alpha_1 + \alpha_2 x* + \alpha_3 x* \\ E(Y_{2j}|X = x*) &= \alpha_0 + \alpha_2 x* . \end{aligned}$$

Let x_0 be the point that the regression lines of both groups cross, we then have

$$\begin{aligned} E(Y_{2j}|X = x_0) - E(Y_{1j}|X = x_0) &= 0 \\ \alpha_1 + \alpha_3 x_0 &= 0 \\ \alpha_1 &= -\alpha_3 x_0. \end{aligned}$$

A simulation is conducted to obtain the power of the 95% confidence interval at 5 different points of X, which are:

(1) minimum value of X (q_0)

(2) $1st$ quartile (q_1)

(3) median of X (q_2)

(4) $3rd$ quartile (q_3)

(5) maximum value of X (q_4)

The confidence intervals are calculated based on the least square and R estimate approaches. The powers of the treatment effects at 10%, 5%, and 1% level of significance are also obtained. The covariate is generated from normal distribution with $\mu = 100$ and $\sigma = 20$. The response variable is generated from:

(1) Normal distribution

(2) Laplace distribution

(3) Cauchy distribution

The sample sizes are 20, 40, and 80. The α_3 ranges from 0 to 1.9. We run 10000 simulations for each scenario.

The power of the treatment effects test based on R estimate is slightly lower than the least squares procedure for all three levels of significance and all different α_3. The power at $\alpha_3 = 0$ at 10% level of significance almost equals to its level of significance; e.g., 10% (LS = 9.94% and WWIL = 9.74%), likewise, at 5% and 1% level of significance. Moreover, the powers of both procedures are higher when α_3 is higher (Table 2).

For the power of the 95% confidence interval at pick-a-point, the power of LS procedure at $\alpha_3 = 0$ is close to 5% at all pick-a-points, while the power of WWIL procedure is slightly lower than 5%. The power based on R estimate is lower than the least square procedure at all points. Both procedures have higher power when α_3 is higher except at the median of X (q_2). The powers at q_2 are all close to 5% for all values of α_3 (Table 3). Figure 2 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R procedures.

Table 2: H_0: $\beta = 0$ when Y is normal, $n = 20$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.0994	0.0506	0.0108
	WWIL	0.0974	0.0462	0.0089
$\alpha_3 = 0.3$	LS	0.1415	0.0768	0.0191
	WWIL	0.1395	0.0758	0.0158
$\alpha_3 = 0.6$	LS	0.2427	0.1535	0.0457
	WWIL	0.2274	0.1395	0.0358
$\alpha_3 = 0.9$	LS	0.4308	0.3032	0.1161
	WWIL	0.4010	0.2683	0.0851
$\alpha_3 = 1.2$	LS	0.7026	0.5644	0.2871
	WWIL	0.6637	0.5142	0.2269
$\alpha_3 = 1.9$	LS	0.9483	0.8949	0.6967
	WWIL	0.9303	0.8594	0.5927

Table 3: Pick-a-point when Y is normal, $n = 20$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0472	0.0505	0.0459	0.0489	0.0485
	WWIL	0.0325	0.0352	0.0320	0.0365	0.0360
$\alpha_3 = 0.3$	LS	0.0697	0.0766	0.0513	0.0560	0.0795
	WWIL	0.0460	0.0510	0.0372	0.0389	0.0519
$\alpha_3 = 0.6$	LS	0.1147	0.1538	0.0483	0.1046	0.1491
	WWIL	0.0790	0.1070	0.0340	0.0738	0.1043
$\alpha_3 = 0.9$	LS	0.2798	0.3030	0.0523	0.1365	0.2697
	WWIL	0.1933	0.2092	0.0359	0.0893	0.1850
$\alpha_3 = 1.2$	LS	0.4438	0.5636	0.0493	0.4296	0.6037
	WWIL	0.3256	0.4225	0.0353	0.3112	0.4599
$\alpha_3 = 1.9$	LS	0.8342	0.8947	0.0508	0.5140	0.8772
	WWIL	0.7005	0.7792	0.0352	0.3723	0.7585

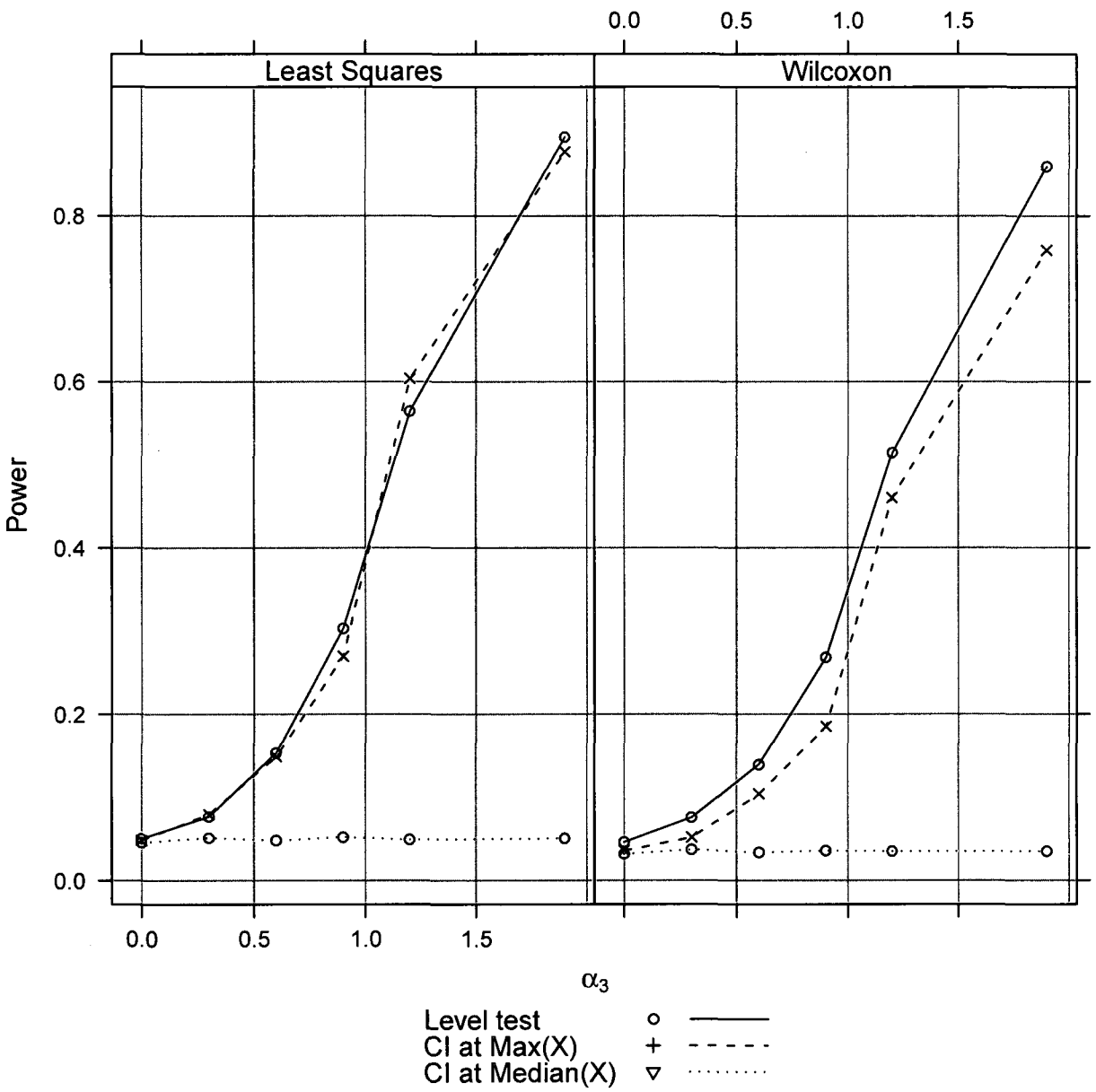

Figure 2: Y is normal, $n = 20$

The power of the treatment effects test based on R estimate is slightly lower than the least squares procedure for all three levels of significance and all different α_3. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 10.27% and WWIL = 9.81%, at 5% level of significance: LS = 4.98% and WWIL = 4.97%, and at 1% level of significance: LS = 0.92% and WWIL = 0.87%. The powers of both procedures are also higher when α_3 is higher (Table 4).

For the power of the 95% confidence interval at pick-a-point, the power of LS procedure at $\alpha_3 = 0$ is close to 5% at the minimum of X (q_0) and at $1st$ quartile (q_1), and slightly higher than 5% at the median of X (q_2), at $3rd$ quartile (q_3), and at the maximum of X (q_4). While the power of WWIL procedure, for all points, is slightly lower than 5%. At all points, the power based on R estimate is lower than the least square procedure. Both procedures have higher power when α_3 is higher except at the median of X (q_2). The powers at q_2 are all close to 5% at all values of α_3 (Table 5). Figure 3 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R procedures at $n = 40$.

Table 4: H_0: $\beta = 0$ when Y is normal, $n = 40$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.1027	0.0498	0.0092
	WWIL	0.0981	0.0497	0.0087
$\alpha_3 = 0.3$	LS	0.1857	0.1074	0.0275
	WWIL	0.1817	0.1053	0.0258
$\alpha_3 = 0.6$	LS	0.4775	0.3568	0.1519
	WWIL	0.4603	0.3382	0.1343
$\alpha_3 = 0.9$	LS	0.8530	0.7695	0.5343
	WWIL	0.8357	0.7422	0.4855
$\alpha_3 = 1.2$	LS	0.9880	0.9689	0.8817
	WWIL	0.9812	0.9588	0.8432

Table 5: Pick-a-point when Y is normal, $n = 40$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0496	0.0497	0.0539	0.0532	0.0514
	WWIL	0.0414	0.0416	0.0457	0.0477	0.0444
$\alpha_3 = 0.3$	LS	0.1008	0.1074	0.0565	0.0727	0.1033
	WWIL	0.0844	0.0877	0.0472	0.0584	0.0849
$\alpha_3 = 0.6$	LS	0.2708	0.3562	0.0493	0.2920	0.3920
	WWIL	0.2299	0.3023	0.0416	0.2399	0.3313
$\alpha_3 = 0.9$	LS	0.7210	0.7690	0.0525	0.4258	0.7631
	WWIL	0.6492	0.7029	0.0446	0.3610	0.6973
$\alpha_3 = 1.2$	LS	0.9544	0.9688	0.0537	0.5016	0.9344
	WWIL	0.9234	0.9484	0.0465	0.4313	0.8925

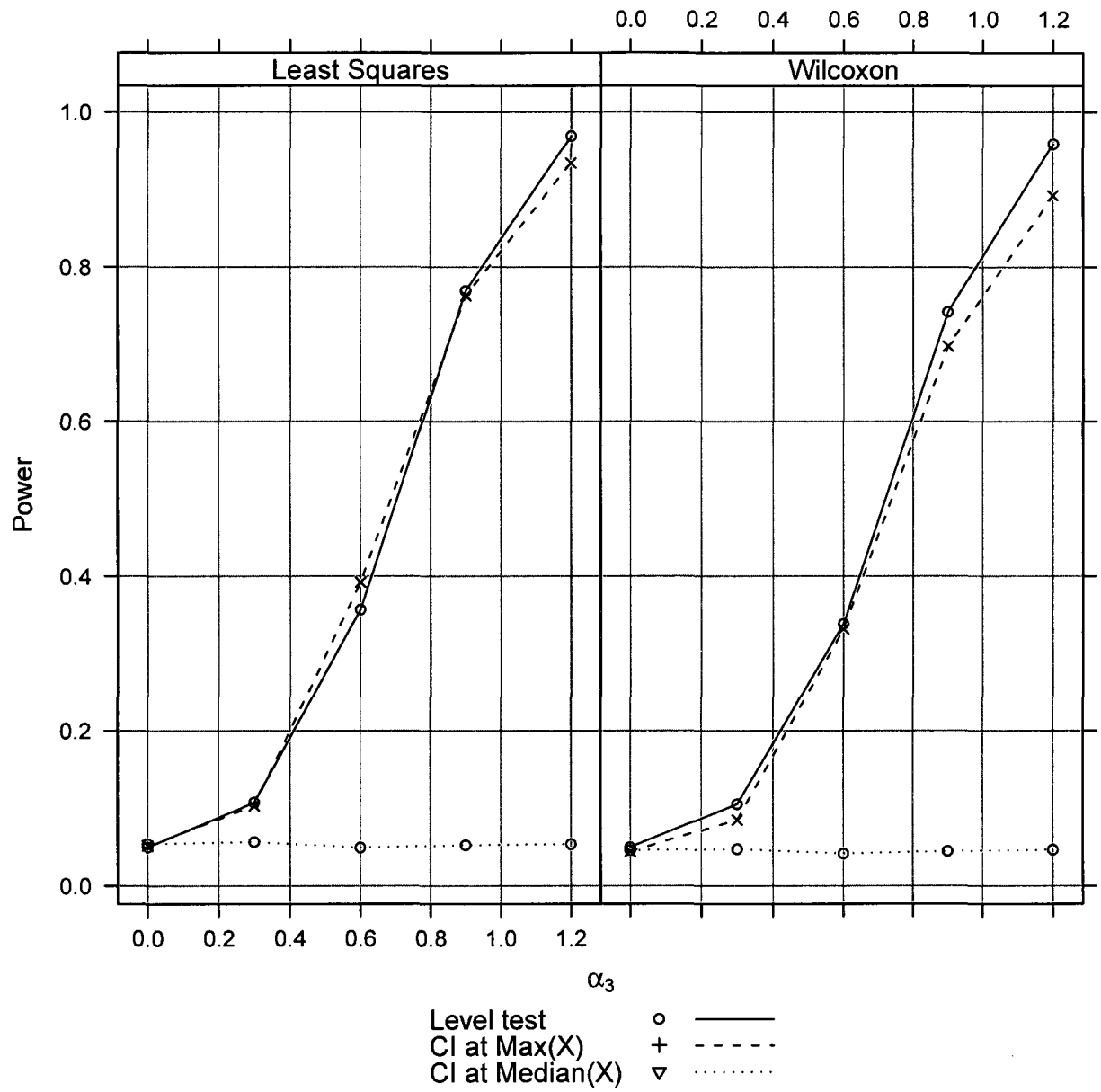

Figure 3: Y is normal, $n = 40$

Table 6: H_0: $\beta = 0$ when Y is normal, $n = 80$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.1010	0.0492	0.0101
	WWIL	0.0994	0.0503	0.0096
$\alpha_3 = 0.3$	LS	0.2923	0.1966	0.0692
	WWIL	0.2849	0.1893	0.0667
$\alpha_3 = 0.45$	LS	0.6629	0.5403	0.2940
	WWIL	0.6414	0.5136	0.2730
$\alpha_3 = 0.6$	LS	0.8499	0.7709	0.5404
	WWIL	0.8361	0.7464	0.5015
$\alpha_3 = 0.9$	LS	0.9921	0.9795	0.9257
	WWIL	0.9885	0.9747	0.9065
$\alpha_3 = 1.2$	LS	0.9998	0.9997	0.9977
	WWIL	0.9998	0.9994	0.9961

Table 7: Pick-a-point when Y is normal, $n = 80$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0484	0.0492	0.0518	0.0508	0.0506
	WWIL	0.0471	0.0463	0.0468	0.0491	0.0470
$\alpha_3 = 0.3$	LS	0.1702	0.1964	0.0481	0.1055	0.1958
	WWIL	0.1529	0.1776	0.0434	0.0962	0.1775
$\alpha_3 = 0.45$	LS	0.4729	0.5397	0.0475	0.2235	0.5453
	WWIL	0.4312	0.4932	0.0460	0.1983	0.5037
$\alpha_3 = 0.6$	LS	0.6947	0.7703	0.0508	0.4283	0.7542
	WWIL	0.6480	0.7268	0.0465	0.3920	0.7103
$\alpha_3 = 0.9$	LS	0.9315	0.9793	0.0510	0.8357	0.9931
	WWIL	0.9091	0.9710	0.0469	0.7990	0.9873
$\alpha_3 = 1.2$	LS	0.9986	0.9997	0.0512	0.9519	0.9998
	WWIL	0.9976	0.9994	0.0448	0.9359	0.9991

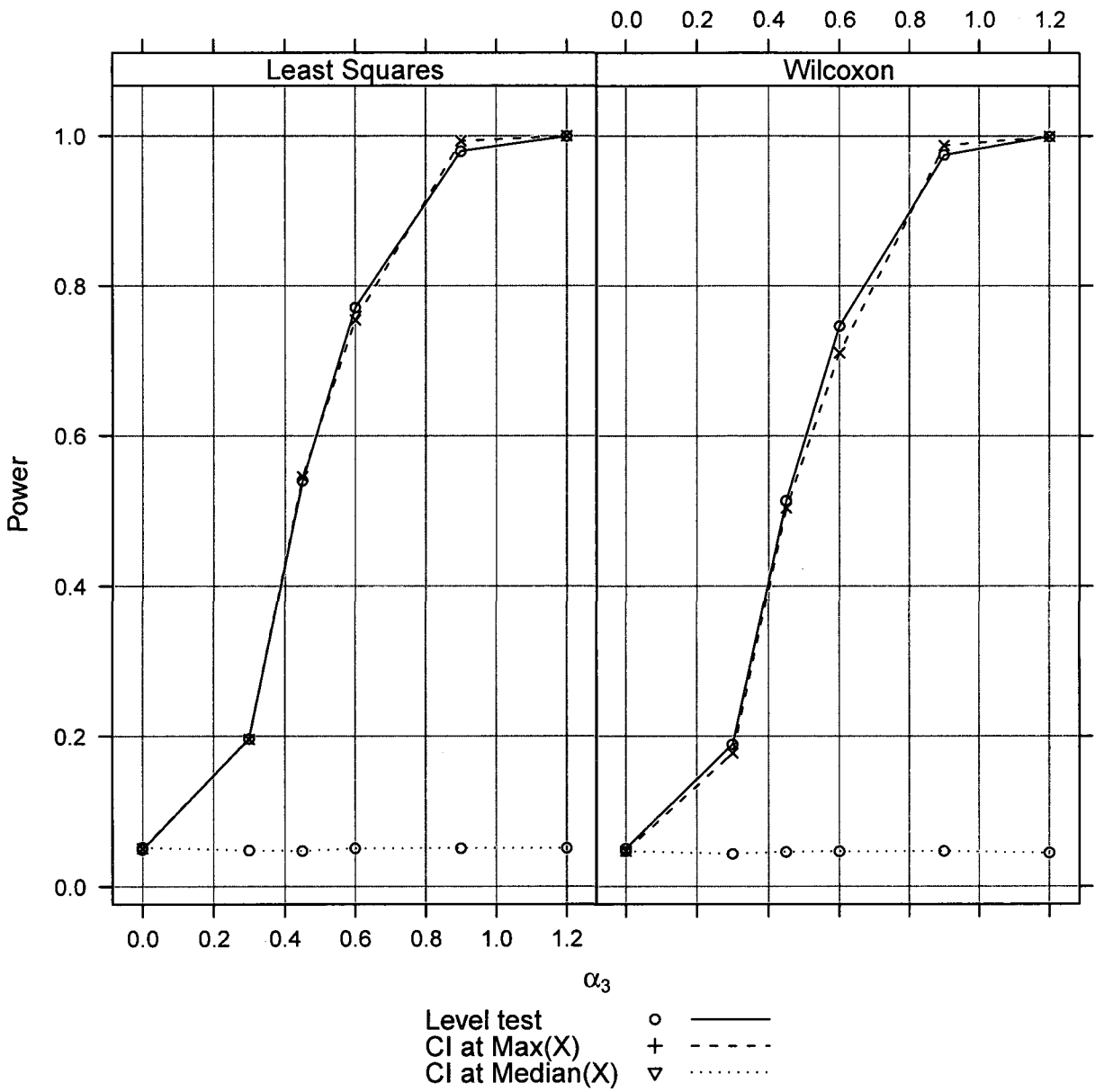

Figure 4: Y is normal, $n = 80$

The power of the treatment effects test based on R estimate is slightly lower than the least squares procedure except at 5% level of significance and $\alpha_3 = 0$, and at 10% level of significance and $\alpha_3 = 1.2$. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 10.10% and WWIL = 9.94%, at 5% level of significance: LS = 4.92% and WWIL = 5.03%, and at 1% level of significance: LS = 1.01% and WWIL = 0.96%. The powers of both procedures are also higher when α_3 is higher (Table 6).

For the power of the 95% confidence interval at pick-a-point, both powers of LS and R procedures at $\alpha_3 = 0$ is close to 5% for all points. At all points, the power based on R estimate is lower than the least square procedure. Both procedures have higher power when α_3 is higher except at the median of X (q_2). The powers at q_2 are all close to 5% at all values of α_3 (Table 7). Figure 4 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R procedures at $n = 80$.

When comparing all different sample sizes, the powers of the treatment test of both procedures reach high power at the lower value of α_3 when the sample size is larger. For instance, at $n = 20$ and 10% level of significance, the power is greater than 90% when $\alpha_3 = 1.9$, while the power of $n = 40$ is greater than 90% when $\alpha_3 = 1.2$, and when $\alpha_3 = 0.9$ when $n = 80$ (Table 2, 4, and 6). For the power of the 95% confidence interval at pick-a-point, the powers of both procedures are higher when the sample size is larger. For instance, at $\alpha_3 = 1.2$, the powers of maximum value of X (q_4) from both procedures are higher as the sample size is higher (Table 3, 5, and 7).

