University of Southern Queensland

Increasing Power of M-test through Pre-testing

A Dissertation submitted by

Rossita Mohamad Yunus

B.Sc.(Hons), M.Sc.

For the award of

Doctor of Philosophy

2010

Abstract

The idea of improving the properties of estimators by pre-testing the uncertain non-sample prior information (NSPI) is adopted in the testing regime to achieve better power of the ultimate test. In this thesis, the studies on increasing power of the ultimate test through pre-testing the uncertain NSPI are carried out for four types of regression models, namely the simple regression model, the multivariate simple regression model, the parallelism model and the multiple linear regression model.

In this thesis, procedures are developed for

- testing the intercept of a simple regression model, when the NSPI on the slope can either be (i) unknown, (ii) certain or (iii) uncertain, or equivalently, when the slope is (i) completely unspecified, (ii) specified to a fixed value, or (iii) suspected to be a fixed value.

- testing the intercept vector of a multivariate simple regression model when the NSPI on the slope vector can either be (i) unknown, (ii) certain or (iii) uncertain, or equivalently, when the slope vector is (i) completely unspecified, (ii) specified to a fixed value, or (iii) suspected to be a fixed value.

- testing the intercepts of p (> 1) simple regression models when the NSPI on the slopes can either be (i) unknown, (ii) certain or (iii) uncertain, or equivalently, when the slopes are (i) completely unspecified, (ii) equal at a fixed value or (iii) suspected to be equal at a fixed value.

- testing a set of parameters of the multiple linear regression when the NSPI on the other set of parameters can either be (i) unknown, (ii) certain or (iii) uncertain, or equivalently, when the other set of parameters is (i) completely unspecified, (ii) zero or (iii) suspected to be zero.

Under the three different scenarios, the (i) unrestricted (UT), (ii) restricted (RT) and (iii) pre-test test (PTT) test functions are used to formulate the M-tests. The M-tests are derived using the score function in the M-estimation methodology. The sensitivity of the M-test to aberrant observations depends on the choice of the score function.

For each regression model, the following steps are carried out: (i) the test statistics for the UT, RT and PTT are proposed, (ii) the asymptotic distributions of the test statistics under the local alternative are derived, (iii) the asymptotic power functions of the tests are derived, (iv) the performance (size and power) of the UT, RT and PTT are compared analytically, (v) the performance of the UT, RT and PTT are compared, computationally using illustrative data of a two-sample case or data simulated using the Monte Carlo method.

Under a sequence of local alternative hypothesis when the sample size is large, the sampling distributions for the UT, RT and PT of the simple regression model follow a normal distribution. However, that of the PTT is a bivariate normal distribution. For the multivariate simple regression model, parallelism model and multiple linear regression model, the sampling distributions of the UT, RT and PT follow a univariate noncentral chi-square distribution under the alternative hypothesis when the sample size is large. However, that of the PTT is a bivariate noncentral chi-square distribution. For all regression models, there is a correlation between the UT and PT but there is no such correlation between the RT and PT. To evaluate the power function of the PTT, a package in R is used to compute the probability integral of the bivariate normal while a new R code is written to compute the probability integral of the bivariate noncentral chi-square distribution.

The robustness properties of the M-test are studied computationally on the simulated data for the simple regression model and the multivariate simple regression model. The power of the M-test using the Huber score function is better than in the least-square (LS) based test because the former is not significantly affected by slight departures from the model assumptions while the latter depends heavily on the normality assumptions. For all regression models, the PTT demonstrates a reasonable domination over the other two tests asymptotically when the suspected NSPI value is not too far away from that under the null hypothesis.

Certification of Dissertation

I certify that the ideas, experimental work, results, analyses, software and conclusions reported in this dissertation are entirely my own effort, except where otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, except where otherwise acknowledged.

....

Signature of Candidate Date

ENDORSEMENT

....

Signature of Supervisor Date

Acknowledgements

Gratitude to Almighty for the completion of this thesis.

I heartily thank my principal supervisor, Professor Shahjahan Khan, for proposing the research topic, clear direction on the project, expert guidance, intellectual challenges and debates, open access and very generous time, continuous support and encouragement throughout the course of my study in the University of Southern Queensland. I am grateful to him for organising the financial assistance and collaboration opportunities as well as emotional support, which has made him much more than merely an advisor to me. I am enormously benefitted from his outstanding scholarship, international professional standing and leadership in gaining scholarly exposition and expanding my academic views and networking.

I am grateful to my present and previous associate supervisors, Dr Stijn Dekeyser, Dr Ashley Plank and Dr Peter Dunn for their cordial support and assistance.

I would like to acknowledge the Government of Malaysia and University of Malaya, Malaysia for financially supporting my PhD study in the University of Southern Queensland. Very special thanks to the Department of Mathematics and Computing, and Australian Centre for Sustainable Catchments (ACSC), USQ for sponsoring my conference travels.

I wish to thank the Heads, Dr. Stijn Dekeyser, Dr. Richard Watson and Associate Professor Ron Addie, and friendly academic staff, particularly those in the area of Statistics, of the Department of Mathematics and Computing, USQ.

I really appreciate the teaching related opportunities and working experiences offered by the Department.

My gratitude also goes to the administrative staff in the Department of Mathematics and Computing, especially Mrs. Kris Lyon and Mrs. Helen Nkansah, the Office of Research and Higher Degrees, the Faculty of Science, the Library, the Division of the ICT services for rapid assistance to administrative and technical problems I encountered. My sincere thanks to Dr. Henk Huijser from the Learning and Teaching Support Unit, USQ for his efforts in proof reading the drafts, are greatly appreciated.

Special thanks to Heads, Professor Ong Seng Huat, Professor Noor Hasnah Moin and Professor Nor Aishah Hamzah, and staff of the Institute of Mathematical Sciences, University of Malaya for encouraging me to pursue my PhD degree overseas. Their support and advice through 'difficult time' kept my spirit up.

My sincere thanks goes to Miss Siti Amni Ismail who introduced me to the idea of *base load* which is use as a motivating example in the thesis and written articles, and Madam Siti Suzlin Supadi who solved many of my queries related to LaTeX. I also thank those who gave valuable comments on the draft of the first chapter and shared ideas at the beginning of my research studies.

Last but not least, my deepest sense of gratitude goes to my beloved family for their unconditional love and support.

List of Abbreviations

PTE	pre-test estimator
UE	unrestricted estimator
RE	restricted estimator
PTT	pre-test test
UT	unrestricted test
RT	restricted test
MLE	maximum likelihood estimator
LSE	least-square estimator
LRT	likelihood ratio test
cdf	cumulative distribution function
d.f.	degree of freedom

Contents

List of Figures

List of Tables

Chapter 1

Overview

1.1 Introduction

In the course of statistical inference, inferences about a population parameter are usually drawn from the sample data using statistical methods such as parameter estimation and hypothesis testing. Inferences about a population parameter could be improved by using information given by any trusted sources. Such information, which is usually provided by experts or experienced researchers, and is not related to any sample data, is known as the non-sample prior information (NSPI). We suspect that the inclusion of NSPI in addition to the sample data, in the statistical methods improves the quality of the estimator and the performance of the test. However, any NSPI value may be uncertain. If the NSPI value is uncertain or unsure, the information can be expressed in the form of a null hypothesis. An appropriate statistical test on this null hypothesis may be useful to eliminate the uncertainty on this suspected information. We also suspect that the preliminary testing (pre-testing) on the uncertain NSPI value used in the parameter estimation or hypothesis testing may improve the quality of the estimator and the performance of the statistical test (Saleh, 2006, Khan and Saleh, 2001, p.1-2).

The NSPI on any parameter may be classified as: (i) unknown if no NSPI is available, (ii) known if the exact value is found from NSPI, and (iii) uncertain if the suspected value is unsure. Under the three different scenarios, three types of estimators: (i) the unrestricted estimator (UE), (ii) the restricted estimator (RE) and (iii) the pre-test estimator (PTE) and three types of statistical tests: (i) the unrestricted test (UT), (ii) the restricted test (RT) and (iii) the pre-test test (PTT) are defined. The UE and UT use the sample data alone. The RE and RT do not use the sample data alone because the NSPI on the parameter is also included in the parameter estimation or hypothesis testing. The PTE and PTT use both the NSPI and the sample data. The PTE is a choice between the UE and the RE. Similarly, the PTT is a choice between the UT and the RT.

The choice depends on the outcome of the pre-testing on the uncertain NSPI value.

To see the idea of UE, RE and PTE in a regression model, we consider a simple regression model. This model has two unknown regression coefficients that are the intercept and the slope that relate a response variable (which value is to be predicted) to a predictor variable (which assumes fixed values). Obviously the estimation of the intercept parameter depends on the conditions on the slope parameter. Therefore, there are three situations considered for slope when the primary interest is to estimate the intercept parameter, namely (i) the slope is completely unspecified or there is no NSPI on the value of slope parameter, (ii) the slope is completely specified and fixed or the NSPI on the value of slope parameter is known with certainty, (iii) the slope is suspected at a fixed constant or the NSPI on the value of the slope parameter is uncertain or doubtful. For these three cases, we respectively define three estimators, namely, (i) UE, (ii) RE and (iii) PTE. The bias quadratic risk and mean square error are some of the statistical criteria to compare the performance of the UE, RE and PTE (Saleh, 2006, Khan et al., 2002).