Table 8: H_0: $\beta = 0$ when Y is Laplace, $n = 20$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.1048	0.0563	0.0129
	WWIL	0.1118	0.0573	0.0097
$\alpha_3 = 0.3$	LS	0.1647	0.0908	0.0214
	WWIL	0.1677	0.0930	0.0204
$\alpha_3 = 0.6$	LS	0.2389	0.1516	0.0478
	WWIL	0.2606	0.1609	0.0485
$\alpha_3 = 0.9$	LS	0.5291	0.3963	0.1799
	WWIL	0.5805	0.4438	0.1921
$\alpha_3 = 1.2$	LS	0.7024	0.5773	0.3227
	WWIL	0.7491	0.6256	0.3282
$\alpha_3 = 1.9$	LS	0.9310	0.8770	0.6919
	WWIL	0.9495	0.8988	0.7085

Table 9: Pick-a-point when Y is Laplace, $n = 20$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0562	0.0560	0.0472	0.0465	0.0504
	WWIL	0.0452	0.0438	0.0279	0.0323	0.0338
$\alpha_3 = 0.3$	LS	0.0786	0.0911	0.0461	0.0597	0.0816
	WWIL	0.0570	0.0673	0.0301	0.0441	0.0600
$\alpha_3 = 0.6$	LS	0.1403	0.1519	0.0452	0.0846	0.1235
	WWIL	0.1152	0.1217	0.0311	0.0594	0.0940
$\alpha_3 = 0.9$	LS	0.2605	0.3960	0.0456	0.1750	0.3877
	WWIL	0.2237	0.3616	0.0305	0.1443	0.3541
$\alpha_3 = 1.2$	LS	0.5280	0.5772	0.0504	0.2640	0.4750
	WWIL	0.4855	0.5388	0.0298	0.2195	0.4366
$\alpha_3 = 1.9$	LS	0.7886	0.8770	0.0474	0.5658	0.8153
	WWIL	0.7616	0.8548	0.0292	0.5217	0.7976

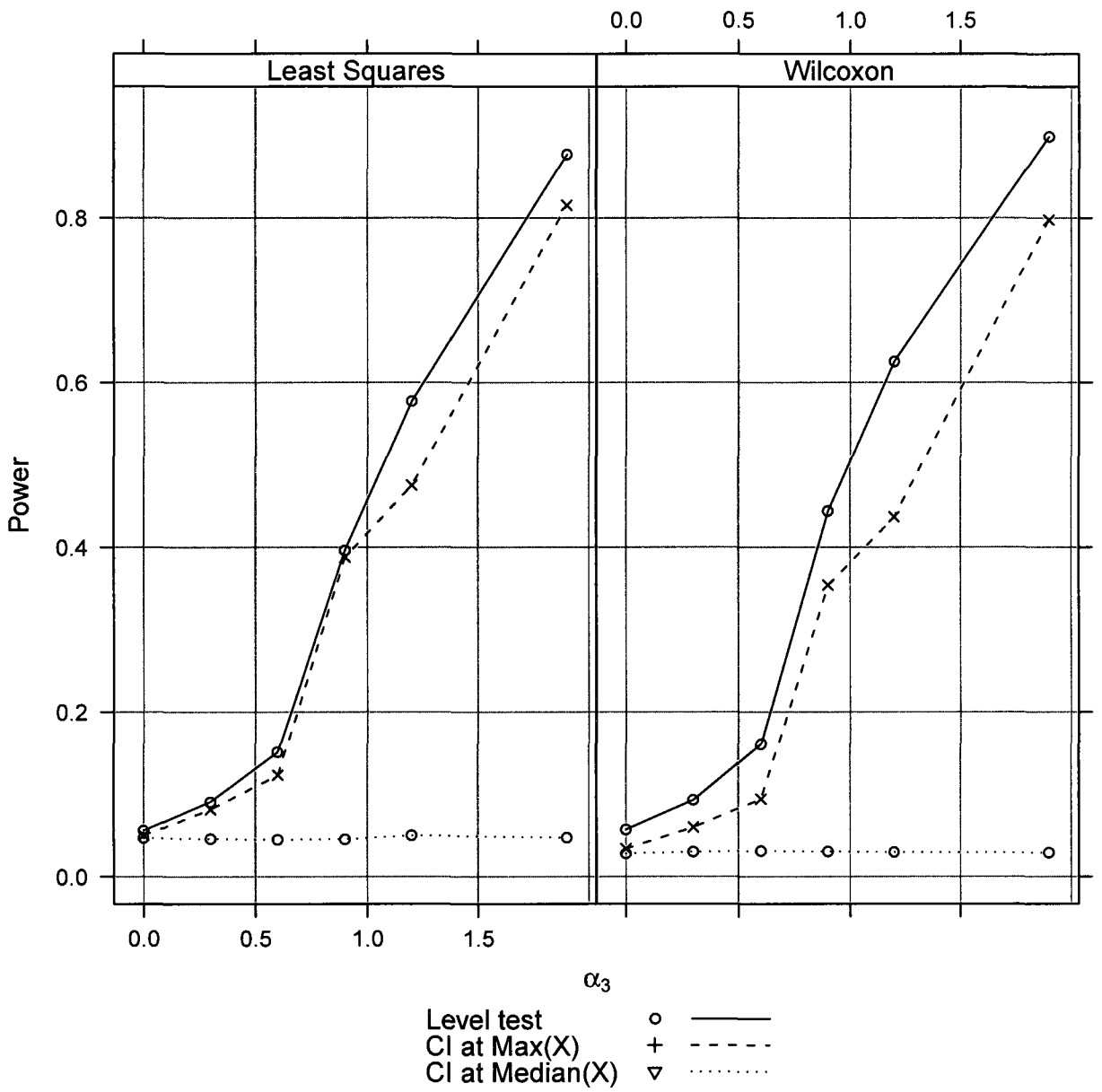

Figure 5: Y is Laplace, $n = 20$

The power of the treatment effects test based on R estimate is slightly higher than the least squares procedure except at 1% level of significance when α_3 equals to 0 and equals to 0.3. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 10.48% and WWIL = 11.18%, at 5% level of significance: LS = 5.63% and WWIL = 5.73%, and at 1% level of significance: LS = 1.29% and WWIL = 0.97%. Moreover, the powers of both procedures are higher when α_3 is higher (Table 8).

For the power of the 95% confidence interval at pick-a-point, the powers of both LS and R procedures at $\alpha_3 = 0$ is close to 5% at all pick-a-points. The power based on R estimate is lower than the least square procedure at all points. Both procedures have higher power when α_3 is higher except at the median of X (q_2). The powers at q_2 are all close to 5% for all values of α_3 (Table 9). Figure 5 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R procedures.

Table 10: H_0: $\beta = 0$ when Y is Laplace, $n = 40$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.1003	0.0491	0.0095
	WWIL	0.1028	0.0524	0.0104
$\alpha_3 = 0.3$	LS	0.2714	0.1781	0.0592
	WWIL	0.3323	0.2188	0.0722
$\alpha_3 = 0.6$	LS	0.6531	0.5201	0.2823
	WWIL	0.7393	0.6170	0.3632
$\alpha_3 = 0.9$	LS	0.6627	0.5416	0.3016
	WWIL	0.7504	0.6407	0.3862
$\alpha_3 = 1.2$	LS	0.9806	0.9604	0.8750
	WWIL	0.9913	0.9815	0.9271

Table 11: Pick-a-point when Y is Laplace, $n = 40$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0501	0.0492	0.0505	0.0430	0.0460
	WWIL	0.0441	0.0454	0.0361	0.0392	0.0429
$\alpha_3 = 0.3$	LS	0.1550	0.1779	0.0459	0.0943	0.1715
	WWIL	0.1661	0.1923	0.0371	0.0871	0.1884
$\alpha_3 = 0.6$	LS	0.4263	0.5193	0.0473	0.1770	0.4687
	WWIL	0.4779	0.5852	0.0371	0.1861	0.5316
$\alpha_3 = 0.9$	LS	0.4153	0.5412	0.0459	0.3151	0.5267
	WWIL	0.4691	0.6123	0.0374	0.3524	0.5969
$\alpha_3 = 1.2$	LS	0.9559	0.9606	0.0478	0.5228	0.9090
	WWIL	0.9768	0.9803	0.0380	0.5904	0.9456

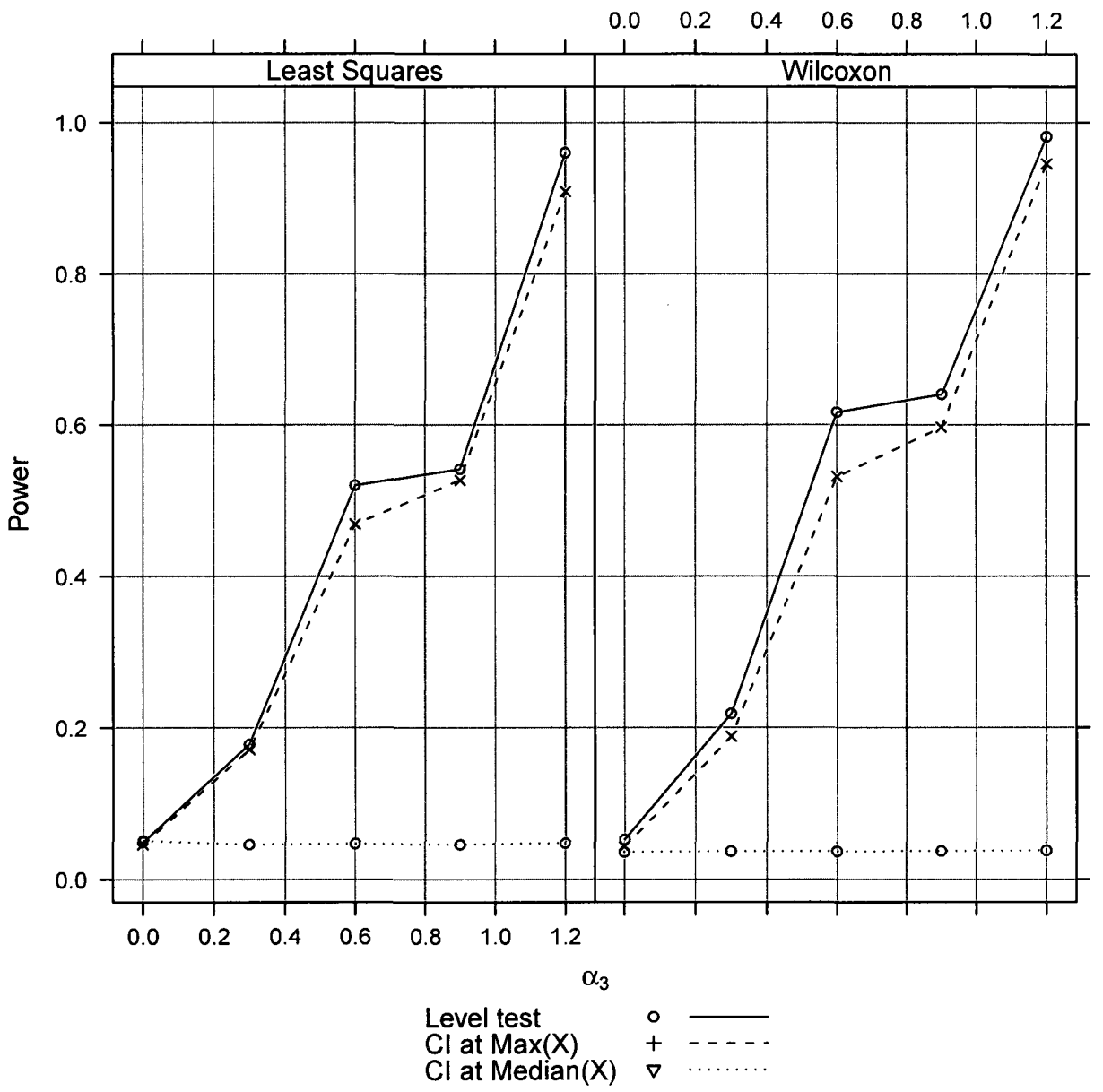

Figure 6: Y is Laplace, $n = 40$

The power of the treatment effects test based on R estimate is slightly higher than the least squares procedure at all values of level of significance and of α_3. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 10.03% and WWIL = 10.28%, at 5% level of significance: LS = 4.91% and WWIL = 5.24%, and at 1% level of significance: LS = 0.95% and WWIL = 1.04%. Moreover, the powers of both procedures are higher when α_3 is higher (Table 10).

For the power of the 95% confidence interval at pick-a-point, the powers of both LS and R procedures at $\alpha_3 = 0$ is close to 5% at all pick-a-points. The power based on R estimate is higher than the least square procedure except at all points when $\alpha_3 = 0$ and at the median of X (q_2) for all values of α_3. Both procedures have higher power when α_3 is higher except at q_2. The powers at q_2 are all close to 5% for all values of α_3 (Table 11). Figure 6 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R procedures for $n = 40$.

Table 12: H_0: $\beta = 0$ when Y is Laplace, $n = 80$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.1049	0.0534	0.0113
	WWIL	0.1061	0.0548	0.0118
$\alpha_3 = 0.3$	LS	0.3294	0.2206	0.0773
	WWIL	0.4078	0.2893	0.1142
$\alpha_3 = 0.45$	LS	0.6974	0.5808	0.3313
	WWIL	0.8078	0.7116	0.4622
$\alpha_3 = 0.6$	LS	0.8272	0.7334	0.4983
	WWIL	0.9120	0.8517	0.6594
$\alpha_3 = 0.9$	LS	0.9611	0.9249	0.7967
	WWIL	0.9884	0.9740	0.9082
$\alpha_3 = 1.2$	LS	0.9992	0.9979	0.9855
	WWIL	1.0000	1.0000	0.9973

Table 13: Pick-a-point when Y is Laplace, $n = 80$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0524	0.0534	0.0482	0.0503	0.0562
	WWIL	0.0511	0.0507	0.0453	0.0480	0.0522
$\alpha_3 = 0.3$	LS	0.2045	0.2207	0.0497	0.0863	0.1895
	WWIL	0.2473	0.2701	0.0440	0.0907	0.2399
$\alpha_3 = 0.45$	LS	0.5498	0.5806	0.0493	0.1461	0.4909
	WWIL	0.6682	0.6991	0.0433	0.1682	0.6001
$\alpha_3 = 0.6$	LS	0.6934	0.7327	0.0473	0.1877	0.6738
	WWIL	0.8128	0.8441	0.0380	0.2310	0.7927
$\alpha_3 = 0.9$	LS	0.8818	0.9242	0.0475	0.6632	0.9420
	WWIL	0.9505	0.9745	0.0398	0.7798	0.9814
$\alpha_3 = 1.2$	LS	0.9919	0.9979	0.0488	0.8769	0.9961
	WWIL	0.9990	1.0000	0.0422	0.9433	0.9996

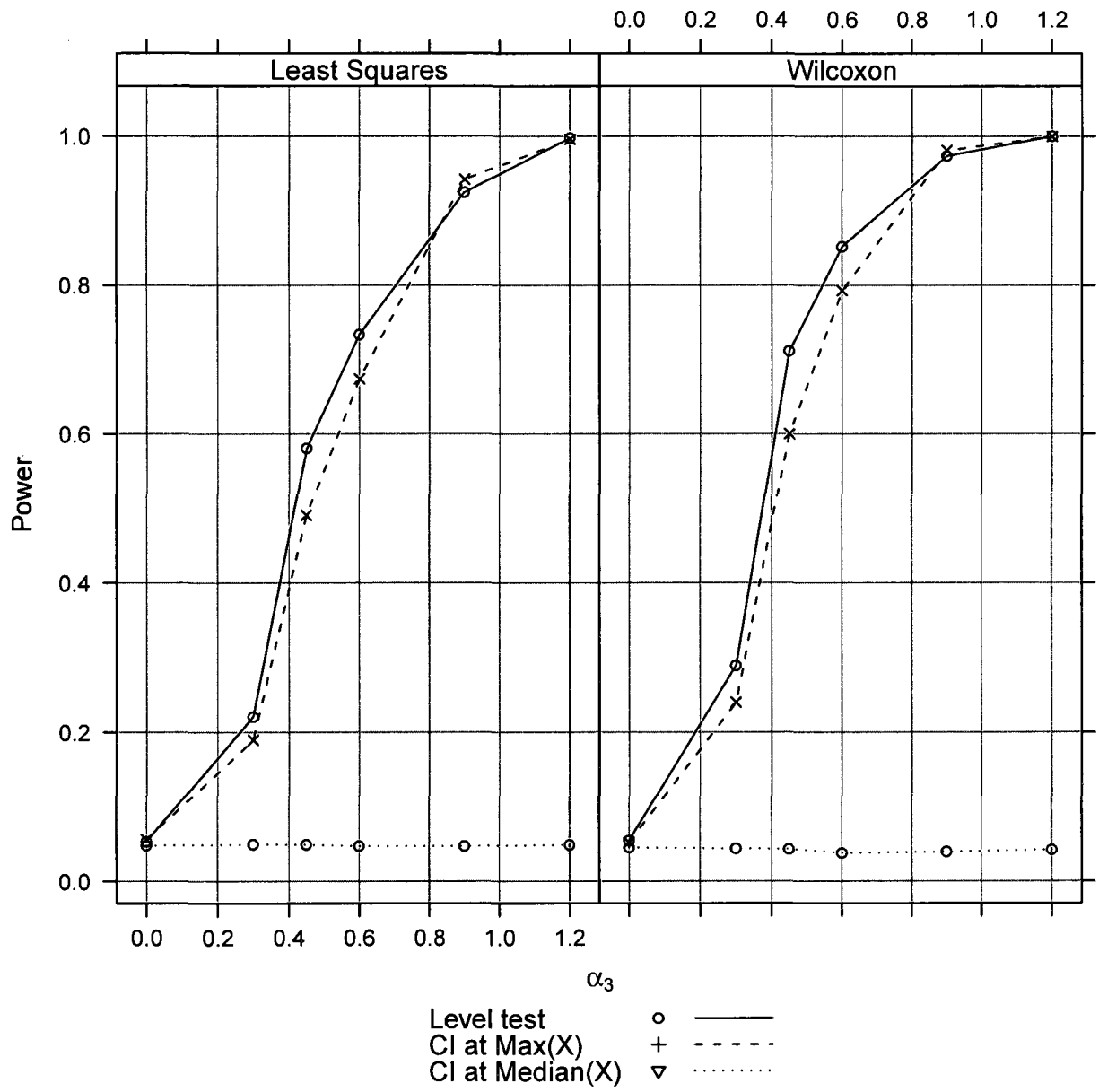

Figure 7: Y is Laplace, $n = 80$

The power of the treatment effects test based on R estimate is slightly higher than the least squares procedure at all values of level of significance and of α_3. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are slightly higher to its level of significance; e.g., at 10% level of significance: LS = 10.49% and WWIL = 10.61%, at 5% level of significance: LS = 5.34% and WWIL = 5.48%, and at 1% level of significance: LS = 1.13% and WWIL = 1.18%. Moreover, the powers of both procedures are higher when α_3 is higher (Table 12).

For the power of the 95% confidence interval at pick-a-point, the powers of both LS and R procedures at $\alpha_3 = 0$ is close to 5% at all pick-a-points. The power based on R estimate is higher than the least square procedure except at all points when $\alpha_3 = 0$ and at the median of X (q_2) for all values of α_3. Both procedures have higher power when α_3 is higher except at q_2. The powers at q_2 are all close to 5% for all values of α_3 (Table 13). Figure 7 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R for procedures for $n = 80$.

Comparing among n = 20, 40, and 80, the power of treatment effects test is higher when the α_3 is higher. The R estimate has slightly higher power than the least square procedure except when n = 20, $\alpha_3 = 0$ and 0.3, and level of significance = 1%. When the sample size is larger, both procedures reach 90% of power at the lower value of α_3 (Table 8, 10, and 12). For 95 % level of significance, the power of R estimate is slightly lower than the least square procedure for all cases of $n = 20$. Some pick-a-points of X when $n = 40$ have higher power than the least square. For $n = 80$, the power of R estimate is higher except all points of X when $\alpha_3 = 0$ and at the point of median of X (q_2) when $\alpha_3 = 0.3, 0.45, 0.6, 0.9$, and 1.2 (Table 9, 11, and 13).

Table 14: H_0: $\beta = 0$ when Y is Cauchy, $n = 20$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.1074	0.0661	0.0156
	WWIL	0.0957	0.0457	0.0083
$\alpha_3 = 0.3$	LS	0.1152	0.0561	0.0069
	WWIL	0.1046	0.0524	0.0085
$\alpha_3 = 0.6$	LS	0.1466	0.0898	0.0290
	WWIL	0.2093	0.1225	0.0284
$\alpha_3 = 0.9$	LS	0.1789	0.1124	0.0346
	WWIL	0.2642	0.1665	0.0454
$\alpha_3 = 1.2$	LS	0.2728	0.1869	0.0746
	WWIL	0.5193	0.3893	0.1654
$\alpha_3 = 1.9$	LS	0.3684	0.2855	0.1580
	WWIL	0.7225	0.6169	0.3512

Table 15: Pick-a-point when Y is Cauchy, $n = 20$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0671	0.0662	0.0210	0.0277	0.0368
	WWIL	0.0433	0.0420	0.0200	0.0194	0.0288
$\alpha_3 = 0.3$	LS	0.0462	0.0560	0.0376	0.0484	0.0543
	WWIL	0.0330	0.0399	0.0249	0.0332	0.0408
$\alpha_3 = 0.6$	LS	0.0732	0.0895	0.0220	0.0311	0.0957
	WWIL	0.0809	0.0973	0.0193	0.0244	0.0902
$\alpha_3 = 0.9$	LS	0.0749	0.1125	0.0226	0.0713	0.1187
	WWIL	0.0717	0.1285	0.0180	0.0638	0.1457
$\alpha_3 = 1.2$	LS	0.1622	0.1867	0.0216	0.1094	0.1566
	WWIL	0.2488	0.3150	0.0194	0.1586	0.2586
$\alpha_3 = 1.9$	LS	0.2930	0.2855	0.0269	0.0465	0.1603
	WWIL	0.5701	0.5440	0.0164	0.0435	0.2616

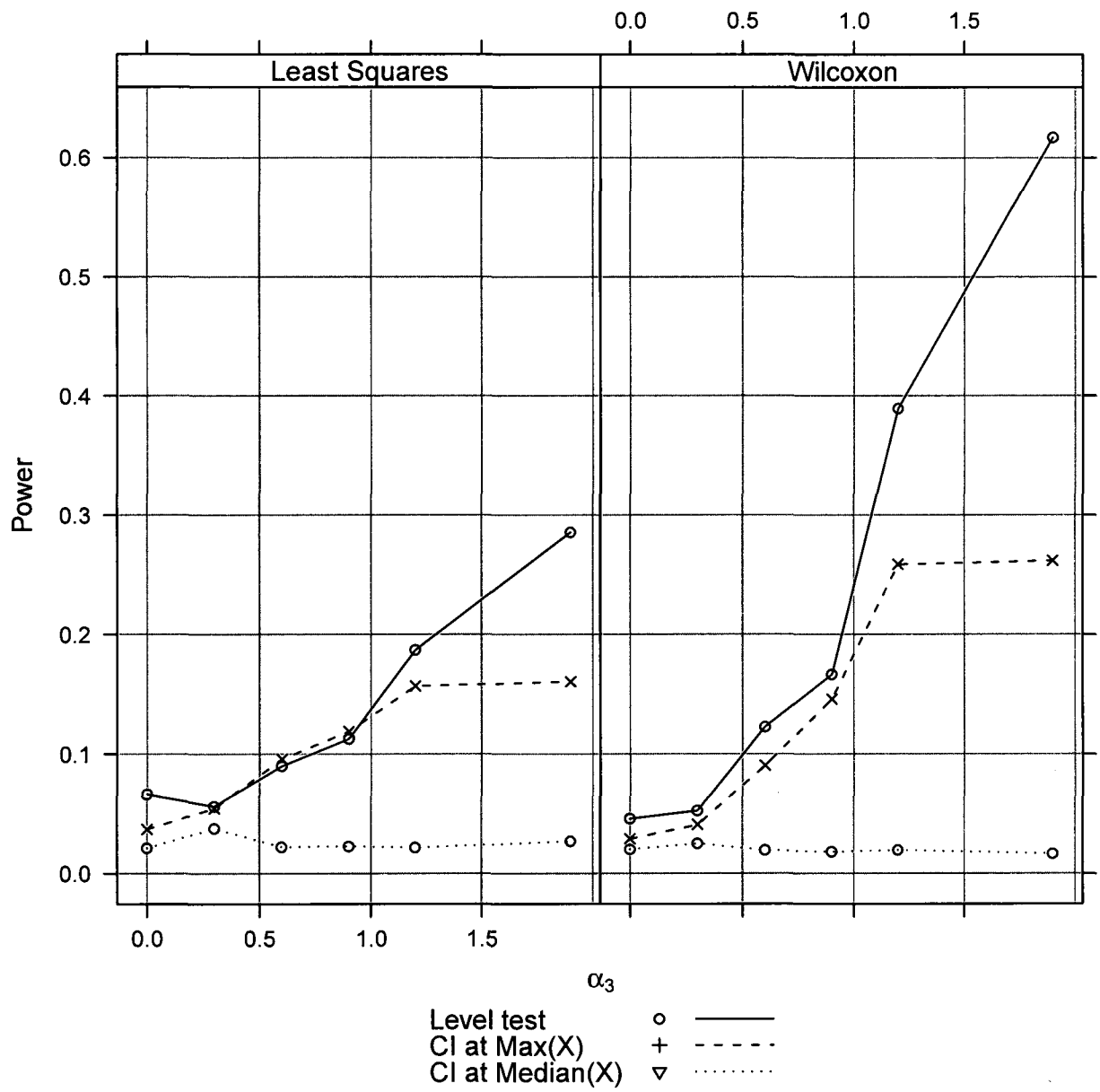

Figure 8: Y is Cauchy, $n = 20$

The power of the treatment effects test based on R estimate is slightly higher than the least squares procedure except at 10% level of significance when $\alpha_3 = 0$ and 0.3, at 5% level of significance when $\alpha_3 = 0$ and 0.3, and at 1% level of significance when $\alpha_3 = 0$. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 10.74% and WWIL = 9.57%, at 5% level of significance: LS = 6.61% and WWIL = 4.57%, and at 1% level of significance: LS = 1.56% and WWIL = 0.83%. Moreover, the powers of both procedures are higher when α_3 is higher. The power of R estimate is approximately 2 times higher than the LS procedure at $\alpha_3 = 1.9$ at all levels of significance. At 10% level of significance and $\alpha_3 = 1.9$, the R estimate already reaches 72% of power, while the power of the least square is only 37% (Table 14).

For the power of the 95% confidence interval at pick-a-point, the powers of both LS and R procedures at $\alpha_3 = 0$ is close to 5% at q_0 and q_1, and is lower at q_2, q_3, and q_4. The power based on R estimate is lower than the least square procedure except at q_0 when $\alpha_3 = 0.6, 1.2$, and 1.9, at q_1 when $\alpha_3 = 0.6$, 0.9, 1.2, and 1.9, at q_3 when $\alpha_3 = 1.2$, and at q_4 when $\alpha_3 = 0.9$, 1.2, and 1.9. Both procedures have higher power when α_3 is higher except at q_2. The powers at q_2 are all lower than 5% for all values of α_3 (Table 15). Figure 8 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R for procedures for $n = 20$.