In recent years, many researchers have contributed to the estimation of parameter(s) in the presence of uncertain NSPI. In spite of a plethora of work in the area of improved estimation using NSPI (Saleh, 2006), there is a very limited number of studies on the testing of parameters in the presence of uncertain NSPI. For testing the intercept of the simple regression model, we define three statistical tests, namely (i) UT, (ii) RT and (iii) PTT based on the knowledge of the NSPI on the slope. The slope can either be: (i) completely unspecified, (ii) specified at fixed value or (iii) uncertain. The statistical criteria that are used to compare the performance of the UT, RT and PTT are the size and power of the tests. A statistical test that has a smaller size is preferable because it guarantees the probability of a Type I error (probability of rejecting the null hypothesis when it is true) to be small. Furthermore, a test that has larger power than the other tests is preferable because it means the probability of Type II error (probability of fail to reject the null hypothesis when it is not true) is small. A test that minimizes both the probability of Type I and Type II errors is desirable though it is impossible to attain these two objectives simultaneously. So, normally an attempt is made to reduce the probability of a Type II error, keeping the probability of a Type I error fixed.

The UE, RE and PTE (as well as the UT, RT and PTT) have been studied for parametric cases (Bechhofer, 1951, Bozivich et al., 1956, Bancroft, 1964, Saleh, 2006) which rely heavily on the model assumptions. The parametric method is no longer an appropriate technique when the model assumptions are not met (Montgomery et al., 2001, p.386). Usually, it is assumed that the error term of a regression model is normally distributed, however, in reality, this may not be true (Goodall, 1983, p.350). The true underlying distribution could be any symmetric distribution such as the Student-t or Cauchy distribution, which is heavier in both tails than that of the Gaussian distribution (Khan and Dellaert, 2004, Rosenberger and Gasko, 1983). The true distribution could be a skewed distribution with one of the tails longer than the other. Also, the presence of outliers in the data can be the cause of poor performance of the estimators and tests. Outliers are observations that lie far away from the majority of the data and probably do not follow the assumed model (Becker and Gather, 1999). Assuming normal distribution on the error variables of the regression model, we find that the maximum likelihood (ML) estimators for the regression coefficients are identical to those of the least square (LS) estimators (Montgomery et al., 2001, p.52). So, both ML and LS estimators are sensitive to outliers or any departures from the assumed normal model. In the same manner, statistical tests based on the ML and LS estimators perform well only when the assumptions of the model hold. Since the parametric methods rely heavily on model assumptions, this have led to an increasing interest in other alternative methods in the literature such as the nonparametric and robust methods.

Nonparametric methods require relatively weak distributional assumptions for their validity, while robust methods make inferences that are little affected by a small number of outliers in the data or slight departures from the distributional assumptions (Garthwaite et al., 2002, p.185). When the model assumptions of the errors under a parametric model are not met, the nonparametric methods and tests are superior to those of the parametric ones. Many nonparametric procedures are based on rank-order statistics (Lindgren, 1993, Flaherty, 1999, p.445). Robust methods are as efficient as the parametric methods when the model assumptions of the parametric methods are met, but are more efficient than the parametric methods when there are departures from the model assumptions (Huber, 1981, p.5). There are three broad fundamental classes of robust estimation - R, M and L-estimation. To be short, the R-estimation is generally applied to some statistical rank test, the M-estimation is based on the so-called generalized maximum likelihood estimation theory and it is strongly related to the LS procedures and the L-estimation is conceived as a linear combination of order statistics (Wang and Wang, 2007, Jurečková and Sen, 1996, p.80). Several robust tests were derived using the robust estimation methodology found in literature. It is suspected that the robustness properties of a robust estimator should be inherited by the respective robust statistical test because both are derived from the same methodology.

Realizing the disadvantages of parametric methods, the UE, RE and PTE (as well as the UT, RT and PTT) for the nonparametric cases (Tamura, 1965, Saleh and Sen, 1982, 1983, Saleh, 2006) were proposed in literature. The rank statistics are mostly used for these nonparametric procedures. When observations are replaced with ranks, the more extreme observations are pulled in closer to the other observations. However, the nonparametric estimation methods and tests (based on ranked data) often preserve information about the order of the data but discard the actual values, thus overlook information that may have led to a better solution. The sign test, for example, uses only the sign of the deviation of each observation from the median. This thesis considers the statistical tests using the M-estimation methodology; the most popular estimation method among the robust estimation methods. To my knowledge, so far no UT, RT and PTT are proposed in literature that uses the M-estimation methodology. Several robust tests derived using the M-estimation methodology were proposed in the literature (Sen, 1982, Schrader and Hettmansperger, 1980, Shiraishi, 1990) but none of them were devoted to testing the parameters following a pre-testing on uncertain NSPI. We suspect some information may be lost when using tests based on the rank or order of the actual data. The statistical test derived using the M-estimation methodology does not suffer the same kind of loss of information as the rank test because the actual data are directly involved in the estimation and test procedures. In M-estimation, the squared of residual is replaced by an objective function of the residuals that could downweight the influence of some observations with large residuals; this makes the M-estimator robust against departures from model assumptions. The M-estimation methodology is chosen for its simplicity, it is well known, and it covers the ML method as a special case (Montgomery et al., 2001, p.405) and is less sensitive to departures from the assumed model or the presence of outliers.

The PTT was investigated for one and two sample problems (Tamura, 1965), for the simple regression model (Saleh and Sen, 1982), for the multivariate simple regression model (Saleh and Sen, 1983) using the nonparametric rank test and for the parallelism model (Lambert et al., 1985a) using the LS based test. However, there is no work investigating the PTT for the multiple linear regression model found in the literature. Apart from the simple regression model, the multivariate simple regression model and parallelism model, this thesis also covers the multiple linear regression model which has never been considered for the PTT. Also, the published articles were devoted to testing the value(s) of intercept(s) and slope(s) at zero(s). This thesis considers the problem of testing the intercept(s) and slope(s) for any arbitrary value, and hence the previous studies are special case studies of the current study. Obviously testing the null hypothesis about any value of the model parameters is more realistic than testing the significance of the parameters.

To carry out a statistical test, the distribution of the test statistic under the null hypothesis must be known (van der Vaart, 1998, p.1). The asymptotic distribution theory of the test statistics that are based on the score function in the M-estimation methodology developed by Jurečková (1977) and Jurečková and Sen (1996) is used in the thesis. Although the asymptotic results of Jurečková and Sen (1996) are used in deriving the distribution of the proposed test statistics in this thesis, these results are used in a different model in the context of testing after pre-testing on uncertain NSPI. Although there are robust tests derived using the other robust estimation methodologies such as the GM-estimation (Markatou and Hettmansperger, 1990, Heritier and Ronchetti, 1994, Gagliardini et al., 2005), the asymptotic distributions of these test statistics are complicated (Muller, 1998). As a result, we believe that it will be difficult to derive the asymptotic power function under the sequence of a local alternative hypothesis for these test statistics. The concept of contiguity probability measures is used to derive the asymptotic distributions under the alternative hypotheses.

The investigations into the comparisons of the UT, RT and PTT for the simple multivariate model by Saleh and Sen (1983) and the parallelism model by Lambert et al. (1985a) are limited to an analytical discussion only; the computational comparisons of the UT, RT and PTT are not provided in these papers. Perhaps, the computational comparison of the UT, RT and PTT could not be given due to the nonexistence tool to compute the power functions at that time. To compute the power of the PTT, the bivariate integral of the non-central chi-square distribution is required. However, the bivariate non-central chi-square distribution, available at the time their papers were published, were very complicated and not practical for computation. In this thesis, we refer to Yunus and Khan (2009) for the computation of the bivariate integral of the noncentral chi-square distribution. For a simple multivariate model and parallelism model, according to Saleh and Sen (1983) and Lambert et al. (1985a), the power of the PTT may be between those of the UT and RT. However, this statement is not clearly supported by arguments in their papers, probably due to the complicated form of the bivariate noncentral chi-square distribution that they used in their papers. No graphical or numerical comparisons of the power functions were provided in the previous studies.

1.2 Contribution of Thesis

The objectives of this thesis are

- to develop robust procedures for testing some parameters of a regression model when the remaining parameters are (i) unspecified (ii) fixed and specified (iii) uncertain values.
- to propose robust test statistics for the UT, RT and PTT using the M-estimation methodology,
- derive the sampling distributions of each test statistics under the sequence of a local alternative hypothesis,
- to derive the asymptotic power functions for each test statistics,
- to compare the performance (size and power) of the UT, RT and PTT analytically and computationally and
- to recommend an optimum test.

Main contributions of the thesis are as follows:

- Propose the UT, RT and PTT for the multiple linear regression (which model has never been considered in the pre-testing uncertain NSPI discipline) as well as the simple regression model, multivariate simple regression model and parallelism model.
- Use the statistical test derived in the M-estimation methodology for the UT, RT and PTT.
  - So far, only an LS based test and rank test are used to propose UT, RT and PTT for nonparametric cases in the literature.
  - So far, a robust statistical test derived in the M-estimation methodology has never been used to define the PTT.
- Improve the investigation into the comparison of power of the UT, RT and PTT through computational results. The investigation into the power function of the UT, RT and PTT that used the rank statistics in the published articles was limited due to the nonexistence tool to compute the power function. In the thesis, the performance of the tests is investigated through a simulated example using a program written in R and is supported by theoretical explanation.
- Consider testing the parameters at any arbitrary value which is more realistic and general.
- Search for an optimum test or a test that is a compromise between minimizing the size and maximizing the power.
- Propose the noncentral bivariate chi-square distribution by using the idea of compounding distribution. This distribution is required for the computation of the power of the PTT for regression models other than the simple regression model. A new program (R-code) for the computation of the bivariate integral of the bivariate noncentral chi-square distribution is written for this thesis.