Table 16: H_0: $\beta = 0$ when Y is Cauchy, $n = 40$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.0949	0.0501	0.0178
	WWIL	0.1089	0.0507	0.0099
$\alpha_3 = 0.3$	LS	0.1021	0.0614	0.0185
	WWIL	0.1700	0.0965	0.0245
$\alpha_3 = 0.6$	LS	0.1472	0.0899	0.0251
	WWIL	0.3651	0.2466	0.0867
$\alpha_3 = 0.9$	LS	0.1446	0.0907	0.0380
	WWIL	0.4316	0.3080	0.1179
$\alpha_3 = 1.2$	LS	0.2145	0.1391	0.0537
	WWIL	0.6670	0.5449	0.3000

Table 17: Pick-a-point when Y is Cauchy, $n = 40$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0491	0.0501	0.0225	0.0380	0.0615
	WWIL	0.0486	0.0499	0.0263	0.0339	0.0495
$\alpha_3 = 0.3$	LS	0.0594	0.0614	0.0240	0.0412	0.0647
	WWIL	0.0765	0.0790	0.0260	0.0413	0.0709
$\alpha_3 = 0.6$	LS	0.0851	0.0900	0.0196	0.0394	0.0727
	WWIL	0.1882	0.2107	0.0266	0.0490	0.1487
$\alpha_3 = 0.9$	LS	0.0870	0.0908	0.0252	0.0456	0.0881
	WWIL	0.2360	0.2658	0.0242	0.0824	0.2280
$\alpha_3 = 1.2$	LS	0.1301	0.1391	0.0222	0.0848	0.1308
	WWIL	0.4468	0.5007	0.0279	0.2216	0.4414

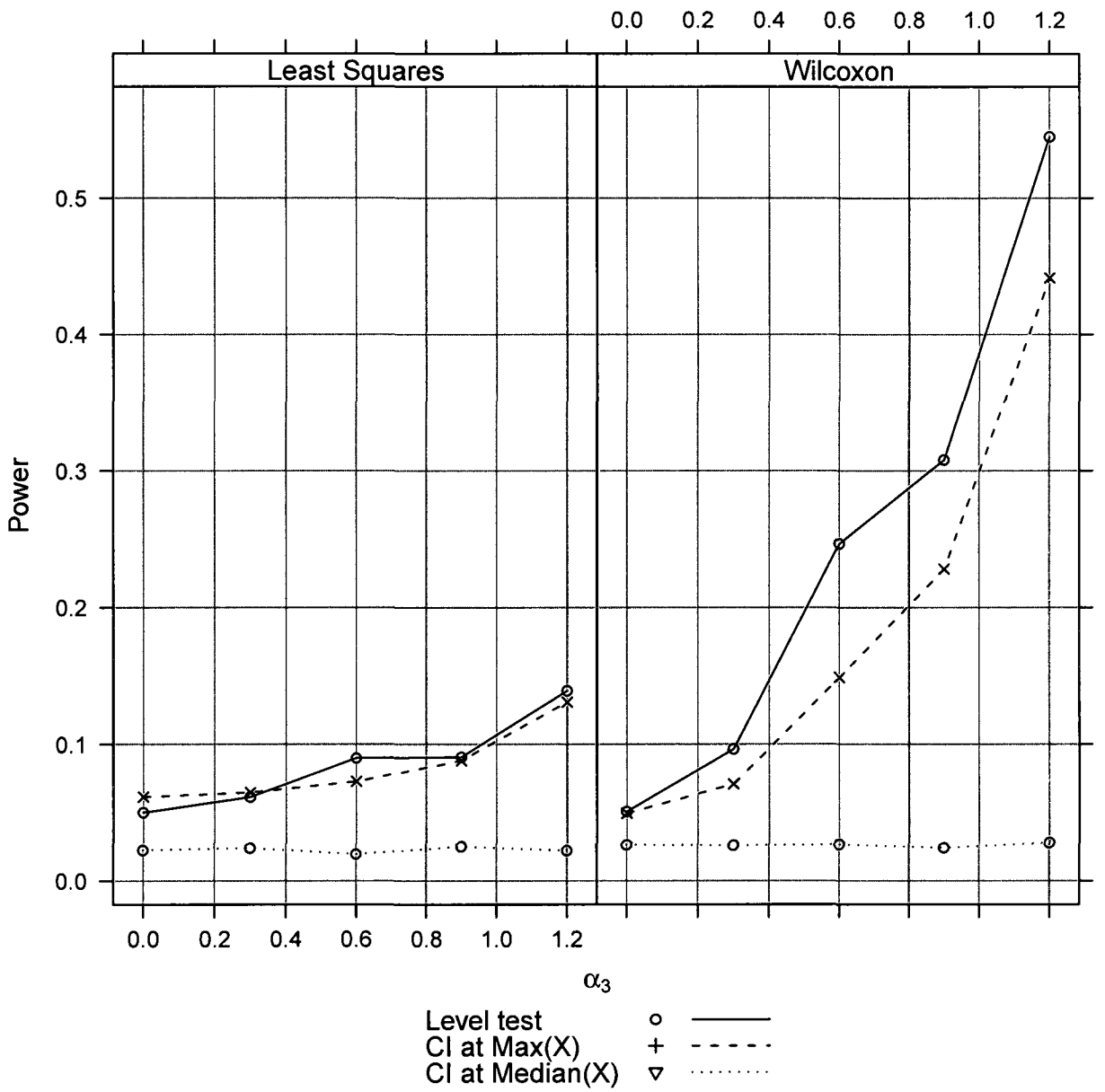

Figure 9: Y is Cauchy, $n = 40$

The power of the treatment effects test based on R estimate is higher than the least squares procedure except at 1% level of significance when $\alpha_3 = 0$. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 9.49% and WWIL = 10.89%, at 5% level of significance: LS = 5.01% and WWIL = 5.07%, and at 1% level of significance: LS = 1.78% and WWIL = 0.99%. Moreover, the powers of both procedures are higher when α_3 is higher. The power of R estimate is 2 times higher at $\alpha_3 = 0.6$ at all levels of significance. At 10% level of significance and $\alpha_3 = 0.6$, the R estimate reaches 36.51% of power, while the power of the least square is only 14.72% (Table 16).

For the power of the 95% confidence interval at pick-a-point, the powers of both LS and R procedures at $\alpha_3 = 0$ is close to 5% at all points. The power based on R estimate is higher than the least square procedure except at q_0 when $\alpha_3 = 0$, at q_1 when $\alpha_3 = 0$, at q_2 when $\alpha_3 = 0.9$, at q_3 when $\alpha_3 = 0$, and at q_4 when $\alpha_3 = 0$. Both procedures have higher power when α_3 is higher except at q_2. The powers at q_2 are all lower than 5% for all values of α_3 (Table 17). Figure 9 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R for procedures for $n = 40$.

Table 18: H_0: $\beta = 0$ when Y is Cauchy, $n = 80$

	Procedure	Power		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\alpha_3 = 0$	LS	0.0861	0.0441	0.0160
	WWIL	0.1051	0.0513	0.0113
$\alpha_3 = 0.3$	LS	0.1246	0.0700	0.0164
	WWIL	0.2395	0.1498	0.0473
$\alpha_3 = 0.45$	LS	0.1267	0.0699	0.0165
	WWIL	0.3566	0.2427	0.0864
$\alpha_3 = 0.6$	LS	0.1438	0.0845	0.0235
	WWIL	0.5896	0.4624	0.2324
$\alpha_3 = 0.9$	LS	0.1641	0.0911	0.0234
	WWIL	0.7501	0.6459	0.4088
$\alpha_3 = 1.2$	LS	0.2323	0.1545	0.0664
	WWIL	0.9448	0.9076	0.7756

Table 19: Pick-a-point when Y is Cauchy, $n = 80$

	Procedure	Power				
		q_0	q_1	q_2	q_3	q_4
$\alpha_3 = 0$	LS	0.0448	0.0441	0.0283	0.0386	0.0480
	WWIL	0.0529	0.0523	0.0370	0.0406	0.0451
$\alpha_3 = 0.3$	LS	0.0654	0.0699	0.0188	0.0439	0.0615
	WWIL	0.1218	0.1367	0.0334	0.0667	0.1256
$\alpha_3 = 0.45$	LS	0.0669	0.0698	0.0202	0.0463	0.0634
	WWIL	0.1916	0.2223	0.0319	0.1012	0.2157
$\alpha_3 = 0.6$	LS	0.0889	0.0845	0.0229	0.0357	0.0555
	WWIL	0.4370	0.4420	0.0335	0.1038	0.2640
$\alpha_3 = 0.9$	LS	0.0749	0.0909	0.0225	0.0654	0.0998
	WWIL	0.5074	0.6289	0.0380	0.3437	0.6518
$\alpha_3 = 1.2$	LS	0.1440	0.1544	0.0222	0.1018	0.1546
	WWIL	0.8863	0.9109	0.0347	0.7126	0.9137

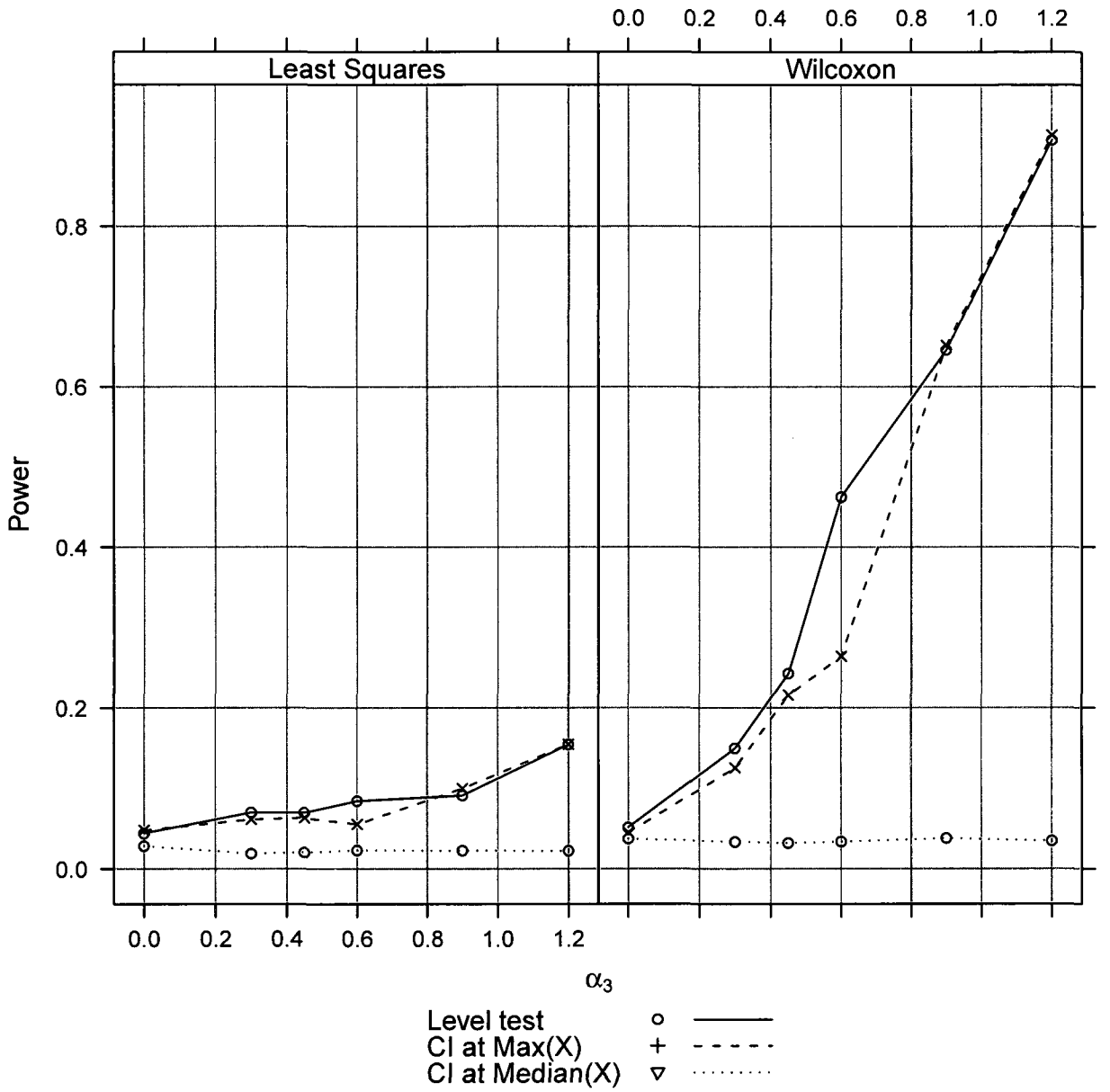

Figure 10: Y is Cauchy, $n = 80$

The power of the treatment effects test based on R estimate is higher than the least squares procedure except at 1% level of significance when $\alpha_3 = 0$. The powers at $\alpha_3 = 0$ at 10%, 5%, and 1% level of significance are similar to its level of significance; e.g., at 10% level of significance: LS = 8.61% and WWIL = 10.51%, at 5% level of significance: LS = 4.41% and WWIL = 5.13%, and at 1% level of significance: LS = 1.60% and WWIL = 1.13%. Moreover, the powers of both procedures are higher when α_3 is higher. The power of the R estimate has 2 times higher of power than the LS procedure when $\alpha_3 = 0.3$, and has much more higher power at $\alpha_3 = 1.2$ for all levels of significance. At 10% level of significance and $\alpha_3 = 1.2$, the R estimate reaches 94.48% of power, while the power of the least square is only 23.23% (Table 18).

For the power of the 95% confidence interval at pick-a-point, the powers of both LS and R procedures at $\alpha_3 = 0$ is close to 5% at all points. The power based on R estimate is higher than the least square procedure except at q_4 when $\alpha_3 = 0$. Both procedures have higher power when α_3 is higher except at q_2. The powers at q_2 are all lower than 5% for all values of α_3 (Table 19). Figure 10 shows the plots of the power for the treatment test and for 95% confidence interval at q_2 and q_4 between LS and R for procedures for $n = 80$.

When comparing among the different sample sizes, the power of the treatment effects test for the R procedure reaches 2 times higher than that of the LS procedure at the lower value of α_3 when the sample size is larger. The powers of both procedures are higher when the α_3 is higher. The R estimate has higher power in some pick-a-points of X in some cases. It seems that R estimate gains more power when the sample size and the α_3 are larger; for instance, at $\alpha_3 = 1.2$, the powers of the maximum of X (q_4) of n = 20, 40, and 80 based on R estimate are 0.2586, 0.4414, and 0.9137, respectively (Table 15, 17, and 19).

When comparing the power among three distributions, the Cauchy distribution seems to be worst for the least square and R methods. However, the power of R is higher

when the sample size and the α_3 are higher.

2.4 Conditional Test

Several conditional tests are used in practice. We describe several such tests. We are now interested in doing the conditional test: (1) for pick-a-point (condition A), and (2) for analysis of variance (condition B). We compare conditional tests with the pick-a-point at the grand mean ($\bar{x}$).

Method I: Condition A

1. Test homogeneity of slopes

$$H_{01} : \beta_1 = \ldots = \beta_k$$

1.1 If the null hypothesis is accepted, we then do the pooled level test (ANCOVA)

$$H_{02} : \beta_i = \beta_j, \quad i \neq j$$

1.2 If the null hypothesis is rejected, we then do the pick-a-point at grand mean of the X's

Method II: Condition B

1. Test homogeneity of slopes

$$H_{01} : \beta_1 = \ldots = \beta_k$$

1.1 If the null hypothesis is accepted, we then do the pooled level test (ANCOVA)

$$H_{02} : \beta_i = \beta_j, \quad i \neq j$$

1.2 If the null hypothesis is rejected, we then do the analysis of variance

Method III

1. Test homogeneity of slopes

$$H_{01} : \beta_1 = \ldots = \beta_k$$

2. Pick-a-point at grand mean of the X's

We are interested in all these questions:

1. Which methods are more powerful?

2. Is there a difference in the powers of the traditional and rank-based analysis?

To answer these questions, we conduct the following empirical study. Recall the matrix X (2.6) for the case of two groups with one covariate, and we consider the point at mean of X instead of median of X. Let $\alpha_1 = -(\mu_x - 4 \times \sigma_x) \cdot \alpha_3$. We conduct a simulation in which the sample sizes are 20, 40, and 80. The covariate is generated from normal distribution with $\mu = 100$ and $\sigma = 20$. The response variable is generated from:

(1) Normal distribution

(2) Laplace distribution

(3) Cauchy distribution

The powers of the homogeneity of slopes at 10%, 5%, and 1% level of significance are also obtained. For each scenario, 10000 simulations are conducted at 5% level of significance.

2.4.1 Condition A and Condition B Simulation Results

Table 20: Condition A when Y is normal

	Procedure	Power		
		$n = 20$	$n = 40$	$n = 80$
$\alpha_3 = 0$	LS	0.0481	0.0525	0.0530
	WWIL	0.0432	0.0501	0.0486
$\alpha_3 = 0.3$	LS	0.7184	0.9535	0.9982
	WWIL	0.6557	0.9364	0.9972
$\alpha_3 = 0.6$	LS	0.9894	1.0000	1.0000
	WWIL	0.9763	1.0000	1.0000
$\alpha_3 = 0.9$	LS	1.0000	1.0000	1.0000
	WWIL	1.0000	1.0000	1.0000

Table 21: Condition B when Y is normal

	Procedure	Power		
		$n = 20$	$n = 40$	$n = 80$
$\alpha_3 = 0$	LS	0.0462	0.0520	0.0527
	WWIL	0.0415	0.0488	0.0482
$\alpha_3 = 0.3$	LS	0.7099	0.9572	0.9981
	WWIL	0.6496	0.9407	0.9969
$\alpha_3 = 0.6$	LS	0.9882	1.0000	1.0000
	WWIL	0.9741	1.0000	1.0000
$\alpha_3 = 0.9$	LS	1.0000	1.0000	1.0000
	WWIL	1.0000	1.0000	1.0000

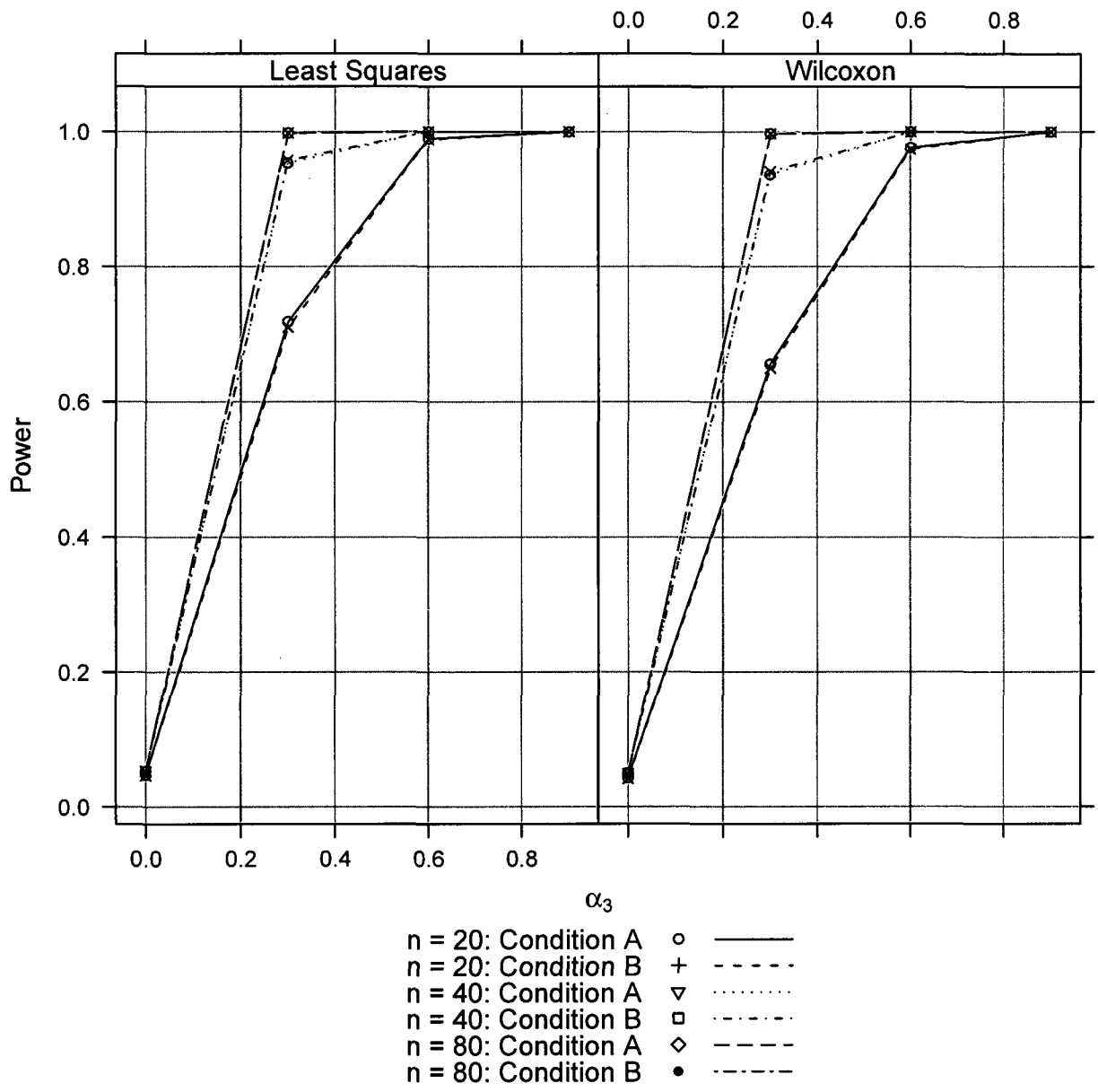

Figure 11: Condition A and B when Y is normal

For normal distribution, the powers of both least squares and R estimate procedures are slightly different in both Condition A and B, they are identical when $\alpha_3 \geq 0.6$ and

$n \geq 40$ (Table 20, 21). At $\alpha_3 = 0$, the powers of both LS and robust procedures are close to 5% at all different sample sizes for both conditions. When the sample size is larger, the power is higher as well except for WWIL at $\alpha_3 = 0$ and $n = 80$ for both procedures. Figure 11 shows the power of Condition A and B at $n = 20$, 40, and 80.