1.3 Thesis Outlines

Chapter 2 presents a review of several topics related to the theory of pre-testing and robust procedures. Chapter 3 focuses on a simple regression model, Chapter 4 on a multivariate simple regression model, Chapter 5 on a parallelism model and Chapter 6 on a multiple linear regression model. In each Chapter, 3 through 6, along with preliminary notions, the method of M-estimation is presented, the M-tests for the UT, RT and PTT are defined and the asymptotic distribution theory involving the statistical tests are given. These results are then used to study the asymptotic size and power of the UT, RT and PTT. The analytical and graphical comparisons of the power of the UT, RT and PTT are provided. Discussions and concluding remarks are given in the final Section of each Chapter. The final Chapter contains the summary of the results, the final conclusions, the limitations of the study, and the recommendations for future work.

Chapter 2

Literature Review

2.1 Introduction

This Chapter covers previous work in the literature about statistical inference using NSPI. On the estimation side, this includes the UE, RE and PTE and in the testing context it deals with the UT, RT and PTT. Some technical concepts such as pre-testing, robust estimation, robust test, contiguity probability measure, bivariate noncentral chi-square distribution and outliers are also revisited. This Chapter also provides some statistical findings that are essential for some of the later derivations in the current study.

2.2 The UE, RE and PTE

The use of prior information in the procedure of estimating the parameter of interest of a statistical model usually improves the quality of the estimator. In many cases, the prior information is available as a suspected value of the parameter interest. However, such a prior value is likely to be uncertain. This has led to the suggestion of pre-testing the suspected value to remove the uncertainty. The outcome of the pre-test is then incorporated into the procedure of estimating the interested parameter or performing test on another parameter.

A variety of topics associated with the theory of pre-test is covered in Saleh (2006). This provides a thorough coverage of the related work published in the area of parameter estimation including the UE, RE and PTE. The parametric and non-parametric estimations are among the topics covered in the book.

An easy way to understand the concept of PTE is to consider a problem of estimating the population mean μ for one-sample data when it is *apriori* suspected that $\mu = \mu_0$ for any $\mu_0 \in \Re$. Khan (1998), Khan and Saleh (2001), Kabir and Khan (2009) are among the authors who consider a one-sample problem for the PTE. Let $X_1, X_2, \ldots, X_n$ be random variable of size n from $N(\mu, \sigma^2)$. Assume that NSPI on the value of μ is available. Then define the RE of μ as $\hat{\mu} = \mu_0$ and the UE of μ as $\tilde{\mu} = \bar{X}$. Based on the likelihood ratio test for testing $H_0 : \mu = \mu_0$ against $H_A : \mu \neq \mu_0$, the PTE of μ is defined by $\hat{\mu}^{PTE} = \tilde{\mu} I(|t| > t_{\alpha/2}) + \hat{\mu} I(|t| < t_{\alpha/2})$, where $I(A)$ is an indicator function of the set A and $t_{\alpha/2}$ is the critical value chosen for a two-sided α-level test based on the Student's t distribution with $n-1$ degrees of freedom. Obviously for a one-sample problem, the PTE is an UE if H_0 is rejected, otherwise it is an RE.

The idea of PTE was first introduced by Bancroft (1944) in his seminal work and later in Bancroft (1964, 1965). The PTE was studied for the classical problem of pooling means in a two-sample situation. Given two samples, the problem is to estimate the mean when it is *apriori* suspected that the two population means may be equal. For example, if we have two samples $\{(X_{i1}, \ldots, X_{in_i})|i = 1, 2\}$ from two normal distributions $N(\theta_1, \sigma^2)$ and $N(\theta_2, \sigma^2)$ respectively, then in order to estimate θ_1 when it is suspected that θ_1 may be equal to θ_2, one may use the first sample mean, $\bar{X}_1$ if the hypothesis $H_0 : \theta_1 = \theta_2$ is rejected or use $\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1+n_2}$ (pooled mean) if H_0 is accepted. This procedure is known as the PTE in the literature (c.f. Kim, 2003, p.2-3). The problem concerning the pooling of data was also discussed in Mosteller (1948). This work was motivated by the problem of testing the difference between two means after testing the equality of unknown variances found in Snedecor (1938). Note that this classical problem is also discussed in many undergraduate statistical books. Bancroft (1944, 1964, 1965) implemented the idea of PTE in the ANOVA setup to study the effect of pre-testing on the estimation of variance.

The PTE was also studied for the simple regression model by Ahsanullah and Saleh (1972) and further extended by Ahmed and Saleh (1989). Consider a simple regression model

$$X_i = \theta + \beta c_i + e_i, \quad i = 1, \ldots, n, \tag{2.2.1}$$

where e_i is the error variable that is identically and independently distributed as normal with mean 0 and variance σ^2, $N(0, \sigma^2)$, c_i is the explanatory variable, X_i is the response variable, and θ and β are the unknown intercept and slope parameters respectively. When it is suspected that $\beta = \beta_0$, the UE, RE and PTE, as defined in Chapter 1, are given as below

- The UE of θ is $\tilde{\theta}^{UE} = \bar{X} - \tilde{\beta}\bar{c}$, where $\tilde{\beta}$ is the maximum likelihood (ML) or least-square (LS) estimators of β, $\bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

- The RE of θ is $\hat{\theta}^{RE} = \bar{X} - \hat{\beta}\bar{c}$, where $\hat{\beta} = \beta_0$ is the RE of β.

- If the NSPI on β is uncertain, the uncertainty of the NSPI on the value of β is removed by testing $H_0 : \beta = \beta_0$ with test statistic, $L_n = \frac{(\tilde{\beta}-\beta_0)^2 Q}{SE(\tilde{\beta})}$, where $Q = \sum_{i=1}^{n} c_i^2 - \frac{1}{n}\sum_{i=1}^{n} c_i^2$. Based on the rejection or acceptance of H_0, the PTE is a choice between the UE and the RE. The PTE of θ is $\hat{\theta}^{PT} = \hat{\theta}^{RE} I(L_n < F_{1,n-2}(\alpha)) + \tilde{\theta}^{UE} I(L_n \geq F_{1,n-2}(\alpha))$, where $F_{1,n-2}(\alpha)$ is the α-level upper critical value of a central F distribution with $(1, n-2)$ degrees of freedom and $I(A)$ is the indicator function of the set A that takes value 1 if A occurs, otherwise it is 0.

In the studies on the comparison of the quadratic bias and MSE, the quadratic bias for the PTE increases to a maximum as $\Delta^2 = \frac{(\beta-\beta_0)^2 Q}{\sigma^2}$ moves away from the origin, and decreases toward zero as $\Delta^2 \to \infty$, while the quadratic bias for the RE is linear in Δ^2 and zero for the UE (c.f. Saleh, 2006, p.61). As for the MSE criterion, the RE performs better than the UE whenever $\Delta^2 < 1$, otherwise the UE performs better (c.f. Saleh, 2006, p.62). The PTE performs better than the UE if $\Delta^2 < \Delta_0^2$, where Δ_0^2 is some positive number that depends on the significance level, otherwise the UE is better than the PTE (c.f. Saleh, 2006, p.63).

The UE, RE and PTE were also studied for other regression models such as the parallelism model (Akritas et al., 1984, Lambert et al., 1985a, Khan, 2003) and the multivariate simple regression model (Sen and Saleh, 1979). The properties of these estimators are also discussed in Saleh (2006).

Some studies proposed alternative estimators to the UE, RE and PTE by introducing the coefficient of distrust in the suspected value. This idea was applied to several models such as the location model (Khan and Saleh, 2001), the simple regression model (Khan et al., 2002) and the two suspected parallel models (Khan, 2006*a*). The coefficient of distrust for the NSPI represents the degree of distrust in the null hypothesis. These studies show that the efficiency of the estimators depends on the departure constant, a function of the difference between the suspected and true value of the parameter and also on the coefficient of distrust.

The developments and investigations into the properties of the PTE have been contributed to by a host of authors, notably, Kitagawa (1963), Saleh and Han (1990), Mahdi et al. (1998), Ali and Saleh (1990), Khan and Saleh (2001) and Khan et al. (2002). All of these studies were carried out assuming a normal model. The normal model assumption is now being criticized more and more. The true underlying distribution of the error variables of the model may not follow the normality assumption in practice. The true underlying distribution of the error variables could be the Student-t distribution which is heavier at both tails than that of the normal distribution. The PTE has also been proposed and studied for the non-normal model. Assuming the Student-t distribution on the error terms of the model, the PTE was proposed and investigated for the non-normal model in the literature (see Khan and Saleh, 1997, Khan, 2005, 2008). This PTE is appropriate only if the assumptions for the underlying distribution of the error terms for this parametric model are met.