Table 22: Condition A when Y is Laplace

	Procedure	Power		
		$n = 20$	$n = 40$	$n = 80$
$\alpha_3 = 0$	LS	0.0487	0.0524	0.0482
	WWIL	0.0395	0.0474	0.0468
$\alpha_3 = 0.3$	LS	0.6347	0.9481	0.9991
	WWIL	0.6797	0.9787	0.9999
$\alpha_3 = 0.6$	LS	0.9684	1.0000	1.0000
	WWIL	0.9779	1.0000	1.0000
$\alpha_3 = 0.9$	LS	1.0000	1.0000	1.0000
	WWIL	1.0000	1.0000	1.0000

Table 23: Condition B when Y is Laplace

	Procedure	Power		
		$n = 20$	$n = 40$	$n = 80$
$\alpha_3 = 0$	LS	0.0467	0.0513	0.0480
	WWIL	0.0372	0.0465	0.0462
$\alpha_3 = 0.3$	LS	0.6242	0.9452	0.9993
	WWIL	0.6693	0.9769	0.9999
$\alpha_3 = 0.6$	LS	0.9667	1.0000	1.0000
	WWIL	0.9765	1.0000	1.0000
$\alpha_3 = 0.9$	LS	1.0000	1.0000	1.0000
	WWIL	1.0000	1.0000	1.0000

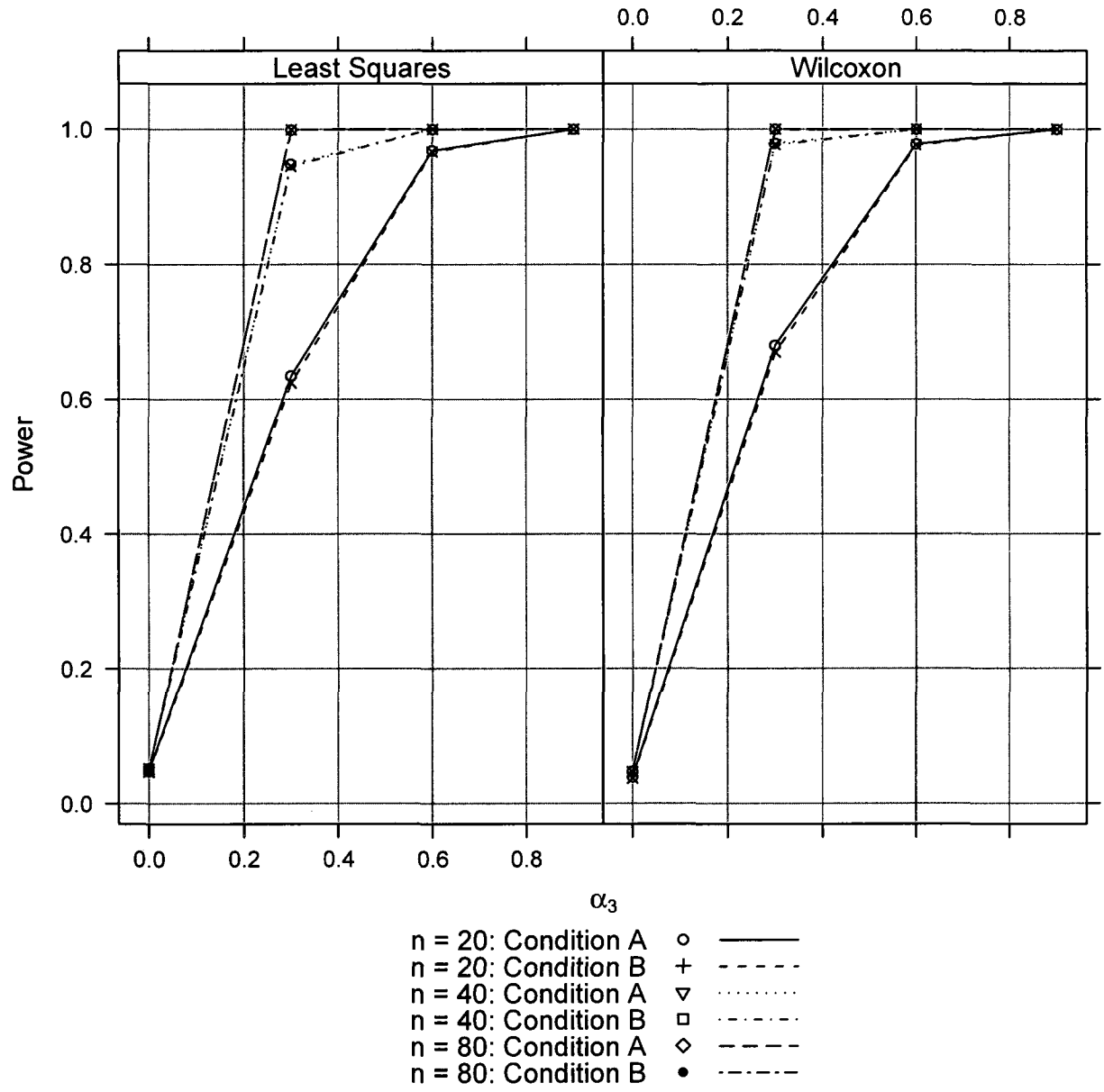

Figure 12: Condition A and B when Y is Laplace

For Laplace distribution, the powers of both least squares and R estimate procedures are slightly different in both Condition A and B, they are identical when $\alpha_3 \geq 0.6$ and

$n \geq 40$ (Table 22, 23). At $\alpha_3 = 0$, the powers of both LS and robust procedures are close to 5% at all different sample sizes for both conditions. When the sample size is larger, the power is higher as well except for LS and WWIL at $\alpha_3 = 0$ and $n = 80$ for both procedures. Figure 12 shows the power of Condition A and B at $n = 20$, 40, and 80.

Table 24: Condition A when Y is Cauchy

	Procedure	Power		
		$n = 20$	$n = 40$	$n = 80$
$\alpha_3 = 0$	LS	0.0500	0.0232	0.0207
	WWIL	0.0680	0.0373	0.0420
$\alpha_3 = 0.3$	LS	0.1686	0.1763	0.1620
	WWIL	0.3862	0.7382	0.9339
$\alpha_3 = 0.6$	LS	0.3553	0.4143	0.3894
	WWIL	0.7305	0.9864	0.9999
$\alpha_3 = 0.9$	LS	0.5426	0.5390	0.5517
	WWIL	0.9289	0.9979	1.0000

Table 25: Condition B when Y is Cauchy

	Procedure	Power		
		$n = 20$	$n = 40$	$n = 80$
$\alpha_3 = 0$	LS	0.0210	0.0222	0.0200
	WWIL	0.0243	0.0367	0.0415
$\alpha_3 = 0.3$	LS	0.1660	0.1738	0.1605
	WWIL	0.3843	0.7389	0.9331
$\alpha_3 = 0.6$	LS	0.3494	0.4127	0.3875
	WWIL	0.7476	0.9867	0.9998
$\alpha_3 = 0.9$	LS	0.5391	0.5387	0.5507
	WWIL	0.9304	0.9980	1.0000

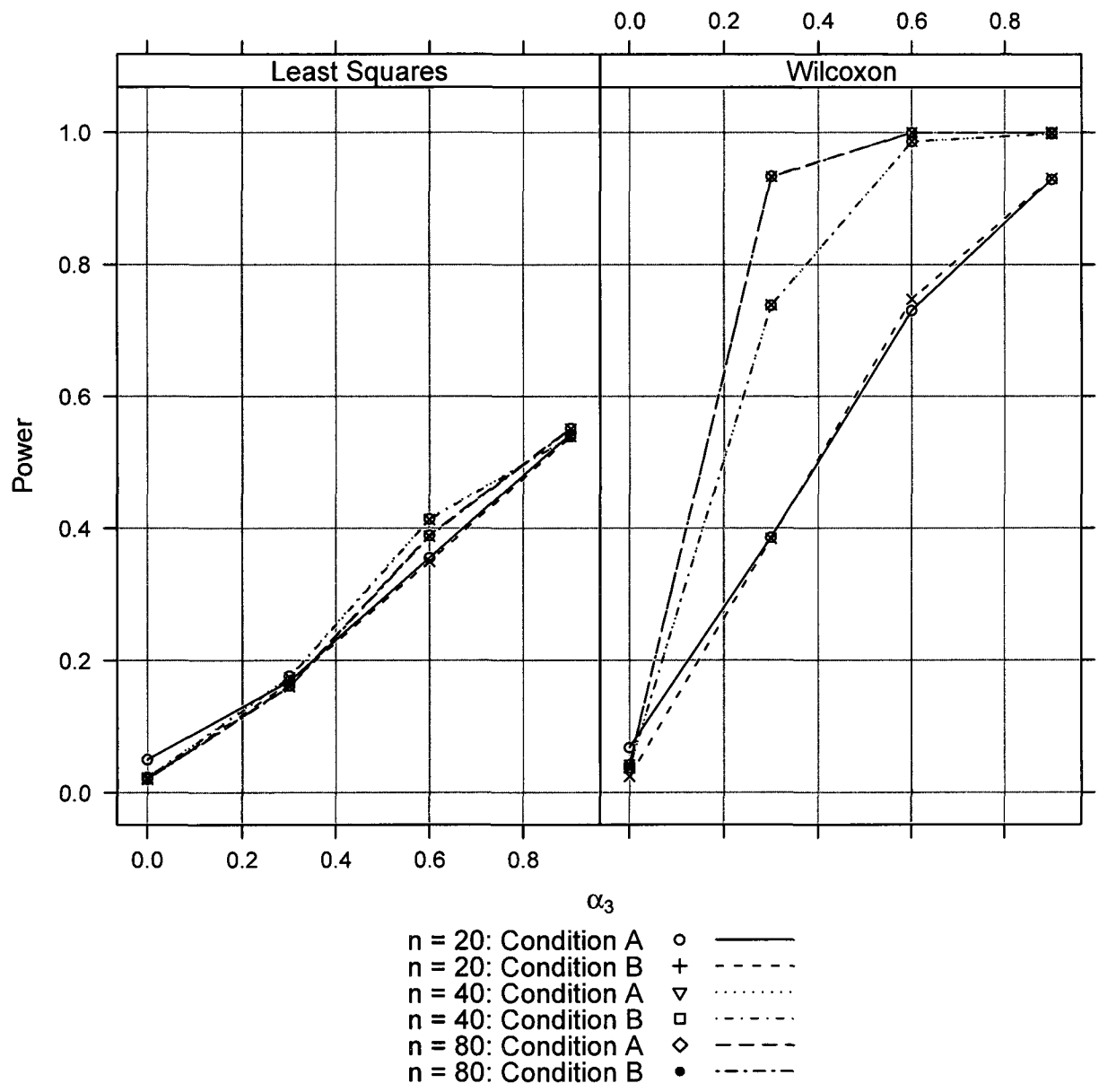

Figure 13: Condition A and B when Y is Cauchy

In case of Cauchy distribution, the power of the R estimate procedures are higher than the LS procedure in both Condition A and B (Table 24, 25). At all values of α_3, the

robust procedures have more powerful for all values of α_3 and all sample size. The power of R procedure is about 2 times higher than that of the least squares when the α_3 and n are higher in both procedures. Figure 13 shows the power of Condition A and B at $n = 20$, 40, and 80.

For normal and Laplace cases, the powers of both least squares and R estimate procedures are slightly different in both Condition A and B, they are identical when $\alpha_3 \geq 0.6$ and $n \geq 40$ (Table 20, 21, 22, and 23). In case of Cauchy, the least squares and the R procedures are close to each other when $\alpha_3 = 0$ in both conditions. However, the power of R procedure is about 2 times higher than that of the least squares when the α_3 and n are higher (Table 24, and 25).

2.4.2 Method III Simulation Results

Table 26: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is normal, $n = 20$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.1028	0.0502	0.0094	0.0459
	WWIL	0.0990	0.0477	0.0090	0.0313
$\alpha_3 = 0.3$	LS	0.1524	0.0830	0.0219	0.7104
	WWIL	0.1431	0.0777	0.0168	0.5643
$\alpha_3 = 0.6$	LS	0.2495	0.1549	0.0465	0.9899
	WWIL	0.2311	0.1371	0.0359	0.9578
$\alpha_3 = 0.9$	LS	0.4493	0.3232	0.1265	1.0000
	WWIL	0.4191	0.2843	0.0984	1.0000
$\alpha_3 = 1.2$	LS	0.7236	0.5879	0.3095	1.0000
	WWIL	0.6901	0.5401	0.2493	1.0000

The power of the LS and R procedures are slightly different for all α_3 and for all levels of significance. The power of the 95% confidence interval at grand mean of both procedures is higher when the α_3 is higher (Table 26, 27, 28). The plots of the power for homogeneity at 5% level of significance and that for pick-a-point at grand mean for normal

Table 27: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is normal, $n = 40$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.0981	0.0488	0.0102	0.0521
	WWIL	0.1033	0.0522	0.0088	0.0444
$\alpha_3 = 0.3$	LS	0.2327	0.1450	0.0478	0.9504
	WWIL	0.2299	0.1366	0.0426	0.9196
$\alpha_3 = 0.6$	LS	0.4915	0.3645	0.1574	1.0000
	WWIL	0.4741	0.3473	0.1418	1.0000
$\alpha_3 = 0.9$	LS	0.7594	0.6451	0.3857	1.0000
	WWIL	0.7367	0.6193	0.3424	1.0000
$\alpha_3 = 1.2$	LS	0.9363	0.8808	0.6957	1.0000
	WWIL	0.9261	0.8636	0.6462	1.0000

Table 28: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is normal, $n = 80$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.1045	0.0507	0.0097	0.0522
	WWIL	0.1034	0.0475	0.0093	0.0458
$\alpha_3 = 0.3$	LS	0.3851	0.2751	0.1101	0.9983
	WWIL	0.3727	0.2602	0.0994	0.9968
$\alpha_3 = 0.6$	LS	0.7692	0.6633	0.4150	1.0000
	WWIL	0.7508	0.6423	0.3834	1.0000
$\alpha_3 = 0.9$	LS	0.9693	0.9380	0.8165	1.0000
	WWIL	0.9608	0.9236	0.7857	1.0000
$\alpha_3 = 1.2$	LS	0.9999	0.9997	0.9880	1.0000
	WWIL	0.9998	0.9997	0.9978	1.0000

distribution is shown in Figure 14 - 16.

In case of Laplace distribution, the power of the LS and R procedures are slightly different for all α_3 and all levels of significance. The power of the 95% confidence interval at grand mean of both procedures is higher when the α_3 is higher (Table 29, 30, 31). The plots of the power for homogeneity at 5% level of significance and that for pick-a-point at grand mean is shown in Figure 17 - 19.

Table 29: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is Laplace, $n = 20$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.1024	0.0515	0.0120	0.0475
	WWIL	0.1083	0.0519	0.0097	0.0299
$\alpha_3 = 0.3$	LS	0.1477	0.0864	0.0253	0.6400
	WWIL	0.1690	0.0987	0.0242	0.6244
$\alpha_3 = 0.6$	LS	0.2520	0.1566	0.0492	0.9704
	WWIL	0.2669	0.1672	0.0456	0.9683
$\alpha_3 = 0.9$	LS	0.6105	0.4738	0.2378	1.0000
	WWIL	0.6563	0.5186	0.2450	1.0000

Table 30: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is Laplace, $n = 40$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.1055	0.0537	0.0099	0.0515
	WWIL	0.1089	0.0563	0.0087	0.0398
$\alpha_3 = 0.3$	LS	0.2310	0.1441	0.0436	0.9480
	WWIL	0.2667	0.1691	0.0545	0.9743
$\alpha_3 = 0.6$	LS	0.5085	0.3833	0.1831	1.0000
	WWIL	0.5898	0.4612	0.2339	1.0000
$\alpha_3 = 0.9$	LS	0.8672	0.7828	0.5621	1.0000
	WWIL	0.9234	0.8607	0.6778	1.0000

Table 31: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is Laplace, $n = 80$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.0972	0.0491	0.0097	0.0478
	WWIL	0.0986	0.0497	0.0099	0.0435
$\alpha_3 = 0.3$	LS	0.3430	0.2423	0.0900	0.9991
	WWIL	0.4339	0.3148	0.1315	0.9999
$\alpha_3 = 0.6$	LS	0.7815	0.6767	0.4338	1.0000
	WWIL	0.8843	0.8090	0.5850	1.0000
$\alpha_3 = 0.9$	LS	0.9931	0.9840	0.9400	1.0000
	WWIL	0.9982	0.9960	0.9791	1.0000

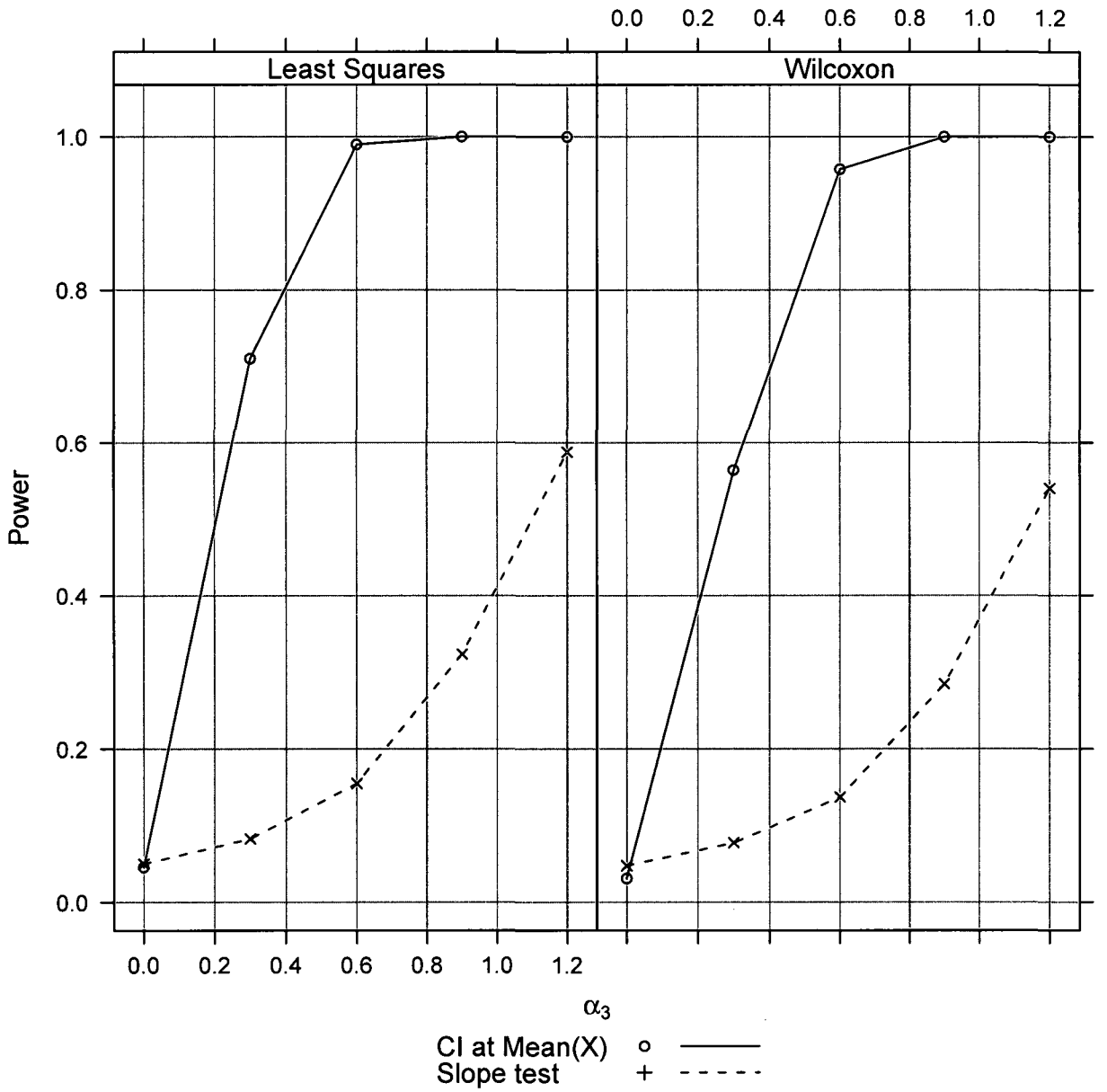

Figure 14: Y is normal, $n = 20$

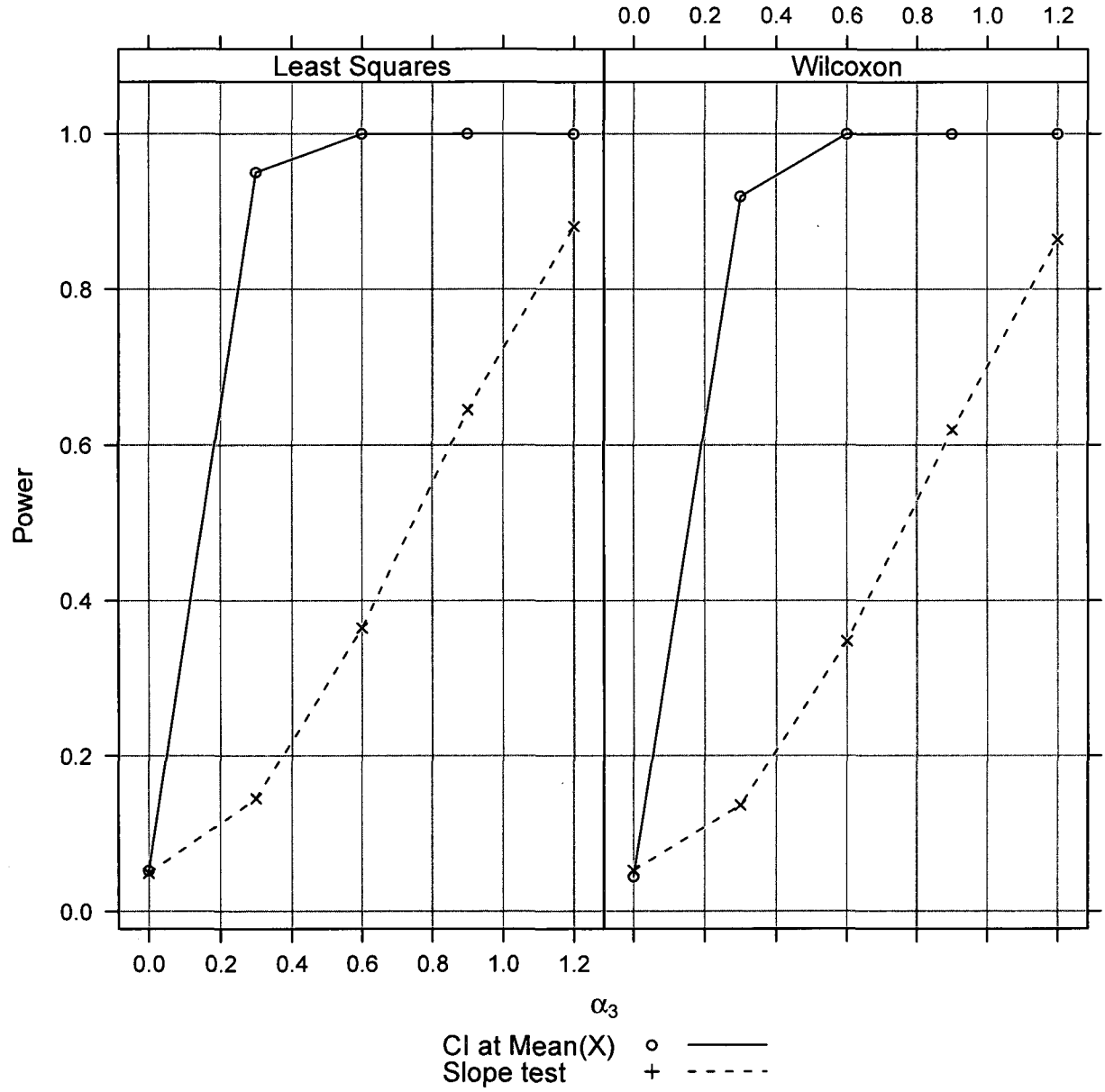

Figure 15: Y is normal, $n = 40$

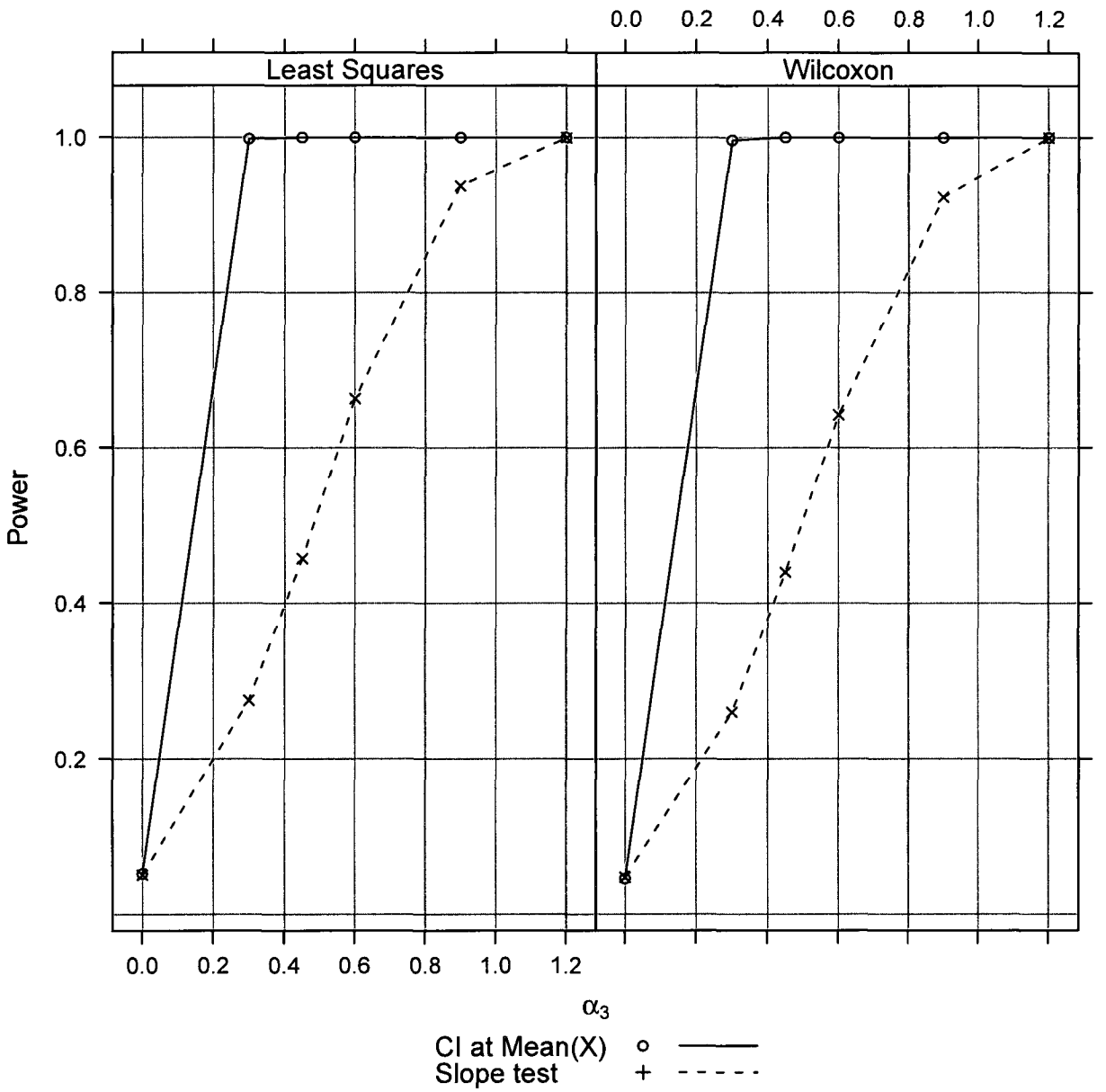

Figure 16: Y is normal, $n = 80$

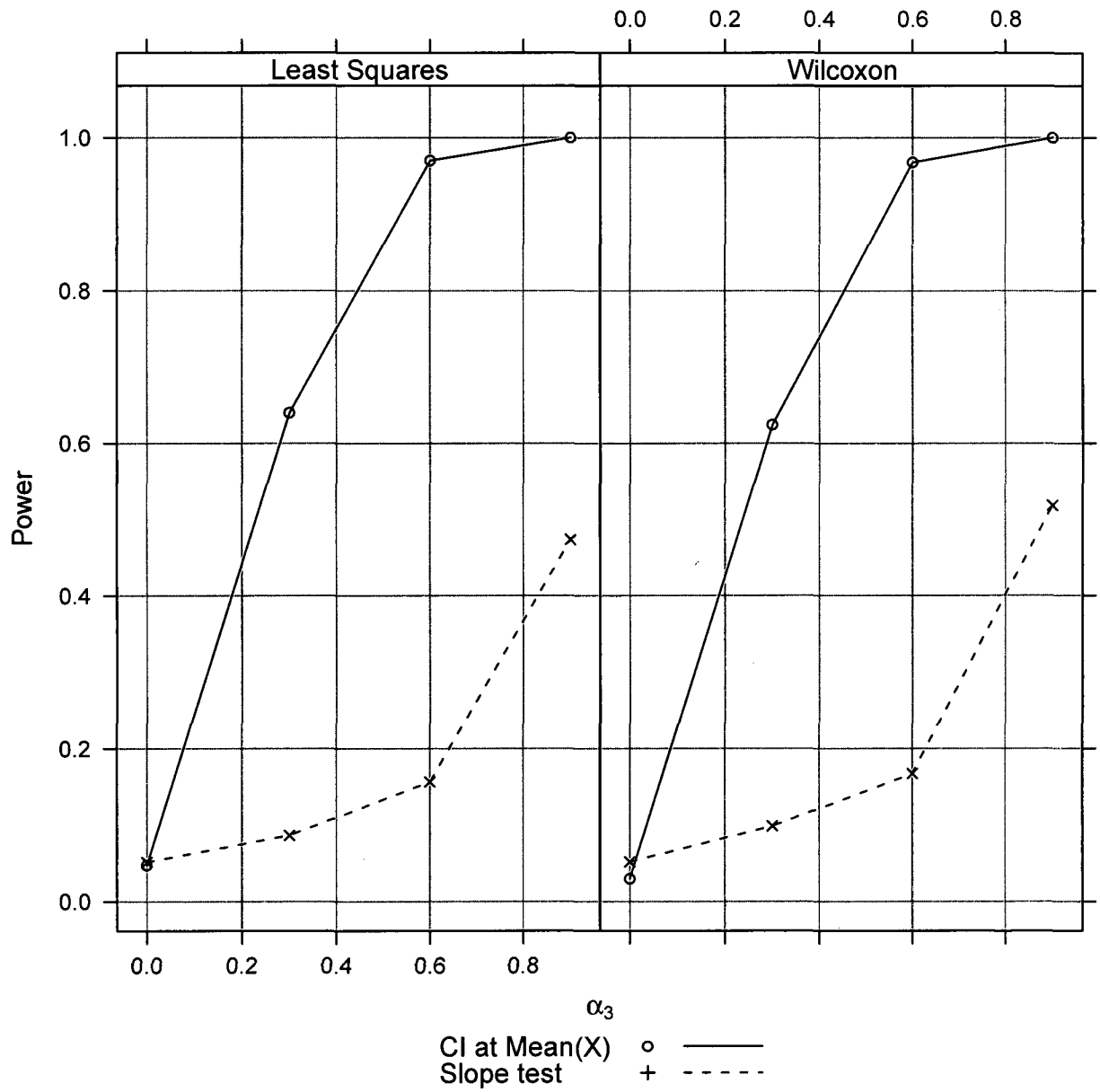

Figure 17: Y is Laplace, $n = 20$

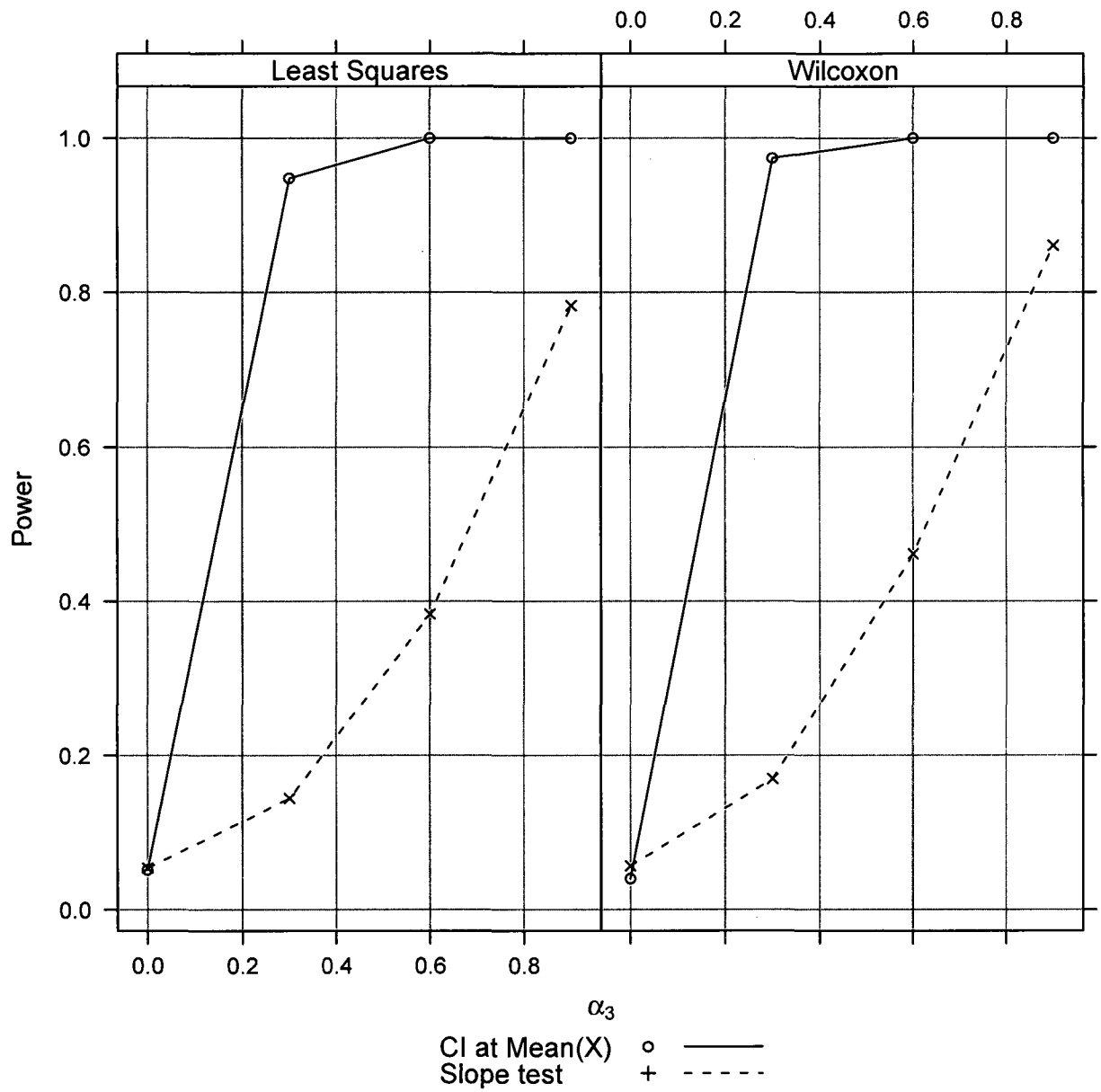

Figure 18: Y is Laplace, $n = 40$

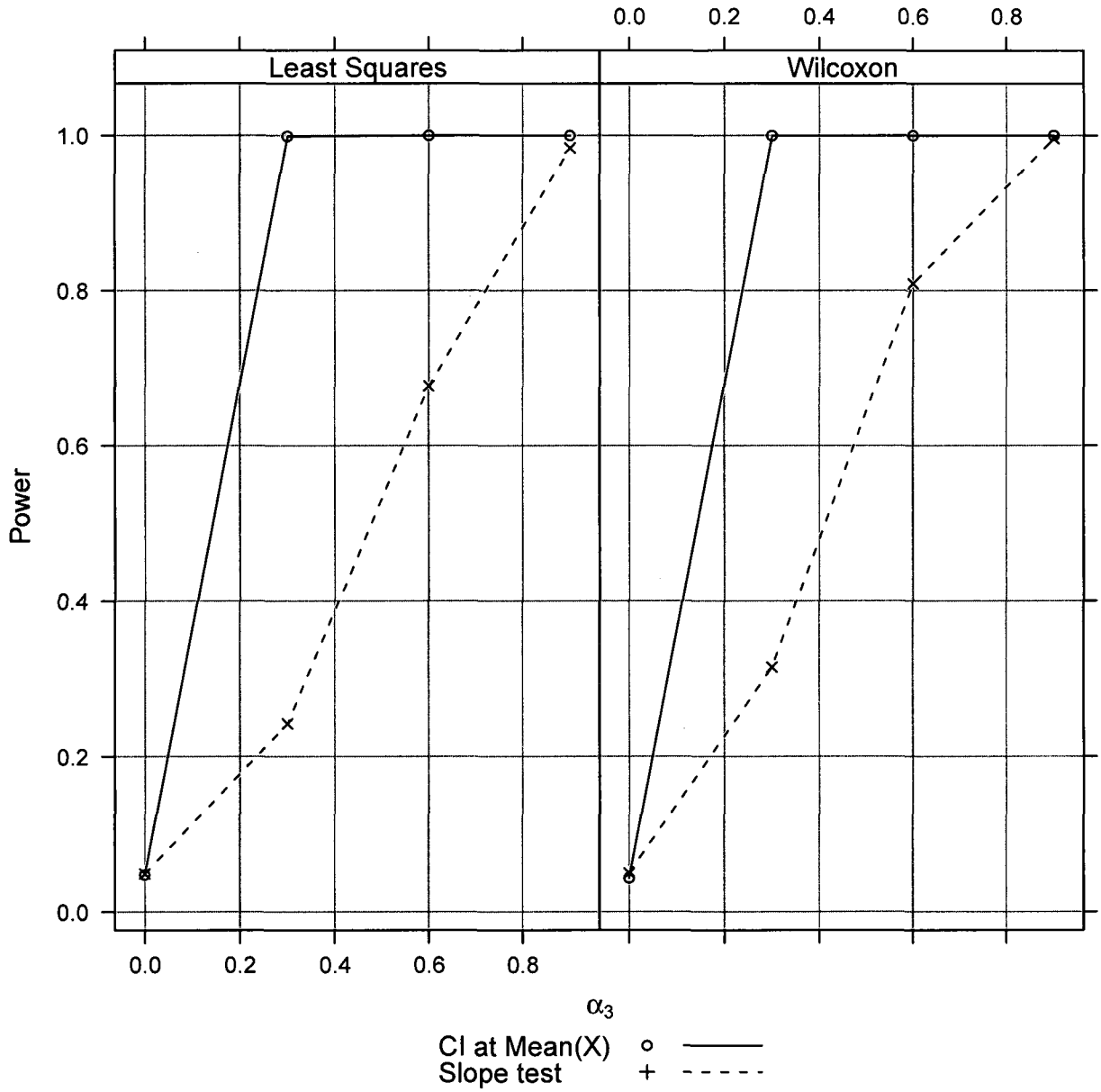

Figure 19: Y is Laplace, $n = 80$

Table 32: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is Cauchy, $n = 20$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.0929	0.0649	0.0387	0.0481
	WWIL	0.1602	0.1057	0.0332	0.0637
$\alpha_3 = 0.3$	LS	0.1024	0.0411	0.0054	0.1637
	WWIL	0.1057	0.0463	0.0069	0.2904
$\alpha_3 = 0.6$	LS	0.1547	0.1088	0.0465	0.3314
	WWIL	0.2412	0.1550	0.0470	0.6371
$\alpha_3 = 0.9$	LS	0.1699	0.0891	0.0218	0.5402
	WWIL	0.2764	0.1718	0.0472	0.9037

Table 33: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is Cauchy, $n = 40$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.1069	0.0533	0.0071	0.0229
	WWIL	0.0906	0.0430	0.0080	0.0278
$\alpha_3 = 0.3$	LS	0.1190	0.0651	0.0102	0.1742
	WWIL	0.1572	0.0856	0.0197	0.6992
$\alpha_3 = 0.6$	LS	0.1469	0.0749	0.0173	0.4136
	WWIL	0.3246	0.2173	0.0746	0.9860
$\alpha_3 = 0.9$	LS	0.1792	0.1071	0.0364	0.5350
	WWIL	0.5485	0.4261	0.2021	0.9976

Table 34: H_0: $\beta_1 = \beta_2$ and pick-a-point at grand mean when Y is Cauchy, $n = 80$

	Procedure	Power			
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Pick-a-point at $\bar{x}$
$\alpha_3 = 0$	LS	0.1061	0.0585	0.0128	0.0207
	WWIL	0.0987	0.0473	0.0091	0.0362
$\alpha_3 = 0.3$	LS	0.1148	0.0571	0.0090	0.1618
	WWIL	0.2520	0.1587	0.0501	0.9352
$\alpha_3 = 0.6$	LS	0.1485	0.0825	0.0247	0.3893
	WWIL	0.6395	0.5151	0.2763	0.9999
$\alpha_3 = 0.9$	LS	0.1709	0.1008	0.0330	0.5515
	WWIL	0.8000	0.7075	0.4745	1.0000

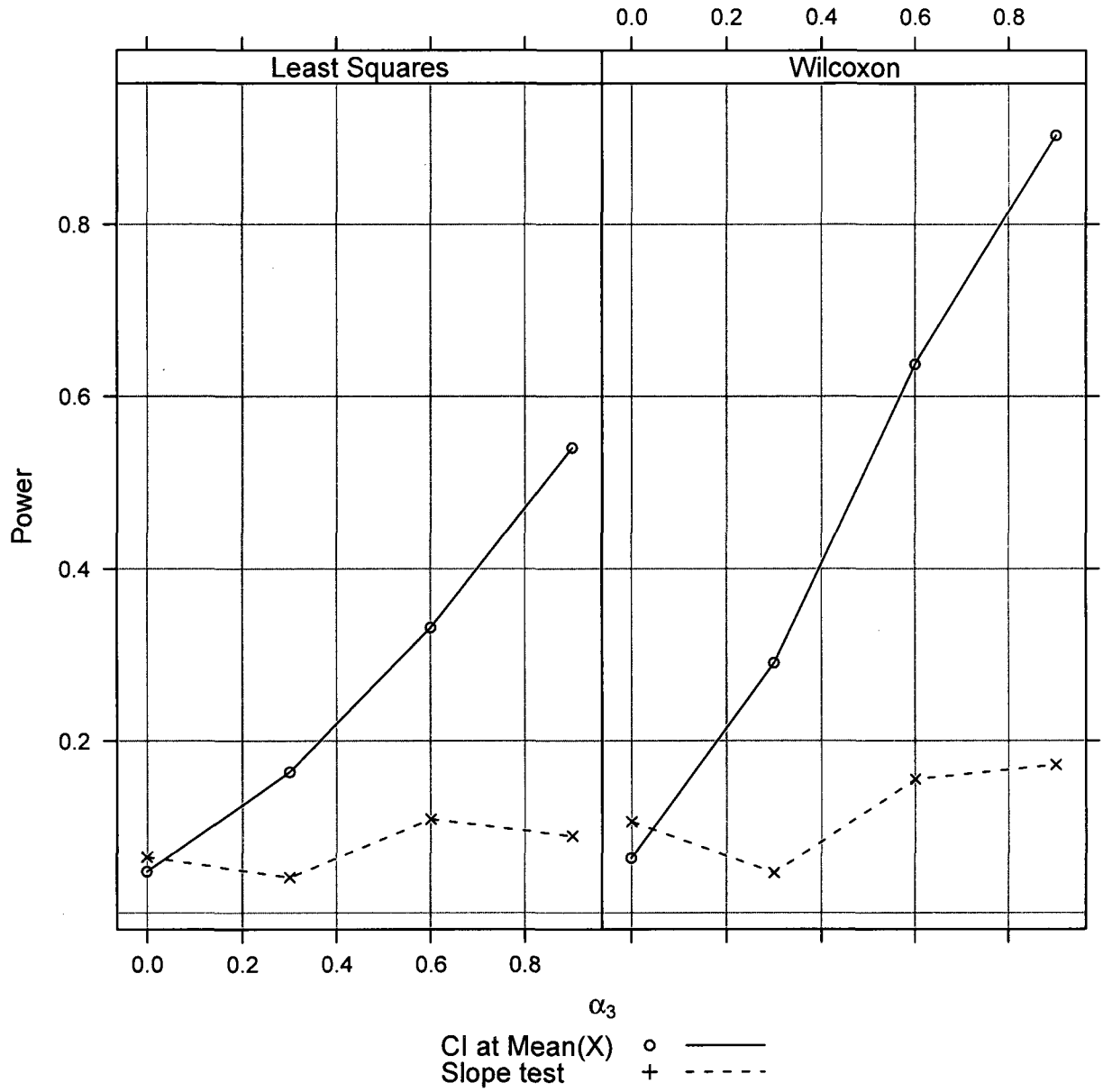

Figure 20: Y is Cauchy, $n = 20$

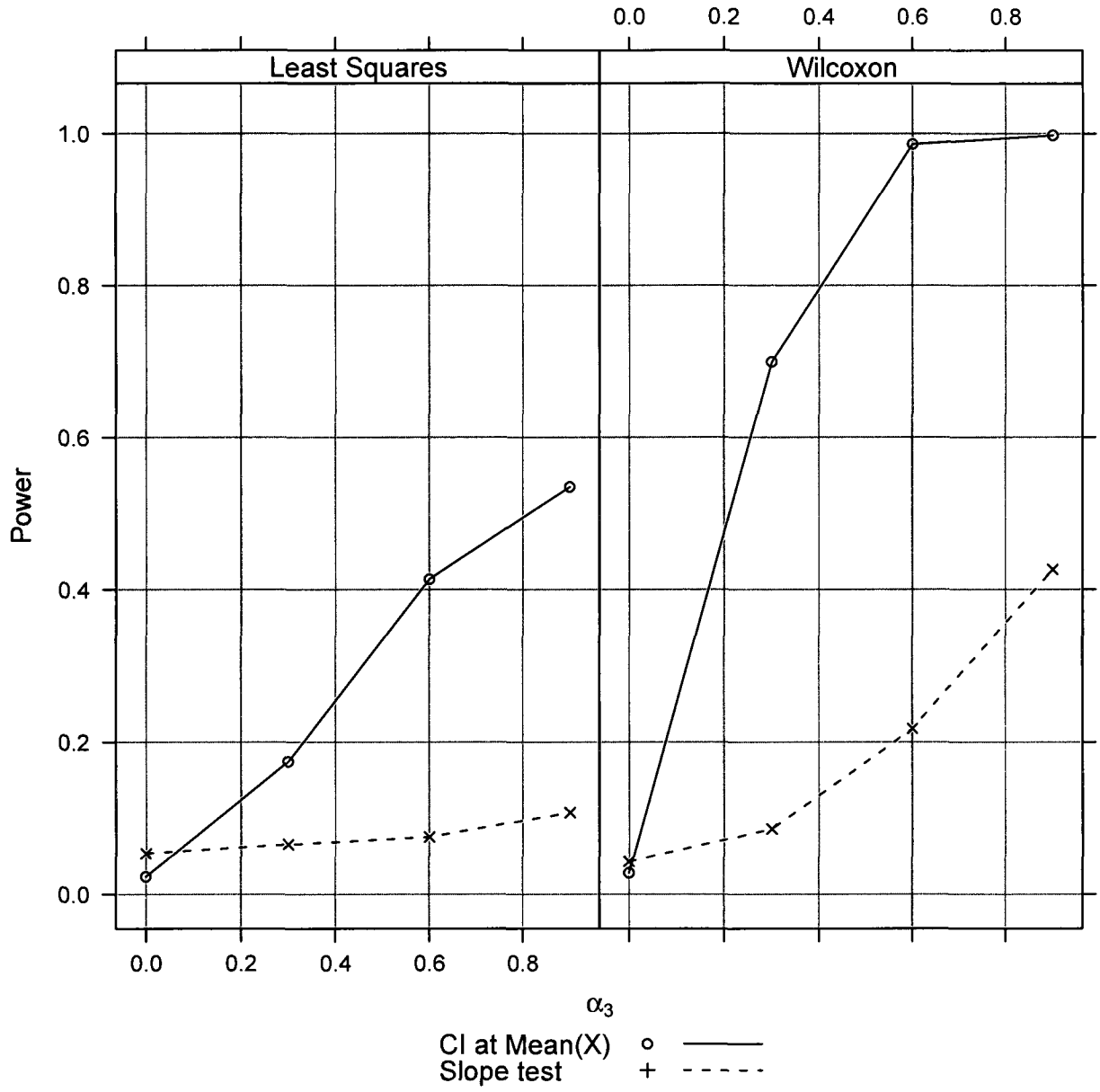

Figure 21: Y is Cauchy, $n = 40$

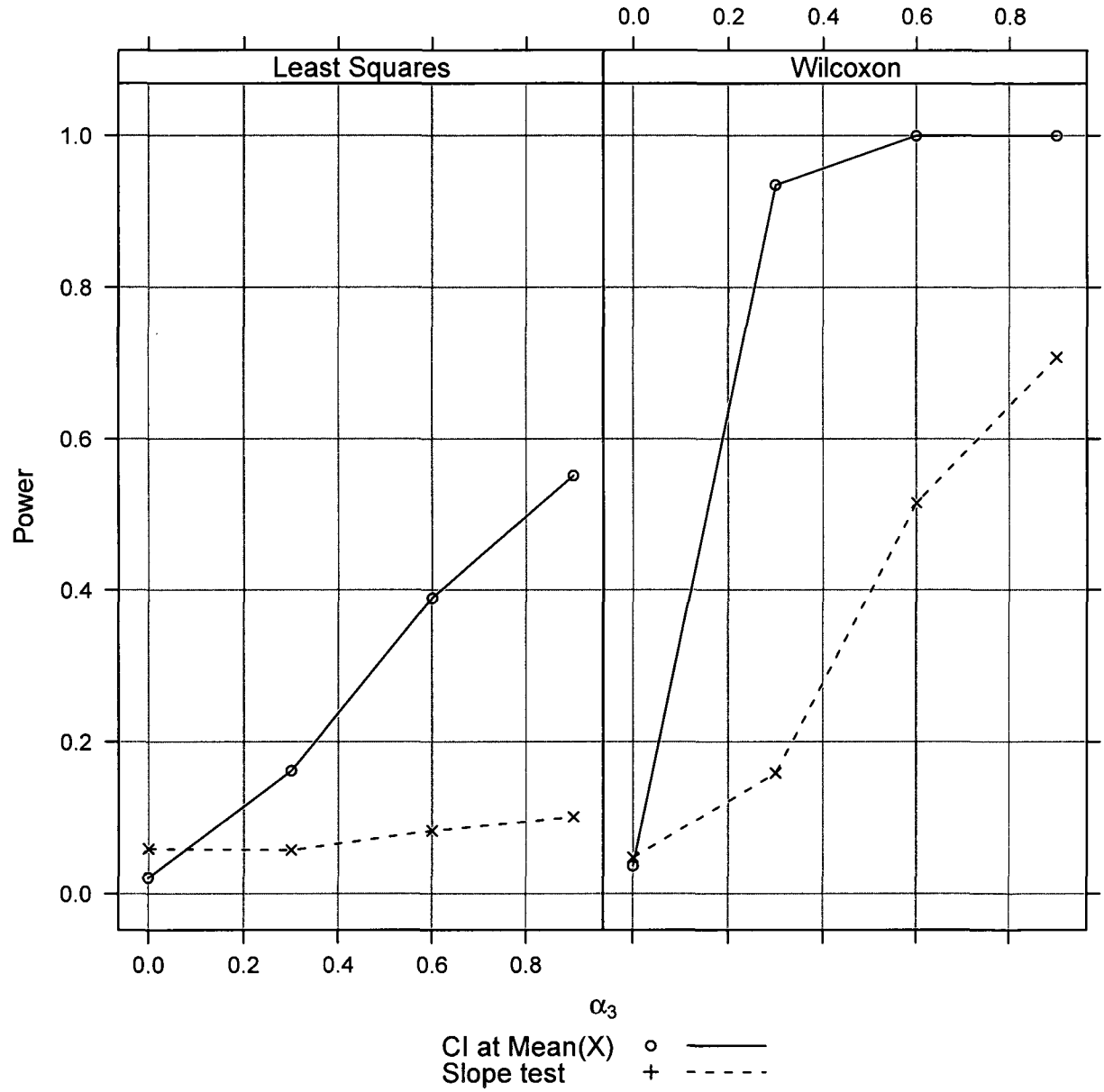

Figure 22: Y is Cauchy, $n = 80$

For Cauchy distribution, the power of robust procedure is much higher than that of the LS procedure when α_3 is higher as well as the case of pick-a-point (Table 32, 33, and 34). The plots of the power for homogeneity at 5% level of significance and that for pick-a-point at grand mean is shown in Figure 20 - 22.