The non-parametric methods do not rely on assumptions that the observations are drawn from a given probability distribution. The asymptotic theory and properties of the PTE were investigated for the nonparametric estimation cases by Saleh and Sen (1978), Sen and Saleh (1979), Saleh and Sen (1985), Saleh and Sen (1987) and Saleh (2006). Most of the studies on the PTE in the non-parametric setting were formulated using rank statistics. These PTEs are also categorized as the robust R-estimators (c.f. Saleh, 2006, p.109). The analytical results for the asymptotic distributional bias and the asymptotic distributional MSE of the nonparametric UE, RE and PTE are given in (Saleh, 2006, p.114) for several regression models. The PTE obviously depends on the choice of the level of significance (c.f. Saleh, 2006, p.58).

The definition of the UE, RE and PTE using statistics from the robust M-estimation methodology were proposed and their behaviors were investigated in only a few papers (Sen and Saleh, 1987, Saleh and Shiraishi, 1989, Ahmed et al., 2006). These studies found that there is no uniform domination among UE, RE and PTE, but the PTE is a compromise between the UE and RE from the risk efficiency point of view. The PTE performs better than the UE when the true parameter is closer to the suspected NSPI value on the parameter, but the RE may still be better than the others. These results invite the question of whether the UT, RT and PTT under the M-estimation methodology deliver the same kind of properties.

The PTE is a choice of the UE or the RE. It does not allow any smooth transition between the two estimators. The PTE is optimum for large values of the significance level ($\alpha \geq 0.20$) and this is practically inappropriate. A shrinkage estimator (SE) does not depend on the significance level of the pre-test and provides a smooth transition between the UE and RE (c.f. Khan and Saleh, 2001, Khan, 2006*b*, 2008, Saleh, 2006, p.76). The shrinkage (Stein-rule) technique was introduced by Stein (1956) and James and Stein (1961). The estimators that are derived using this technique are also known as the Stein-type estimators. The shrinkage estimator dominates the MLE for its smaller quadratic risk and this important result is recommended in the study of estimating mean for a multivariate normal population. This shrinkage technique is widely used in this area of study for a variety of multivariate models. The pre-test approach to Stein-rule estimation regarding the robust M-estimation method was explored by Saleh and Sen (1985), Sen and Saleh (1987) and Ahmed et al. (2006). The Stein-type estimators were also investigated for a family of Student-t distribution by Khan and Saleh (1995) and Khan and Saleh (1997). Many authors have contributed to this area, notably Sclove et al. (1972), Judge and Bock (1978), Stein (1981), Maatta and Casella (1990) and Khan (1998) to mention a few.

2.3 Robust Estimation – General Idea and Properties

Statistical inference involves two analysis methods: estimation and hypothesis testing. The development of robustness statistical inference was first established significantly by Huber starting in 1964 (Huber, 1964). Good coverage of the basic theory developed during the 1960 and 1970s is provided in his text book.

Huber (1981, p.1) says, "The word *robust* is loaded with many ... connotations. We use it in a relatively narrow sense: for our purposes, robustness signifies insensitivity to small deviations from the assumptions."

The word 'deviations' in the previous paragraph suggests two kinds of robustness, namely distributional robustness and robustness against contamination (c.f. Barnett and Lewis, 1995, p.55). The former suggests procedures that are robust against the possibility that the entire sample comes from some other distributions while the latter are procedures that are robust against the presence of outliers arising from contamination.

The differences among outliers, contaminants and extreme values are discussed in Barnett and Lewis (1995, p.7-9). Extreme values may or may not be outliers but any outlier is always an extreme value. Whether we declare the extreme value as an outlier depend on consideration of how it appears in relation to the assumed model. Outliers may or may not be contaminants and contaminants may or may not be outliers. Outliers are observations which appear to be inconsistent with the remainder of the data while contaminants are observations that arise from distribution other than the assumed one (c.f. Barnett and Lewis, 1995, p.7-9).

Outliers are observations that lie far away from the majority of the data and possibly do not follow the assumed model (c.f. Becker and Gather, 1999). A routine data set typically contains about 1-10% outliers, and even the highest quality data set can not be guaranteed free of outliers (c.f. Hampel et al., 1986, p.26-28). It is well known that they can strongly influence the classical inferential procedure and even cause misleading results. In particular, some classical parametric test and estimator performances are affected by the influential outlying observations (c.f. Becker and Gather, 1999).

Desirable properties of robust procedures are that they should be nearly as efficient as (or perform as well as) the classical procedures in the assumed model and more efficient over (or perform better than) the classical procedures in the presence of small deviations or violations from the model assumptions (c.f. Huber, 1981, Ryan, 1997, p.5). The performance of robust procedures is measured by means of a so-called efficiency criterion. Examples of efficiency criteria are the asymptotic variance of an estimate, the level and power of a test (Huber, 1981, p.5), the mean square error (Montgomery et al., 2001, p.401), the ratio of squared distance (Imon, 2003) and the relative efficiency (Hoaglin et al., 1983, p.327). The efficiency criterion should be close to the nominal value calculated within the assumed model.