Getting back to the two questions above, the results based on normal and Laplace distributions show that the power of Condition A, which is the pick-a-point, is equal to or slightly higher than the power of Condition B except the power from LS procedure at $\alpha_3 = 0.3$ and $n = 40$ in case of normal distribution, and at $\alpha_3 = 0.3$ and $n = 80$ in case of Laplace distribution. For Cauchy distribution, the power of Condition A of both LS and R procedures is slightly higher than the power of Condition B except the power of WWIL at $\alpha_3 = 0.3$ and $n = 40$.

Second question can be answered based on the simulation results from Condition A, Condition B, and Method III, the powers of the traditional and rank-based analysis is different especially when the distribution is the heavy-tailed distribution.

CHAPTER III

ADJUSTED MEANS

Consider a data set with k groups in which group i has sample size n_i, for $i = 1, 2, \ldots, k$. Note that the total sample size is $n = \sum_{i=1}^{k} n_i$. Let Y_{ij} denotes the jth response from the ith group and let X_{ij} denotes the value of the jth covariate of the ith group. We assume that the slopes are homogenous in this chapter. The centered design is

$$Y_{ij} = \alpha_i + \beta(X_{ij} - \overline{X}_{..}) + e_{ij}, \quad j = 1, 2, \ldots, n_i, \ i = 1, 2, \ldots, k,$$

where α_i is the effect of the ith treatment and β is the slope. The mean of the ith treatment is

$$\overline{Y}_{i.} = \alpha_i + \beta(\overline{X}_{i.} - \overline{X}_{..}) + e_{ij}$$

Now suppose $\widehat{\beta}$ is an estimate of β. Then an estimator of α_i is

$$\widehat{\alpha}_i = \overline{Y}_{i.} - \widehat{\beta}(\overline{X}_{i.} - \overline{X}_{..}) \tag{3.1}$$

The second term on the right-hand side is the adjustment introduced by the covariance analysis (Snedecor and Cochran, 1980, p.367). If we take the expected value of $\widehat{\alpha}_i$ we get

$$E[\widehat{\alpha}_i] = \alpha_i + \beta(\overline{X}_{i.} - \overline{X}_{..}) - \beta(\overline{X}_{i.} - \overline{X}_{..}) = \alpha_i.$$

Hence, the covariance adjustment removes all the bias if we have random sample and the regression of Y on X is linear with the same slope in each population; see also Snedecor and Cochran, 1980, p.380. Note in (3.1), the mean of the ith group is adjusted

by the covariate term. Hence, $\widehat{\alpha}_i$ is called the adjusted mean of the ith group. We denote it as

$$\overline{Y}_{i,adj} = \overline{Y}_i - \widehat{\beta}(\overline{X}_{i.} - \overline{X}_{..}) \tag{3.2}$$

where $\overline{Y}_{i,adj}$ denotes the adjusted mean of the ith group, $\overline{Y}_i$ denotes the unadjusted mean, $\widehat{\beta}$ denotes the estimator of the slope parameter, $\overline{X}_{i.}$ denotes the covariate mean of the ith group, and $\overline{X}_{..}$ denotes the grand covariate mean. The adjusted mean is employed to reduce the bias by adjusting the mean of treatments to the point that we would expect the mean response variable to occur when the covariate means of all groups are the same as the grand covariate mean (Huitema, 1980, p.15).

These estimators motivate some simple naive robust analogs. For example, consider the ith (naive) adjusted median given by

$$\widetilde{Y}_{i,adj} = \widetilde{Y}_i - \widehat{\beta}_R(\overline{X}_{i.} - \overline{X}_{..})$$

where $\widetilde{Y}_i$ denotes the ith groups median and $\widehat{\beta}_R$ denotes a robust estimator of common slope. Another example is the adjusted Hodges-Lehmann (HL) estimate of location. The HL estimator of the ith group is

$$\widetilde{Y}_i^+ = \text{med}_{1 \le k \le j \le n_i} \frac{Y_{ij} + Y_{ik}}{2} \tag{3.3}$$

Hence, the ith adjusted naive HL estimate is

$$Y^{+}_{i,adj} = \widetilde{Y}^{+}_{i} - \widehat{\beta}_R(\overline{X}_{i.} - \overline{X}_{..}) \tag{3.4}$$

where $\widehat{\beta}_R$ denotes a robust estimator of the common slope parameter. The properties of these naive robust adjusted "means" are not apparent because the Y_i's are not identically distributed. In particular, the standard errors for these simple robust analogs are difficult to obtain. For example, for the median analog, we need the asymptotic distribution of the sample median of non-identically distributed random variables. Other one-sample R estimators of center (and hence adjusted center estimators) are discussed below.

As we discuss next, the LS adjusted mean estimators (3.2) are easily obtained from the full model LS fit of model (3.6). From this point of view, standard errors for these estimators readily follow. Even though they are the same for LS, we call these adjusted means for the design. Our robust analog based on the design will differ from the naive ones given above. We then consider the robust analogues of adjusted means based on the design and establish the asymptotic theory for these robust estimators. Besides efficiency comparisons between the robust and LS estimators, the theory leads to the standard errors for these robust estimators.

3.1 LS Adjusted Means via the Design Matrix

The design matrix implementation of LS adjusted means is given on page 65 of Huitema (1980) without proof. To gain an understanding of this development design, we present a proof of it now. As we show later, this proof motivates the robust analog.

Consider the design matrix

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \alpha & \beta_1 & \beta_2 & \cdots & \beta_{k-1} & \beta_k \end{matrix} \\ \begin{bmatrix} 1 & 1_{n_1} & 0 & \cdots & 0 & x \\ 1 & 0 & 1_{n_2} & \cdots & 0 & x \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & \cdots & 1_{n_{k-1}} & x \\ 1 & 0 & 0 & 0 & 0 & x \end{bmatrix} \end{array}. \tag{3.5}$$

A second way of writing it is $\mathbf{X} = \begin{bmatrix} \mathbf{1}_n & \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_{k-1} & \mathbf{x} \end{bmatrix}$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones, $\mathbf{c}_i$ is the $n \times 1$ dummy vector for the ith group, and $\mathbf{x}$ is the vector of covariate. Write $\mathbf{X}$ as $\mathbf{X} = [\mathbf{1}_n \ \mathbf{x}]$. Let $\mathbf{Y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_k')'$, where $\mathbf{y}_i$ is the vector of responses for the ith group. Let $\mathbf{b} = (\alpha, \beta')'$ denote the vector of parameters. The LS estimates satisfy the normal equations given by

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y} \tag{3.6}$$

Equivalently, this expression can be written as

$$\begin{bmatrix} n & n_1 & n_2 & \cdots & n_{k-1} & \sum\sum x_{ij} \\ n_1 & n_1 & 0 & \cdots & 0 & \sum x_{1j} \\ n_2 & 0 & n_2 & \cdots & 0 & \sum x_{2j} \\ \vdots & & & & & \vdots \\ n_{k-1} & 0 & 0 & \cdots & n_{k-1} & \sum x_{k-1j} \\ \sum x_{ij} & \sum x_{1j} & \sum x_{2j} & \cdots & \sum x_{k-1j} & \sum\sum x_{ij}^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum\sum y_{ij} \\ \sum y_{1j} \\ \vdots \\ \sum y_{k-1j} \\ \sum\sum x_{ij}y_{ij} \end{bmatrix}$$

where $\sum x_{ij}$ denote the summation of the covariate for the ith group. This leads to

the system of equations

$$n\alpha + n_1\beta_1 + \cdots + n_{k-1}\beta_{k-1} + \beta_k \sum\sum x_{ij} = \sum\sum y_{ij} \tag{1}$$

$$n_1\alpha + n_1\beta_1 + \beta_k \sum x_{1j} = \sum y_{1j} \tag{2}$$

$$n_2\alpha + n_2\beta_2 + \beta_k \sum x_{2j} = \sum y_{2j} \tag{3}$$

$$\vdots \tag{:}$$

$$n_{k-1}\alpha + n_{k-1}\beta_{k-1} + \beta_k \sum x_{k-1j} = \sum y_{k-1j} \tag{k}$$

$$\alpha \sum\sum x_{ij} + \beta_1 \sum x_{1j} + \cdots + \beta_{k-1} \sum x_{k-1j} + \beta_k \sum\sum x_{ij}^2 = \sum\sum x_{ij}y_{ij} \tag{k+1}$$

Dividing both sides of equation (2) by n_1, we have

$$\alpha + \beta_1 + \beta_k \bar{x}_{1.} = \bar{y}_{1.}$$

Then subtract off $\beta_k(\bar{x}_{1.} - \bar{x}_{..})$ from both sides to get

$$\alpha + \beta_1 + \beta_k \bar{x}_{1.} - \beta_k(\bar{x}_{1.} - \bar{x}_{..}) = \bar{y}_{1.} - \beta_k(\bar{x}_{1.} - \bar{x}_{..}).$$

Therefore, the adjusted mean of the $1st$ group is

$$\alpha + \beta_1 + \beta_k \bar{x}_{..} = \bar{y}_1 - \beta_k(\bar{x}_{1.} - \bar{x}_{..}) = \bar{y}_{1,adj}. \tag{3.7}$$

For the $2nd$ group, divide both sides of equation (3) by n_2

$$\alpha + \beta_2 + \beta_k \bar{x}_{2.} = \bar{y}_{2.}$$

Then subtract off $\beta_k(\bar{x}_{2.} - \bar{x}_{..})$ from both sides.

$$\alpha + \beta_2 + \beta_k \bar{x}_{2.} - \beta_k(\bar{x}_{2.} - \bar{x}_{..}) = \bar{y}_{2.} - \beta_k(\bar{x}_{2.} - \bar{x}_{..})$$

The adjusted mean of the $2nd$ group is

$$\alpha + \beta_2 + \beta_k \bar{x}_{..} = \bar{y}_2 - \beta_k(\bar{x}_{2.} - \bar{x}_{..}) = \bar{y}_{2,adj}. \tag{3.8}$$

Continuing this was, we can show that i adjusted mean, for $i = 1, \cdots, k-1$ is

$$\alpha + \beta_i + \beta_k \bar{x}_{..} = \bar{y}_{k,adj}.$$

For the kth adjusted mean, we proceed as follows. Because $n = n_1 + n_2 + \cdots + n_k$, we can rewrite equation (1) as

$$\begin{aligned} (n_1 + \cdots + n_k)\alpha + n_1\beta_1 + \cdots + n_{k-1}\beta_{k-1} + \beta_k \sum\sum x_{ij} &= \sum\sum y_{ij} \\ n_1\alpha + \cdots + n_k\alpha + n_1\beta_1 + \cdots + n_{k-1}\beta_{k-1} + \beta_k \sum\sum x_{ij} &= \sum\sum y_{ij} \end{aligned} \tag{3.9}$$

Taking the difference of equation (3.9) from the sum of equation (2), equation (3), $\cdots$, and equation (k), we have

$$\begin{aligned} & n_1\alpha + \cdots + n_k\alpha + n_1\beta_1 + \cdots + n_{k-1}\beta_{k-1} + \beta_k \sum\sum x_{ij} \\ & -(n_1\alpha + n_1\beta_1 + \beta_k \sum x_{1j}) - (n_2\alpha + n_2\beta_1 + \beta_k \sum x_{2j}) \\ & \quad - \cdots - (n_{k-1}\alpha + n_{k-1}\beta_{k-1} + \beta_k \sum x_{k-1j}) &&= \sum\sum y_{ij} - \sum y_{1j} - \cdots - \sum y_{k-1j} \\ & n_k\alpha + \beta_k \sum\sum x_{ij} - \beta_k \sum x_{1j} - \cdots - \beta_k \sum x_{k-1j} &&= \sum\sum y_{ij} - \sum y_{1j} - \cdots - \sum y_{k-1j} \\ & n_k\alpha + \beta_k(\sum\sum x_{ij} - \sum x_{1j} - \cdots - \sum x_{k-1j}) &&= \sum\sum y_{ij} - \sum y_{1j} - \cdots - \sum y_{k-1j}. \end{aligned} \tag{3.10}$$

Because

$$\sum\sum x_{ij} - \sum x_{1j} - \cdots - \sum x_{k-1j} = \sum x_{kj},$$

and

$$\sum\sum y_{ij} - \sum y_{1j} - \cdots - \sum y_{k-1j} = \sum y_{kj},$$

Equation (3.10) becomes

$$n_k\alpha + \beta_k \sum x_{kj} = \sum y_{kj}. \tag{3.11}$$

Dividing both sides of equation (3.11) by n_k, we have

$$\alpha + \beta_k \bar{x}_{k.} = \bar{y}_{k.}$$

Then subtracting from both sides with $\beta_k(\bar{x}_{k.} - \bar{x}_{..})$, we have an expression for the kth adjusted mean; i.e.,

$$\alpha + \beta_k \bar{x}_{..} = \bar{y}_{k.} - \beta_k(\bar{x}_{k.} - \bar{x}_{..}) = \bar{y}_{k,adj}. \tag{3.12}$$

Because the LS estimates solve the normal equations, we rewrite equations (3.7), (3.8), and (3.12) with the estimates; i.e.,

$$\bar{y}_{i,adj} = \bar{y}_i - \hat{\beta}_k(\bar{x}_{i.} - \bar{x}_{..}) = \hat{\alpha} + \hat{\beta}_i + \hat{\beta}_k \bar{x}_{..}, \quad i = 1, 2, \ldots, k-1 \tag{3.13}$$

and

$$\bar{y}_{k,adj} = \bar{y}_k - \hat{\beta}_k(\bar{x}_{k.} - \bar{x}_{..}) = \hat{\alpha} + \hat{\beta}_k \bar{x}_{..} \tag{3.14}$$

Thus, we have expressed the adjusted means in terms of the LS estimator based on the design matrix $\mathbf{X}$ (3.5).

Notice from this formulation that the standard error of the LS adjusted mean is easily obtained. For $i = 1, 2, \ldots, k-1$, the ith adjusted mean is $\bar{y}_{i,adj} = a_i' \hat{b}_{LS}$ where the

jth component of the vector a_i is given by

$$a_{ij} = \begin{cases} 1 & j = 1 \\ 1 & j = i + 1 \\ \bar{x} & j = k + 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The kth adjusted mean is $\bar{y}_{k,adj} = a_k^{'} \hat{b}_{LS}$ where the components of a_k are

$$a_{kj} = \begin{cases} 1 & j = 1 \\ \bar{x} & j = k + 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The standard error of the ith adjusted mean is the square root of

$$V(\bar{y}_{i,adj}) = a_i^{'} \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1} a_i, \quad i = 1, 2, \ldots, k.$$

3.2 Preliminary Notation

Rewrite model (2.3) as

$$\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_1 \beta + \mathbf{e} \tag{3.15}$$

where $\mathbf{X}$ of (2.5) is $\mathbf{X} = [1 \ \ \mathbf{X}_1]$. As in Chapter 2, let $\varphi(u)$ be a given nondecreasing score function defined on the interval $(0, 1)$. Recall that $F(t)$ is the c.d.f of the random errors and $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$. Let $\widehat{\boldsymbol{\beta}}_\varphi$ denote the R estimator β; i.e.,

$$\widehat{\boldsymbol{\beta}}_\varphi = \text{Argmin} \, \| \mathbf{Y} - \mathbf{X}_1 \boldsymbol{\beta} \|_\varphi$$

A result that will prove useful is the asymptotic representation of $\widehat{\boldsymbol{\beta}}_\varphi$ which is given by

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_\varphi - \beta) = \tau_\varphi(n^{-1}\mathbf{X}'\mathbf{X})^{-1}n^{-1/2}\mathbf{X}'\varphi(\mathbf{F}(\mathbf{Y} - \mathbf{X}\beta)) + o_p(1); \tag{3.16}$$

see page 163 of Hettmansperger and McKean (1998).

The R estimator of the intercept that we discussed in Chapter 2 is the median of the residuals,

$$\widehat{\boldsymbol{\alpha}}_S = \text{med}\left\{\mathbf{Y} - \mathbf{X}'\widehat{\boldsymbol{\beta}}_\varphi\right\}.$$

It solves the equation

$$S_1(\mathbf{Y} - \mathbf{1}_n\alpha - \mathbf{X}\widehat{\boldsymbol{\beta}}_\varphi) = \sum_{i=1}^{n} \text{sgn}(Y_i - \alpha - x_i'\widehat{\boldsymbol{\beta}}_\varphi) = 0. \tag{3.17}$$

We next show the asymptotic representation of $\widehat{\boldsymbol{\alpha}}_S$. Without loss of generality, the true intercept and slopes are assumed as 0. According to Lemma 3.5.8 of Hettmansperger and McKean (1998, p.165), for all $S(a)$,

$$n^{-1/2}\left|S_1(\mathbf{Y} - an^{-1/2}\mathbf{1} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\varphi) - S_1(\mathbf{Y} - an^{-1/2}\mathbf{1})\right| \xrightarrow{P} 0.$$

Letting $a = 0$, $n^{-1/2}\left|S_1(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\varphi) - S_1(\mathbf{Y})\right| \xrightarrow{P} 0$, which implies that the asymptotic distribution of $n^{-1/2}S_1(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\varphi)$ and $n^{-1/2}S_1(\mathbf{Y})$ are the same.

Furthermore, Lemma 3.5.8 leads to the asymptotic linearity result for the equation (3.17),

$$n^{-1/2}S_1(\mathbf{Y} - an^{-1/2}\mathbf{1} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\varphi) = n^{-1/2}S_1(\mathbf{Y}) - a\tau_S^{-1} + o_p(1),$$

see Hettmansperger and McKean (1998).

The intercept solves $S_1(\mathbf{Y} - \mathbf{1}_n\alpha - \mathbf{X}\widehat{\boldsymbol{\beta}}_\varphi) = 0$ and $n^{1/2}\widehat{\boldsymbol{\alpha}}_S$ is bounded in probability,

hence, we have

$$0 = n^{-1/2} S_1(\mathbf{Y}) - n^{1/2} \widehat{\boldsymbol{\alpha}}_S \tau_S^{-1} + o_p(1).$$

This we can write as

$$0 = n^{-1/2} \sum_{i=1}^{n} \text{sgn}(Y_i) - n^{1/2} \widehat{\boldsymbol{\alpha}}_S \tau_S^{-1} + o_p(1).$$

This yield the asymptotic representation

$$\widehat{\boldsymbol{\alpha}}_S = n^{-1/2} \tau_S n^{-1/2} \sum_{i=1}^{n} \text{sgn}(Y_i) + o_p(1).$$

Thus, the asymptotic representation of the R estimate of the intercept with α as the true intercept is

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_S - \alpha) = \tau_S n^{-1/2} \sum_{i=1}^{n} \text{sgn}(Y_i - \alpha) + o_p(1).$$

The intercept can also be estimated based on signed-rank location process. To discuss this, first consider the simple location model

$$Z_i = \theta + e_i \tag{3.18}$$

where $e_1, \ldots, e_n$ are iid with pdf $f(x)$ and cdf $F(x)$. For theory, we need to assume also that f is symmetric about 0. Recall that the optimal rank regression scores are generated by the score function

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

But then by symmetry $\varphi_f(u)$ is odd about $\frac{1}{2}$; i.e.,

$$\varphi_f(u) = -\varphi_f(1-u).$$

Note that such scores satisfy: $\varphi(\frac{1}{2}) = 0$ and $\varphi(u) \geq 0$ for $u \geq \frac{1}{2}$ (Hettmansperger and McKean, 1998, p.101). The corresponding signed-rank scores are $a^+(i) = \varphi^+(\frac{i}{n+1})$, where $\varphi^+(u) = \varphi(\frac{u+1}{2})$. These scores are positive and decreasing. Consider the norm

$$\|\mathbf{v}\|_+ = \sum_{i=1}^{n} a^+(R\,|v_i|)\,|v_i|\,, \quad \mathbf{v} \in R^n$$

The signed-rank estimate of θ in the location model, (3.18), is

$$\widehat{\theta} = \text{Argmin}\,\|\mathbf{Z} - \theta\mathbf{1}\|_+\,,$$

where $\mathbf{Z}' = (Z_1, \ldots, Z_n)'$.

The gradient function of this norm is

$$S^+(\theta) = \sum_{i=1}^{n} a^+(R\,|Z_i - \theta|)\text{sgn}(Z_i - \theta).$$

Hence, the estimate also satisfies the equation

$$S^+(\widehat{\theta}) = 0.$$

If sign scores are used then $\widehat{\theta} = \text{med}Z_i$, while if Wilcoxon scores are used then

$$\widehat{\theta} = \text{med}_{i \leq j} \frac{Z_i + Z_j}{2}.$$

The above representation is under the assumption that the true location parameter is θ_o is 0. There is no loss in generality because $\widehat{\theta}$ is an equivariant location estimator, which is true for any norm-based estimator; see Chapter 1 of Hettmabsperger and McKean

(1998). In general, the representation is

$$\widehat{\theta}^{+} = \theta_o + \tau_\varphi n^{-1} - S^{+}(\mathbf{Z} - \theta_o \mathbf{1}) + o_p(1/\sqrt{n}).$$

Returning to the linear model, the estimate of the intercept is based on the residuals of the R fit using the score function φ, where $\varphi^{+}(u) = \varphi(\frac{u+1}{2})$. Denote the residuals $\widehat{e}$ = $Y - X_1\widehat{\beta}$ and $\widehat{\beta}$ is the R estimate of β in model (3.15). That is, using the gradient formulation, the estimate of the intercept is $\widehat{\alpha}$ such that

$$S^{+}(\widehat{\alpha}) = \sum_{i=1}^{n} a^{+}(R\,|e_i - \widehat{\alpha}|)\mathrm{sgn}(e_i - \widehat{\alpha}) = 0.$$

The function above is applied to the residuals for obtaining the estimate of the intercept. Therefore, we have

$$S^{+}(\alpha) = \sum_{i=1}^{n} a^{+}(R\,|e_i - \alpha|)\mathrm{sgn}(e_i - \alpha)$$

where $e_i = y_i - x_i'\widehat{\beta}_\varphi$. The intercept estimator is $\widehat{\alpha}_\varphi^{+}$ that solves $S^{+}(\alpha) = 0$. The estimate of HL estimator based on residuals, if we use Wilcoxon scores; i.e, $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$, while if we use sign scores then the estimate of intercept is the median of the residuals. As shown in Theorem A.2.11 of Hettmansperger and McKean (1998, p.410), the representation for the intercept is similar as $\widehat{\theta}^{+}$; i.e.,

$$\widehat{\alpha} = \alpha_o + \tau_\varphi n^{-1} S^{+}(e) + o_p(1/\sqrt{n}),$$

$$\widehat{\alpha}_\varphi^{+} = \alpha_o + \tau_\varphi n^{-1} \sum_{i=1}^{n} \varphi[F(e_i)] + o_p(1/\sqrt{n}), \tag{3.19}$$

where $e_i = y_i - x_i'\beta_o$ are the true errors.

3.3 Adjusted Means via Signed-Rank Estimate of the Intercept

Consider the analysis of covariance model (2.2). Let X be the design matrix (3.5). Recall LS development is based on the normal equations (3.6)

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}.$$

Write $\mathbf{X}$ as $\mathbf{X} = [\mathbf{1}\ \mathbf{X}_1]$. Then the right-hand side is

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \mathbf{1}' \\ \mathbf{X}_1' \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{X}_1'\mathbf{Y} \end{bmatrix}.$$

Also left-hand side is

$$\begin{aligned} \mathbf{X}'\mathbf{X}\mathbf{b} &= \begin{bmatrix} n & \mathbf{1}'\mathbf{X}_1 \\ \mathbf{X}_1'\mathbf{1} & \mathbf{X}_1' \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ &= \begin{bmatrix} n\alpha + \mathbf{1}'\mathbf{X}_1\beta \\ \mathbf{X}_1'\mathbf{1}\alpha + \mathbf{X}_1'\mathbf{X}_1\beta \end{bmatrix} \\ &= \begin{bmatrix} n\alpha + n\bar{\mathbf{x}}'\beta \\ n\bar{\mathbf{x}}\alpha + \mathbf{X}_1'\mathbf{X}_1\beta \end{bmatrix}. \end{aligned}$$

So putting these two results together, we have

$$\begin{bmatrix} n\alpha + n\bar{\mathbf{x}}'\beta \\ n\bar{\mathbf{x}}\alpha + \mathbf{X}_1'\mathbf{X}_1\beta \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{X}_1'\mathbf{Y} \end{bmatrix}. \tag{3.20}$$

Next, recall that the asymptotic representation for the R estimators $n\widehat{\alpha}_\varphi^+$ (3.19) and $\widehat{\beta}_\varphi$ (3.16), namely

$$n\widehat{\alpha}_\varphi^+ \doteq \mathbf{1}'\tau_\varphi\varphi[F(\mathbf{e})] \tag{3.21}$$

$$\mathbf{X}_c'\mathbf{X}_c\widehat{\boldsymbol{\beta}}_\varphi \doteq \mathbf{X}_c'\tau_\varphi\varphi[F(\mathbf{e})] \tag{3.22}$$

However with the design matrix $\mathbf{X} = [\mathbf{1}\ \mathbf{X}_1]$, we want the uncentered intercept; i.e., $\widehat{\boldsymbol{\alpha}}_\varphi = \widehat{\alpha}_\varphi^+ - \bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_\varphi$. Therefore, the asymptotic representation of the robust estimates of the regression coefficients is

$$n\widehat{\alpha}_\varphi^+ + n\bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_\varphi = \mathbf{1}'\tau_\varphi\varphi[F(\mathbf{e})] \tag{3.23}$$

$$\mathbf{X}_c'\mathbf{X}_c\widehat{\boldsymbol{\beta}}_\varphi = \mathbf{X}_c'\tau_\varphi\varphi[F(\mathbf{e})]. \tag{3.24}$$

Recall that $\mathbf{X}_c = \mathbf{X}_1 - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}_1 = \mathbf{X}_1 - \mathbf{1}\bar{\mathbf{x}}'$. Hence the right side of (3.24) is

$$\mathbf{X}_c'\tau_\varphi\varphi[F(\mathbf{e})] = \mathbf{X}_1'\tau_\varphi\varphi[F(\mathbf{e})] - \bar{\mathbf{x}}\tau_\varphi\mathbf{1}'\varphi[F(\mathbf{e})],$$

While the left side of (3.24) is

$$\begin{aligned}\mathbf{X}_c'\mathbf{X}_c\widehat{\boldsymbol{\beta}}_\varphi &= (\mathbf{X}_1' - \bar{\mathbf{x}}\mathbf{1}')(\mathbf{X}_1 - \mathbf{1}\bar{\mathbf{x}}')\widehat{\boldsymbol{\beta}}_\varphi \\ &= (\mathbf{X}_1'\mathbf{X}_1 - n\bar{\mathbf{x}}\bar{\mathbf{x}}')\widehat{\boldsymbol{\beta}}_\varphi \\ &= \mathbf{X}_1'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_\varphi - n\bar{\mathbf{x}}\bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_\varphi.\end{aligned}$$

So (3.24) is equivalent to

$$\mathbf{X}_1'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_\varphi - n\bar{\mathbf{x}}\bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_\varphi = \mathbf{X}_1'\tau_\varphi\varphi[F(\mathbf{e})] - \bar{\mathbf{x}}\tau_\varphi\mathbf{1}'\varphi[F(\mathbf{e})].$$

But by (3.23)

$$\bar{\mathbf{x}}\tau_\varphi\mathbf{1}'\varphi[F(\mathbf{e})] = n\bar{\mathbf{x}}\widehat{\boldsymbol{\alpha}}_\varphi + n\bar{\mathbf{x}}\bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_\varphi.$$

So (3.24) is equivalent to

$$n\bar{\mathbf{x}}\widehat{\boldsymbol{\alpha}}_{\varphi} + \mathbf{X}_1'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_{\varphi} = \mathbf{X}_1'\tau_{\varphi}[F(\mathbf{e})].$$

Therefore, the R asymptotic representation is equivalent to

$$\begin{aligned} n\widehat{\boldsymbol{\alpha}}_{\varphi} + n\bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_{\varphi} &= \mathbf{1}'\tau_{\varphi}\varphi[F(\mathbf{e})] \\ n\bar{\mathbf{x}}\widehat{\boldsymbol{\alpha}}_{\varphi} + \mathbf{X}_1'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_{\varphi} &= \mathbf{X}_1'\tau_{\varphi}\varphi[F(\mathbf{e})]. \end{aligned}$$

By (3.22) we have

$$\mathbf{X}'\mathbf{X}\widehat{\mathbf{b}}_{\varphi} = \mathbf{X}'\tau_{\varphi}\varphi[F(\mathbf{e})] + o(1/\sqrt{n}) \tag{3.25}$$

where $\widehat{\mathbf{b}}_{\varphi} = \left[\widehat{\boldsymbol{\alpha}}_{\varphi} \;\; \widehat{\boldsymbol{\beta}}_{\varphi}\right]$. Note that by replacing $\mathbf{Y}$ in (3.6) with $\tau_{\varphi}\varphi[F(\mathbf{e})]$, we get equation (3.25).

Denote the errors for the ith group as $e_{ij}, j = 1, \ldots, n_i$. Therefore, the asymptotic representation for our adjusted means follows as in (3.13) by replacing $\mathbf{Y}$ with $\tau_{\varphi}\varphi[F(\mathbf{e})]$. For $i = 1, \ldots, k-1$, the asymptotic representation of the ith adjusted mean is

$$\widehat{\boldsymbol{\alpha}}_{\varphi} + \widehat{\boldsymbol{\beta}}_{\varphi,i} + \widehat{\boldsymbol{\beta}}_k\bar{x} = \tau_{\varphi}\frac{1}{n_i}\sum_{j=1}^{n_i}\varphi[F(e_{ij})] - \widehat{\boldsymbol{\beta}}_{\varphi,k}(\bar{x}_{i\cdot} - \bar{x}) + o(1/\sqrt{n}) \tag{3.26}$$

Likewise, the asymptotic representation for the kth adjusted mean is

$$\widehat{\boldsymbol{\alpha}}_{\varphi} + \widehat{\boldsymbol{\beta}}_{\varphi,k}\bar{x} = \tau_{\varphi}\frac{1}{n_k}\sum_{j=1}^{n_k}\varphi[F(e_{kj})] - \widehat{\boldsymbol{\beta}}_{\varphi,k}(\bar{x}_{k\cdot} - \bar{x}) + o(1/\sqrt{n}) \tag{3.27}$$

By (3.19), $\tau_{\varphi}\frac{1}{n_k}\sum_{j=1}^{n_k}\varphi[F(e_{kj})] + o(1/\sqrt{n})$ is the asymptotic representation for the

signed-rank estimate of location for the ith group. Hence (3.26) and (3.27) are indeed the asymptotic representations for the signed-rank adjusted mean for the ith group.

We can use this process and obtain the asymptotic representation for the LS adjusted mean. It is

$$\frac{1}{n}\sum_{j=1}^{n_k} e_{kj} - \widehat{\boldsymbol{\beta}}_{LS,k}(\bar{x}_k - \bar{x}) + o(1/\sqrt{n}).$$

3.3.1 Standard Error

It follows immediately that the standard error of the ith signed-rank adjusted mean is the square root of

$$V(\bar{y}_{\varphi_{i,adj}}) = a_i^{'}\hat{\tau}^2(\mathbf{X}^{'}\mathbf{X})^{-1}a_i, \quad i = 1, 2, \ldots, k,$$

where, for $i = 1, 2, \ldots, k-1$, the vector a_i is given by

$$a_{ij} = \begin{cases} 1 & j = 1 \\ 1 & j = i+1 \\ \bar{x} & j = k+1 \\ 0 & \text{elsewhere.} \end{cases}$$

The components of a_k are

$$a_{kj} = \begin{cases} 1 & j = 1 \\ \bar{x} & j = k+1 \\ 0 & \text{elsewhere.} \end{cases}$$

3.4 Example

A sample of 30 freshmen biology students are randomly selected (Huitema, 1980, p.38) to analyze the difference among three type of behavioral objectives, which consist of (1) General, (2) Specific, and (3) Specific with study time allocations. The sample sizes of the three groups are equal with the common value 10. The response (Y) is the biology achievement test scores, and the covariate (X) is the aptitude test scores.

Table 35: Behavioral Objectives Data

1		2		3	
X	Y	X	Y	X	Y
29	15	22	20	33	14
49	19	24	34	45	20
48	21	49	28	35	30
35	27	46	35	39	32
53	35	52	42	36	34
47	39	43	44	48	42
46	23	64	46	63	40
74	38	61	47	57	38
72	33	55	40	56	54
67	50	54	54	78	56

The adjusted means for each group are computed using: (1) the simple way (formula (3.2)), and (2) the design matrix way (formula (3.11) and (3.12)). In addition, the adjusted means based on R estimates are considered. For the simple way, we estimate: (1) adjusted means by LS (AMLS), (2) robust adjusted median naive (RAMMN), and (3) robust adjusted Hodges-Lehmann naive (RAHL). Note that the standard errors for these simple robust analogs, which are RAMMN and RAHL, are difficult to obtain. In case of the design matrix way, we estimate: (1) adjusted means by LS (AMLS), (2) robust adjusted median design (RAMMD), and (3) robust adjusted signed-rank (RASR). Recall that the AMLS from both ways are the same. The results are in Table 36.

As can be seen, the adjusted means from both ways give exactly the same value.

Table 36: Estimate (Standard Error)

	Simple Way				Design Matrix Way		
	$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$		$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$
AMLS	28.48	40.33	36.19	AMLS	28.48	40.33	36.19
	(0.152)	(0.133)	(0.019)		(0.152)	(0.133)	(0.019)
RAMMN	28.57	42.25	36.18	RAMMD	27.14	39.35	36.16
	(NA)	(NA)	(NA)		(0.286)	(0.035)	(0.016)
RAHL	28.07	41.25	36.18	RASR	28.09	40.29	37.10
	(NA)	(NA)	(NA)		(0.191)	(0.129)	(0.110)

The adjustments based on R estimate are slightly different but close. The R estimate is different from LS primarily because the R estimate is more resistant to y-outliers. (McKean, Naranjo, and Sheather, 1999). To investigate the robustness of our adjusted means, we replace several the response values with outliers.

First, we replace the biology achievement test score of the first group with 100 ($Y_{11} = 100$). The outlier has a big impact on the AMLS of the first group ($\overline{Y}_{1,adj}$) (Table 37). When comparing between the simple and the design matrix ways, the RAMMN and RAHL in the simple way have both $\overline{Y}_{1,adj}$ and $\overline{Y}_{2,adj}$ higher than the RAMMD and RASR in the design matrix way; however, $\overline{Y}_{3,adj}$ in the simple way is lower and it is more closer to the original value.

Table 37: Estimate (Standard Error): 1 outlier

	Simple Way				Design Matrix Way		
	$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$		$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$
AMLS	37.89	39.53	36.08	AMLS	37.89	39.53	36.08
	(0.789)	(0.053)	(0.008)		(0.789)	(0.053)	(0.008)
RAMMN	32.73	42.11	36.16	RAMMD	30.37	40.68	36.89
	(NA)	(NA)	(NA)		(0.037)	(0.168)	(0.089)
RAHL	32.73	41.11	36.16	RASR	30.11	40.43	36.64
	(NA)	(NA)	(NA)		(0.011)	(0.143)	(0.064)

Next, another outlier is applied but in a different group. The biology achievement test score of the second group is replaced by the outlier; that is, $Y_{21} = 150$. Now, we then

have two outliers, one is in the first group, another is in the second group. The results in Table 38 show that the AMLS is very sensitive to the outliers. Both $\overline{Y}_{1,adj}$ and $\overline{Y}_{2,adj}$ of the RAMMN and RAHL in the simple way are higher than the RAMMD and RASR in the design matrix way.

Table 38: Estimate (Standard Error): 2 outliers

	Simple Way				**Design Matrix Way**		
	$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$		$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$
AMLS	39.41	51.20	35.89	AMLS	39.41	51.20	35.89
	(0.941)	(1.220)	(0.011)		(0.941)	(1.220)	(0.011)
RAMMN	33.17	43.73	36.10	RAMMD	30.73	41.85	35.67
	(NA)	(NA)	(NA)		(0.073)	(0.285)	(0.033)
RAHL	33.17	43.73	36.10	RASR	31.26	42.39	36.20
	(NA)	(NA)	(NA)		(0.126)	(0.339)	(0.020)

We then add another outlier to the biology achievement test score of the third group; that is, $Y_{31} = 180$. Therefore, all groups have one outlier. The outliers have an impact on the AMLS of both simple and design matrix ways. The RAMMD and RASR are less sensitive to the outliers (Table 39).

Table 39: Estimate (Standard Error): 3 outliers

	Simple Way				**Design Matrix Way**		
	$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$		$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$
AMLS	40.66	50.11	52.33	AMLS	40.66	50.11	52.33
	(1.066)	(1.111)	(1.633)		(1.066)	(1.111)	(1.633)
RAMMN	33.37	43.55	39.08	RAMMD	30.20	41.55	38.37
	(NA)	(NA)	(NA)		(0.020)	(0.255)	(0.237)
RAHL	33.37	43.55	40.08	RASR	31.37	42.73	39.55
	(NA)	(NA)	(NA)		(0.137)	(0.373)	(0.355)

3.5 Simulation Study

We next present the results of a simulation study for one covariate, and three groups with 10 observations in each group. Both covariate and response variables are generated

from: (1) standard normal distribution, and (2) contaminated normal distribution. The group means for the covariate are 4, 5, and 6, respectively; the standard deviation is 1 for all groups. Recall that we assume the slopes are homogenous. We run 10000 simulations. Let β = [0 4 6 1]. Recall the model

$$Y_{ij} = \alpha + \beta_i + X_{ij} + e_{ij},$$

where $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, and the adjusted means is

$$\overline{Y}_{i,adj} = \overline{Y}_{i.} - \widehat{\beta}(\overline{X}_{i.} - \overline{X}_{..})$$

Note that we use the same X for all 10000 simulations. The grand mean of X ($\overline{X}_{..}$) is 5.627, the means for each group are 3.076, 6.250, and 7.557, respectively.

Therefore, we have

$$\begin{array}{lrcccl}
Group1: & E(Y_{1n_1}) & = & 0+4+(1)(3.076)+0 & = & 7.076 \\
 & E(\overline{Y}_{1,adj}) & = & 7.076-(1)(3.076-5.627) & = & 9.627 \\
Group2: & E(Y_{2n_2}) & = & 0+6+(1)(6.250)+0 & = & 12.250 \\
 & E(\overline{Y}_{2,adj}) & = & 12.250-(1)(6.250-5.627) & = & 11.627 \\
Group3: & E(Y_{3n_3}) & = & 0+0+(1)(7.557)+0 & = & 7.557 \\
 & E(\overline{Y}_{1,adj}) & = & 7.557-(1)((7.557-5.627) & = & 5.627
\end{array}$$

We then have the true adjusted means for $1st$, $2nd$, and $3rd$ groups as 9.63, 11.63, and 5.63, respectively. The average of each adjusted estimate and the error are obtained as shown in Table 40.

All estimates from the design matrix way are the identical except for AMLS of the $3rd$ group. As a check for LS, for the simple way the AMLS estimates are the same as the

Table 40: Adjusted Mean Simulation: Standard Normal Distribution

	Simple Way				**Design Matrix Way**		
	$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$		$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$
AMLS	9.63	11.63	5.62	AMLS	9.63	11.63	5.62
	(0.12)	(0.10)	(0.11)		(0.12)	(0.10)	(0.11)
RAMMN	10.30	12.60	6.29	RAMMD	9.63	11.63	5.63
	(NA)	(NA)	(NA)		(0.14)	(0.12)	(0.13)
RAHL	9.80	11.59	6.02	RASR	9.63	11.63	5.63
	(NA)	(NA)	(NA)		(0.12)	(0.11)	(0.12)

estimates of the design matrix way. The RAMMN and RAHL are a little higher than the true mean except the RAHL of the $2nd$ group.

Next simulation is conducted in similar scenario except the response variable is generated from contaminated normal with $\epsilon = 0.3$, and $\sigma = 9$. The covariate we use here is the same as for standard normal distribution. Note that the true adjusted means of $1st$, $2nd$, and $3rd$ groups for this contaminated normal distribution are exactly the same as for standard normal distribution. The average of each adjusted estimate and the error are obtained (Table 41).

Table 41: Adjusted Mean Simulation: Contaminated Normal Distribution

	Simple Way				**Design Matrix Way**		
	$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$		$\overline{Y}_{1,adj}$	$\overline{Y}_{2,adj}$	$\overline{Y}_{3,adj}$
AMLS	9.63	11.64	5.63	AMLS	9.63	11.64	5.63
	(3.04)	(2.46)	(2.74)		(3.04)	(2.46)	(2.74)
RAMMN	10.07	12.08	6.17	RAMMD	9.63	11.64	5.64
	(NA)	(NA)	(NA)		(0.67)	(0.54)	(0.63)
RAHL	9.75	11.70	5.86	RASR	9.63	11.63	5.64
	(NA)	(NA)	(NA)		(0.70)	(0.57)	(0.64)

The AMLS estimates from both ways are identical. The errors of AMLS are much higher than those of RAMMD and RASR. The R estimates of the design matrix way are lower and are close to the true mean than those of the simple way.

CHAPTER IV

THE JOHNSON-NEYMAN TECHNIQUE

As we stated before, the important assumption of analysis of covariance is that the regression slopes are homogenous. If this assumption is not true, then the pick-a-point method of Chapter 2 can be used to investigate differences among treatment levels. In this chapter, we discuss an alternative test which is called the Johnson-Neyman technique. The purpose of the Johnson-Neyman procedure is to identify the values of X that are associated with significant group differences on Y (Huitema, 1980, p.271). Recall that the treatment effects in the analysis of covariance are assessed at the grand mean. However, the Johnson-Neyman technique determines a region in X-space where these differences are significant. The Johnson-Neyman technique was originally designed for the situation with two groups; however, it can be used to the case of multiple groups by comparing the pairs of groups.

In this study, we will consider the case of two groups and one covariate. The Johnson-Neyman technique based on LS is discussed. We then develop the robust analog to the Johnson-Neyman technique based on R estimates. We follow this development with a simulation study comparing the LS and robust Johnson-Neyman procedures.

4.1 Traditional and Robust Johnson-Neyman Procedures

Assume that the sample size of group i is n_i, $i = 1, 2, \ldots, k$. Let Y_{ij} denote the jth response from the ith group and let X_{ij} be the covariate. The model is written as

$$Y_{ij} = \alpha_i + \beta_i X_{ij} + e_{ij},$$

where α_i is the intercept and β_i is the slope parameters for the ith group. The errors e_{ij} are independent and identically distributed with pdf $f(t)$ and cdf $F(t)$. The conditional distribution of Y given X is

$$E(Y_{ij}|X_{ij}) = \alpha_i + \beta_i X_{ij}.$$

The difference at point X between both groups is

$$\begin{aligned}\Delta(X) &= E(Y_{2j}|X) - E(Y_{1j}|X) \\ &= (\alpha_2 + \beta_2 X) - (\alpha_1 + \beta_1 X) \\ &= (\alpha_2 - \alpha_1) + (\beta_2 - \beta_1)X.\end{aligned}$$

Given a fitting procedure, the estimator of the difference is

$$\widehat{\Delta}(X) = (\widehat{\alpha_2} - \widehat{\alpha_1}) + (\widehat{\beta_2} - \widehat{\beta_1})X. \tag{4.1}$$

First we outline the Johnson-Neyman procedure based on LS estimates. The Johnson-Neyman technique is used to obtain a point set or "region of significance" of values of the X variables for which one would reject, at a specified level α, the null hypothesis that the two groups have the same expected Y value (Potthoff (1964, 1983)). This "region of significance," which will be referred to as R, consists of the set of all points X such that

$$[\widehat{\Delta}(X)]^2 - t^2_{f,1-(1/2)\alpha} v(X) S_e^2 > 0,$$

where

$$
\begin{aligned}
X_{r\times 1} &= (X_1, X_2, \ldots, X_r)' \\
v(X) &= \sum_{i=1}^{2}[(1/n_i) + (X - \bar{X}_i)'C_i^{-1}(X - \bar{X}_i)] \\
S_e^2 &= \sum_{i=1}^{2}[Y_i'Y_i - (1/n_i)(Y_i'\mathbf{1})^2 - \widehat{\beta}_i{}'W_i] \\
C_{r\times r} &= X_iX_i' - (1/n_i)(X_i\mathbf{1})(X_i\mathbf{1})' \\
W_{r\times 1} &= X_iY_i - (1/n_i)(X_i\mathbf{1})(Y_i\mathbf{1})' \\
f &= \sum_{i=1}^{2}(n_i - r - 1).
\end{aligned}
$$

The region R allows one to reject the null hypothesis ($\Delta(X) = 0$). This implies that one can be at least $100(1-\alpha)\%$ confident in making a statement about the difference between the two groups for any specified individual point X in R. However, one cannot be $100(1-\alpha)\%$ confident in making the statements about the differences simultaneously for all points in R (Potthoff, 1964, 1983). For simultaneous inference, consider the region R' defined by the set of all points X such that

$$[\widehat{\Delta}(X)]^2 - (r+1)F_{r+1,f,1-\alpha}v(X)S_e^2 > 0.$$

With confidence coefficient $\geq 100(1-\alpha)\%$ one can state simultaneously for all points X in R' that the two groups differ (Potthoff, 1964, 1983). The region R' is smaller than R at a given α level; however, R' can be larger by choosing a larger α level for R' than that for R. In summary, a $100(1-\alpha)\%$ confidence interval for $\Delta(X)$ for any specified individual point X (whether inside or outside R) is given by

$$\widehat{\Delta}(X) \pm t_{f,1-(1/2)\alpha}[v(X)S_e^2]^{1/2},$$

while $100(1-\alpha)\%$ simultaneous confidence intervals for the functions $\Delta(X)$ for all possible points X in the r-dimensional X-space are given by

$$\widehat{\Delta}(X) \pm [(r+1)F_{r+1,f,1-\alpha}v(X)S_e^2]^{1/2},$$

(Potthoff, 1964, 1983). Note that for exact confidence, we need to assume that the errors have a normal distribution. Otherwise the confidences are approximate.

It is common practice to make a preliminary test of the hypothesis $\beta_1 = \beta_2$ before using the Johnson-Neyman technique: If this hypothesis is rejected by the test, then the Johnson-Neyman technique is applied, but if it is not rejected, then analysis of covariance is applied rather than the Johnson-Neyman technique (Potthoff, 1964, 1983).

Next we want to develop the robust analog to the Johnson-Neyman technique. This is easier to do if the non-incremental model is used. So consider the matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & X_1 & 0 \\ 0 & 1 & 0 & X_2 \end{bmatrix}.$$

We then center X

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & X_1 - \bar{X}_1 & 0 \\ 0 & 1 & 0 & X_2 - \bar{X}_2 \end{bmatrix}.$$

Therefore,

$$\mathbf{X'X} = \begin{bmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_2 & 0 & 0 \\ 0 & 0 & c_1 & 0 \\ 0 & 0 & 0 & c_2 \end{bmatrix}, \tag{4.2}$$

where $c_j = \sum_{i=1}^{n_j}(X_{ij} - \bar{X}_j)^2$ for $j = 1, 2$.

The inverse of matrix 4.2 is

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n_1} & 0 & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 & 0 \\ 0 & 0 & \frac{1}{c_1} & 0 \\ 0 & 0 & 0 & \frac{1}{c_2} \end{bmatrix}.$$

Let X_0 be a point of interest, then it easily follows that

$$v(X_0) = \sum_{i=1}^{2}[(1/n_i) + (X_0 - \bar{X}_i)'C_i^{-1}(X - \bar{X}_i)].$$

Let $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ denote the parameters for this design matrix X. Note that β_3 and β_4 for this design are the same as for the incremental design but the intercepts; β_1 and β_2, differ from α_1 and α_2 as discuss below.

Let $h' = (-1, 1, -(X_0 - \bar{X}_1), (X_0 - \bar{X}_2))$ and $\hat{\beta}' = (\beta_1, \beta_2, \beta_3, \beta_4)$, then the estimate of interest ($\widehat{\Delta}(X_0)$) is the linear combination $h'\hat{\beta}$,

$$\begin{aligned} h'\hat{\beta} &= \hat{\beta}_2 - \hat{\beta}_1 + (X_0 - \bar{X}_2)\hat{\beta}_4 - (X_0 - \bar{X}_1)\hat{\beta}_3 \\ &= (\hat{\beta}_2 - \bar{X}_2\hat{\beta}_4) - (\hat{\beta}_1 - \bar{X}_1\hat{\beta}_3) + (\hat{\beta}_4 - \hat{\beta}_3)X_0. \end{aligned} \tag{4.3}$$

Recall that the R estimate of β we discuss in Chapter 2 is given by

$$\widehat{\boldsymbol{\beta}}_{\varphi} = \text{Argmin} D_{\varphi}(\beta),$$

where $D_{\varphi}(\beta)$ utilizes the pseudo-norm: $\|w\|_{\varphi} = \sum_{i=1}^{n} a[R(w_i)]w_i$. A weighted Wilcoxon estimate corresponds to a minimum of the following function

$$D_{W\varphi}(\beta) = \sum_{1 \le i < j \le n} b_{ij} \left| w_i - w_j \right|,$$

where b_{ij} denotes a weight to be used in the $(i,j)th$ comparison (Terpstra and McKean, 2005).

The estimate of interest is $\widehat{\Delta}_{LS}(X_0) = h'\hat{\beta}_{LS}$ and for the R estimate $\widehat{\Delta}_R(X_0) = h'\hat{\beta}_R$. Comparing the estimator of the difference (4.1) with $h'\hat{\beta}$ above (4.3), we have

$$\begin{aligned} \hat{\alpha}_2 &= \hat{\beta}_2 - \overline{X}_2\hat{\beta}_4 \\ \hat{\alpha}_1 &= \hat{\beta}_1 - \overline{X}_1\hat{\beta}_3. \end{aligned}$$

Note that the LS intercept estimator is the mean of the LS residuals; therefore,

$$\begin{aligned} \hat{\alpha}_2 &= \hat{\beta}_2 - \bar{X}_2\hat{\beta}_4 = \text{mean}(Y_2 - \hat{\beta}_4 X_2) \\ \hat{\alpha}_1 &= \hat{\beta}_1 - \bar{X}_1\hat{\beta}_3 = \text{mean}(Y_1 - \hat{\beta}_3 X_1). \end{aligned}$$

For the Wilcoxon estimate, let $\hat{Y} = X\hat{\beta}_\varphi$. The intercept α can be estimated by a location estimate based on the residuals $\hat{e} = Y - \hat{Y}$ (Hettmansperger and McKean, 1998, p.147). Let $\varphi(u)$ is a nondecreasing function on $(0,1)$. The scores are generated by $a(i) = \varphi[i/(n+1)]$, $i = 1, 2, \ldots, n$, where $\varphi(u) = \sqrt{12}\left(u - \frac{1}{2}\right)$. Let $\varphi^+(u) = \varphi[(u+1)/2] = \sqrt{3}u$. Let $a^+(i) = \varphi^+[i/(n+1)]$. Consider the norm

$$\|v\|_W^+ = \sum a^+(R\left|v_i\right|)\left|v_i\right|, \quad i = 1, 2, \ldots, n,$$

(Hettmansperger and McKean, 1998, p.42).

Let $\hat{e}_\varphi = Y - \hat{Y}_\varphi$. The signed-rank procedure is applied to the residuals in order to obtain an estimate of the intercept. That is, the estimate of the intercept minimizes the

norm $\|\hat{e}_\varphi - \alpha\mathbf{1}\|_\varphi^+$ or equivalently solves $S(\hat{e}_\varphi - \alpha) = 0$ where

$$S_\varphi(\hat{e}_R - \alpha\mathbf{1}) = \sum a(R\,|\hat{e}_{Ri} - \alpha|)\text{sgn}(\hat{e}_{Ri} - \alpha\mathbf{1});$$

(Hettmansperger and McKean, 1998, p.169).

For Wilcoxon scores, the estimate is the median of the Walsh averages (Hettmansperger and McKean, 1998, p.169), which is called the HL estimator based on residuals. Let $\widehat{\Delta}_\varphi(X_0) = (\widehat{\alpha}_{\varphi_2} - \widehat{\alpha}_{\varphi_1}) + (\widehat{\beta}_{\varphi_2} - \widehat{\beta}_{\varphi_1}X_0$. Hence, we have

$$\begin{aligned}\hat{\alpha}_{\varphi_2} &= \text{HL}(Y_2 - \hat{\beta}_4 X_2)\\ \hat{\alpha}_{\varphi_1} &= \text{HL}(Y_1 - \hat{\beta}_3 X_1),\end{aligned}$$

where $HL(Z) = \text{median}_{i\le j}(\frac{Z_i+Z_j}{2})$.

For LS, the point X_0 is significant if

$$[\widehat{\Delta}(X_0)]^2 - t^2_{(\alpha/2,n-4)}v(X_0)\hat{\sigma}^2 > 0, \tag{4.4}$$

and X_0 is simultaneous significant if

$$[\widehat{\Delta}(X_0)]^2 - 2F_{(\alpha,2,n-4)}v(X_0)\hat{\sigma}^2 > 0. \tag{4.5}$$

For R estimate, we replace $\hat{\sigma}^2$ in the residual part in the equation 4.4 and 4.5 with $\hat{\tau}^2$; therefore, the equations become

$$[\widehat{\Delta}_\varphi(X_0)]^2 - t^2_{(\alpha/2,n-4)}v(X_0)\hat{\tau}^2 > 0, \tag{4.6}$$

and

$$[\widehat{\Delta}_{\varphi}(X_0)]^2 - 2F_{(\alpha,2,n-4)}v(X_0)\hat{\tau}^2 > 0. \tag{4.7}$$

4.2 Simulation Study

Table 42 shows the aggression scores on behavioral checklist data (Huitema, 1980, p.272). The data are based on an experiment in which two methods of therapy are the treatments, and scores on a sociability scale are employed as the covariate. The response is the aggressiveness score on behavioral checklist. Based on a 5% level of significance, the treatment effects test indicates no difference; however, the homogeneity of slopes is significant. That is, the slopes are heterogeneous. The scatterplot shows heterogeneous regression slopes in which the regression line of Therapy 1 is higher than that of Therapy 2 at low values of covariate (X) and is lower at high values of X than that of Therapy 2 (Figure 23).

Table 42: Therapy Data

Therapy 1		**Therapy 2**	
X	**Y**	**X**	**Y**
1	10	1	5
2	10	1.5	6
2	11	2.5	6
3	10	3.5	7
4	11	4.5	8
5	11	4.5	9
5	10	5	9
6	11	6	9
6	11.5	6	10.5
7	12	7	11
8	12	7	12.5
8	11	7.5	12.5
9	11	8	14
10	12.5	9	14.5
11	12	10	16

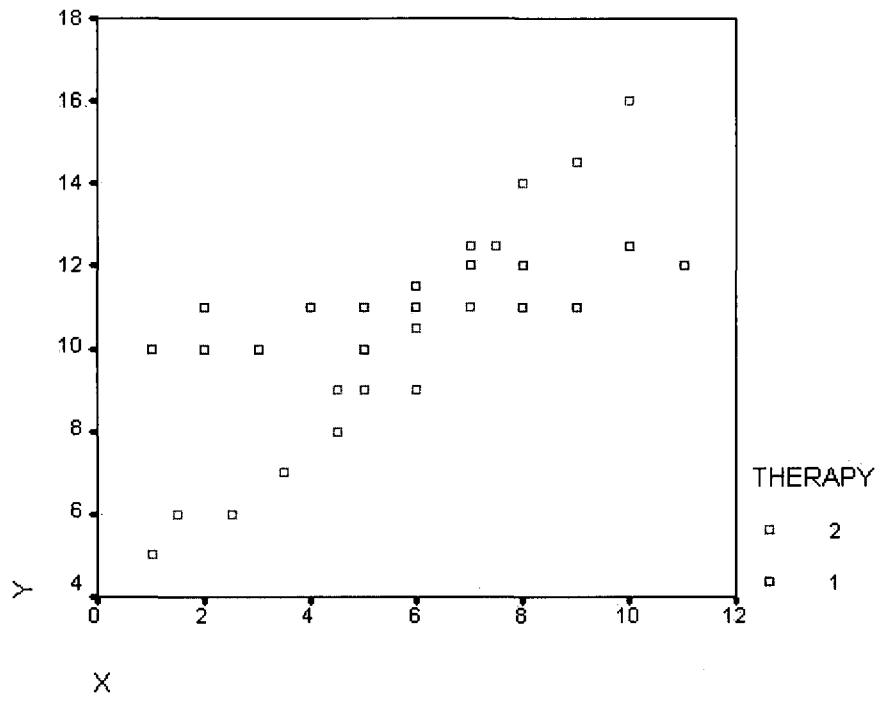

Figure 23: Plot of the behavioral checklist data

We do 10000 simulations for the response variable under three different distributions of error, which are (1) Standard normal distribution, (2) Normal distribution at $\mu = 0$, $\sigma = 30$, and (3) Contaminated normal distribution with $\epsilon = 0.3$ and $\sigma = 9$, as following step:

Response Variable Simulation Steps

1. Obtain $\hat{e}$ using the original X and Y
2. Compute $\hat{Y} = Y - \hat{e}$
3. Simulate 10000 of $\hat{e}$ from
   - **3.1** Standard normal distribution
   - **3.2** Normal distribution with $\mu = 0$ and $\sigma = 30$

3.3 Contaminated normal distribution with $\epsilon = 0.3$ and $\sigma = 9$

4. Obtain Y for each distribution by

4.1 $Y = \hat{Y} + \hat{e}_{standard\ normal}$

4.2 $Y = \hat{Y} + \hat{e}_{normal}$

4.3 $Y = \hat{Y} + \hat{e}_{contaminated\ normal}$

Note that the covariate (X) is the same for all 10000 simulations. Let a_1, a_2, b_1, and b_2 be the intercept of Therapy 1, intercept of Therapy 2, slope of Therapy 1, and slope of Therapy 2, respectively. Let X_0 be the point that the regression lines of the groups cross. Then we have

$$\begin{aligned} a_1 + b_1 X_0 &= a_2 + b_2 X_0 \\ X_0 &= \frac{a_2 - a_1}{b_1 - b_2} \\ &= 6.516461 \end{aligned}$$

At the crossing point, the power of the region of significance (4.2) and simultaneous region of significance (4.3), called JNLS method, is compared with that of the robust region of significance (4.4) and robust simultaneous region of significance (4.5), called JNWL method. The power between JNLS and JNWL of the other points: (1) Minimum of X $(Min(X))$, (2) $1st$ quartile of X (Q_1), (3) Point between $1st$ quartile and X_0 $(Q_{1.5})$, (4) Point at $\frac{Q_1}{4} + \frac{3X_0}{4}$ $(Qnew1)$, and (5) Point at $\frac{Q_1}{8} + \frac{7X_0}{8}$ $(Qnew2)$, is also investigated. Note that we use 5% level of significance for all cases here.

The power of the simultaneous region of significance is lower than that of the region of significance except at $Min(X)$ for both JNLS and JNWL methods, and at Q_1 for JNLS method only. At X_0, the power of the region of significance based on JNLS and JNWL is close to 5%, while the power of the simultaneous region of significance is about 1%-2%.

Table 43: Standard Normal Distribution

Method	Type	Power					
		X_0	$Min(X)$	Q_1	$Q_{1.5}$	$Qnew1$	$Qnew2$
JNLS	Region	0.0492	1.0000	1.0000	0.9778	0.5365	0.1685
	Simultaneous	0.0155	1.0000	1.0000	0.9314	0.3332	0.0757
JNWL	Region	0.0392	1.0000	1.0000	0.9503	0.4480	0.1308
	Simultaneous	0.0125	1.0000	0.9994	0.8588	0.2557	0.0519

The power is lower for the points which are closer to X_0. The closer points to X_0 from nearest to and further from X_0 are $Qnew2$, $Qnew1$, $Q_{1.5}$, Q_1, and $Min(X)$. The power of $Qnew2$ is the lowest among the other points, and the power of $Min(X)$ is the highest for both regions and both methods. When comparing between JNLS and JNWL, both the region and the simultaneous region, the empirical power of JNLS are slightly higher than those of JNWL, except at $Min(X)$ and at Q_1 (Table 43).

Table 44: Normal Distribution with $\mu = 0$, $\sigma = 30$

Method	Type	Power					
		X_0	$Min(X)$	Q_1	$Q_{1.5}$	$Qnew1$	$Qnew2$
JNLS	Region	0.0488	0.0561	0.0535	0.0490	0.0473	0.0487
	Simultaneous	0.0139	0.0193	0.0182	0.0166	0.0150	0.0138
JNWL	Region	0.0361	0.0456	0.0432	0.0382	0.0367	0.0377
	Simultaneous	0.0093	0.0151	0.0137	0.0123	0.0105	0.0097

For a normal distribution with $\mu = 0$ and $\sigma = 30$, the power of the simultaneous region of significance is lower than that of the region of significance at all points for both JNLS and JNWL methods. The power of the region of significance at all points based on JNLS and JNWL is close to 5%, while the power of the simultaneous region of significance is about 1%-2%. When comparing between JNLS and JNWL, both region and simultaneous region of the empirical power of JNLS are slightly higher than those of JNWL at all points (Table 44).

In case of contaminated normal distribution, the power of the simultaneous region

Table 45: Contaminated Normal Distribution with $\epsilon = 0.3$, $\sigma = 9$

Method	Type	Power					
		X_0	$Min(X)$	Q_1	$Q_{1.5}$	$Qnew1$	$Qnew2$
JNLS	Region	0.0428	0.3860	0.2998	0.1501	0.0717	0.0473
	Simultaneous	0.0088	0.2322	0.1663	0.0650	0.0214	0.0119
JNWL	Region	0.0318	0.8235	0.7223	0.3672	0.1138	0.0464
	Simultaneous	0.0093	0.6908	0.5540	0.2039	0.0415	0.0139

of significance is lower than that of the region of significance at all points both JNLS and JNWL methods. At X_0, the power of the region of significance based on JNLS and JNWL is close to 5%, while the power of the region of significance is about 1%. The power is lower when the points are closer to the X_0. The closer points to X_0 from nearest to further are $Qnew2$, $Qnew1$, $Q_{1.5}$, Q_1, and $Min(X)$, respectively. The empirical power at $Qnew2$ is the lowest among the other points, and the power at $Min(X)$ is the highest for both regions and both methods. The further the point is from X_0, the higher power. When comparing between JNLS and JNWL, the region of significance of the empirical power of JNWL is higher than that of JNLS for all points except at X_0. The power is even much higher at the points that further from X_0. The power of the simultaneous region of JNWL is also much more higher than that of JNLS at the further points from X_0 (Table 45).

CHAPTER V

CONCLUSIONS

Robust procedures for the analysis of covariance (ANCOVA) model are provided in this study. For the traditional ANCOVA, our empirical results show that the validity of our robust analog is similar to the least squares procedures.

The main thrust of the study develops robust analogs of alternative methods to the ANCOVA. There are important methods associated with the ANCOVA. One such procedure developed is the pick-a-point method. Rank-based analogs are developed for the pick-a-point. The simulation study is conducted and compared between the traditional and R procedures under differences in distribution of response variable, slopes, and sample sizes at different pick-a-point. The simulations results of the treatment effects test and the 95% confidence interval for pick-a-point show that the power is higher when the absolute difference in slopes (α_3) is larger. When n = 20, the power of R estimate is close to that of the least square procedure but the power of R estimate is approximately 2 times higher at $\alpha_3 = 1.9$, at all level of significances. For $n = 40$, the power of R estimate is more than 2 times higher at $\alpha_3 = 0.6$ at all level of significances. In case of $n = 80$, R estimate has 2 times higher of power when $\alpha_3 = 0.3$. Comparing the power among three distributions, the simulation results indicate that the R procedure can handle the Cauchy distribution, which is the heavy-tailed distribution, better than the LS procedure.

An second important method associated with the ANCOVA is the set of adjusted means. Two ways (simple and design matrix) to obtain the adjusted means are illustrated. We also develop R analogues for adjusted means, which are the robust naive adjusted median, robust naive adjusted Hodges-Lehmann, robust adjusted median design, and robust

adjusted signed-rank. Our empirical studies show that the LS procedure is sensitive to outliers. The comparisons of different way of computing adjusted means indicate that the design matrix way is superior, and that it is the preferred method. Simulation results also show that the robust procedures are useful for longer-tailed distributions.

The third procedure used in the analysis of covariance that we are interested in is the Johnson-Neyman technique. The robust procedure for the region and simultaneous region of significance is developed and compared with the traditional procedure at different points of X. Simulation results show that the robust procedure is more powerful for the heavy-tailed error distributions. The power of the simultaneous region of significance is lower than that of the region of significance at all points in both JNLS and JNWL methods. When comparing between JNLS and JNWL, the region of significance of JNLS equals to that of JNWL in case of standard normal distribution at $Min(X)$ and Q_1, and is slightly higher than that of JNWL at the other points of standard normal and all points of normal distribution with $\mu = 0$ and $\sigma = 30$. Likewise, the simultaneous region of JNLS equals to JNWL in case of standard normal distribution at $Min(X)$, and is slightly higher than JNWL at the other points of standard normal and all points of normal with $\mu = 0$ and $\sigma = 30$. However, for contaminated normal distribution, the region of significance of JNWL has higher power than JNLS except at X_0 and $Qnew2$. The simultaneous region of JNWL is higher than that of JNLS for all points of contaminated normal distribution.

REFERENCES

D'Alonzo, K.T. (2004). "The Johnson-Neyman Procedure as an Alternative to ANCOVA." Western Journal of Nursing Research, 26(7), 804-812.

Hettmansperger, T.P., and McKean, J.W. (1998). Robust Nonparametric Statistical Methods. Great Britain: Arnold.

Hettmansperger, T.P., and McKean, J.W. (1977). "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models." Technometrics, 19(3), 275-284.

Hettmansperger, T.P., and McKean, J.W. (2000). "Robust Nonparametric Methods." Journal of the American Statistical Association, 95(452), 1308-1312.

Hettmansperger, T.P., McKean, J.W., and Sheather, S.J. (2003). "Testing Symmetry of the Errors of a Linear Model." Mathematical Statistics and Applications: Festschrift for Constance van Eeden, Moore, M., Froda, S&Leger, C. (ed.), Institute of Mathematical Statistics, 99-112.

Hogg, R.V., McKean, J.W., and Craig, A.T. (2005). Introduction to Mathematical Statistics. Pearson Education, Inc., sixth edition.

Huitema, B.E. (1980). The Analysis of Covariance and Alternatives. New York: Wiley.

Johnson, P.O., and Neyman, J. (1936). "Tests of Certain Linear Hypothesis and their Application to Some Educational Problems." Statistical Research Memoirs, 1, 57-93.

Keselman, H.J., Huberty, C.J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donahue, B., Kowalchuk, R.K., Lowman, L.L., Petoskey, M.D., and Keselman, J.C. (1998). "Statistical Practices of Educational Researchers: An Analysis of Their ANOVA, MANOVA, and ANCOVA Analyses." Review of Educational Research, 68(3), 350-386.

Kloke, J.D., and McKean, J.W. (2010). "R-fit: An R Package for Rank-Based Analyses of Linear Models under General Scores", In preparation.

Koul, H.L., Sievers, G.L., and McKean, J.W. (1987). "An Estimator of the Scale Parameter for the Rank Analysis of Linear Models under General Score Functions." Scandinavian Journal of Statistics, 14, 131-141.

Kowalski, C.J., Schneiderman, E.D., and Willis, S.M. (1994). "ANCOVA for Nonparallel Slopes: the Johnson-Neyman Technique." International Journal of Bio-Medical Computing, 37, 273-286.

McKean, J.W. (2004). "Robust Analysis of Linear Models." Statistical Science, 19(4), 562-570.

McKean, J.W., Naranjo, J.D., and Sheather, S.J. (1999). "Diagnostics for Comparing Robust and Least Squares Fits." Nonparametric Statistics, 11, 161-188.

McKean, J.W., and Vidmar, T.J. (1994). "A Comparison of Two Rank-Based Methods for the Analysis of Linear Models." The American Statistician, 48(3), 220-229.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. (1996). Applied Linear Statistical Models. McGraw-Hill, Washington, D.C., fourth edition.

Potthoff, R.F. (1964). "On the Johnson-Neyman Technique and Some Extension Thereof." Psychometrika, 29, 241-256.

Potthoff, R.F. (1983). Johnson-Neyman Technique. In S. Kotz & N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences (Vol.4). New York: Wiley.

R Development Core Team (2005). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

Rogosa D. (1980). "Comparing Nonparallel Regression Lines." Psychological Bulletin, 88(2), 307-321.

Snedecor, G.W., and Cochran, W.G. (1980). Statistical Methods ($7th$ ed.). The Iowa State University Press.

Terpstra, J.T., and McKean, J.W. (2005). "Rank-Based Analyses of Linear Models Using R." Journal of Statistical Software, 14(7), http://www.jstatsoft.org/

White, C.R. (2003). "Allomatric Analysis beyond Heterogeneous Regression Slopes: Use of the Johnson-Neyman Technique in Comparative Biology." Physiological and Biochemical Zoology 76(1):135-140.