There are several basic ways to measure robustness of an estimator, namely, the influence curve, the gross error sensitivity and the breakdown point (see Hampel, 1971, 1974, Hampel et al., 1986). Basically, the breakdown point is simply the smallest fraction of contamination (or largest possible fraction of the alterations of observations) that can completely ruin an estimator (c.f. Donoho and Huber, 1983, Efron and Tibshirani, 1993, Rousseeuw, 1984, p.157-184, Montgomery et al., 2001, p.400).

Several definitions of empirical influence functions are found in the literature, for instance in Hampel et al. (1986, p.93) and Davison and Hinkley (1997, p.46). In spite of the differences, all of them are trying to measure, in some way, the influence that one observation has on the value of an estimate (c.f. Amado and Pires, 2004). The influence function (Hampel, 1974) gives the amount of changes in the estimator that can be wrought by an infinitesimal amount of contamination (c.f. Coakley and Hettmansperger, 1993). The gross error sensitivity expresses the maximum effect a contaminated observation can have on the estimator (c.f. Goodall, 1983, p.358). There are yet other tools dealing with this robustness measurement such as the sensitivity curves, the local shift sensitivity and the rejection point. Readers could refer to Goodall (1983) and Wilcox (2005) for an excellent dissertation of this subject matter.

One of the earliest and well known estimators for the purpose of making statistical inference is the LS estimator (c.f. Chakraborty, 1999), which is defined as the estimator that minimizes the sum of squared of residuals. The LS estimator is sensitive to possible variations from the assumed normal model (c.f. Rousseeuw, 1984, Coakley and Hettmansperger, 1993). It is clear that this estimator is not robust, because only one observation can change the estimator to any value. The LS estimator has both a low breakdown point and an unbounded influence function. However, it has the advantage of having the highest possible efficiency when normality assumptions hold (c.f. Coakley and Hettmansperger, 1993). The ML estimator is another classical estimator derived by maximizing the likelihood (probability) function of observations which are assumed to follow a known distribution (c.f. Montgomery et al., 2001, p.50). Thus, it is not robust when the assumed model is not met. As stated in Yohai and Zamar (1988), one of the goals of robust regression estimation is to simultaneously achieve a breakdown point of roughly .50, a bounded influence function and a high efficiency compared to LS when the underlying distribution of the model errors is normal (c.f. Coakley and Hettmansperger, 1993).

The L-, R- and M-estimators have played an important role in the study of robust estimations. L-estimators are linear combinations of order statistics (c.f. Rosenberger and Gasko, 1983, p.306). For example, suppose we wish to estimate the location parameter of a distribution. Let the order statistics of a sample of size n be $x_{[1]} \leq \ldots \leq x_{[n]}$. Let $a_1, \ldots, a_n$ be real numbers, $0 \leq a_i \leq 1$, $i = 1, \ldots, n$, such that $\sum_{i=1}^{n} a_i = 1$. An L-estimator with weights $a_1, \ldots, a_n$ is given by $L_n = \sum_{i=1}^{n} a_i x_{[i]}$. Sample mean and trimmed mean are two examples of L-estimators. The sample mean of a sample size n is an L-estimator with all weights equal to $1/n$. A trimmed mean is identified by the proportion that is trimmed off from each of the ordered samples.

R-estimates are derived from the nonparametric rank tests. Let $X_1, \ldots, X_n$ be n independently and identically distributed random variables with continuous distribution function F_θ, where θ is the median of the distribution and is assumed to be unique. Then, R_n, an R-estimator of θ is the value of t such that

$$R_n = \frac{1}{2}\left[\sup\{t : S_n(t) > 0\} + \inf\{t : S_n(t) < 0\}\right], \tag{2.3.2}$$

where $S_n(t) = \sum_{i=1}^{n} \mathrm{sign}(X_i - t), \quad t \in \Re$. Here, $S_n(t)$ is nondecreasing in $t \in \Re$. This R-estimator of location is derived from the sign test. See Jurečková and Sen (1996, p.103-4) for more general cases of R-estimators. Important references on R-estimation in regression include Adichie (1967), Jurečková and Sen (1996, Ch.3 & 6) and Montgomery et al. (2001, p.407).

The M-estimation method is a generalization of the maximum likelihood estimation (MLE) (c.f. van der Vaart, 1998, p.61, Thisted, 1988, p.150). The method minimizes a function of a residual or solving for the root of an estimating equation (c.f. Carroll and Ruppert, 1988, p.209).

There are other robust estimators studied in the literature, such as the least median of squares estimator (LMS) (Rousseeuw, 1984, Rousseeuw and Leroy, 1987), the least trimmed of squares estimator (LTS) (Rousseeuw, 1984, Jung, 2005, Rousseeuw and Van Driessen, 2006), the generalized M-estimators (GM or also known as the bounded influence estimators) (Krasker and Welsch, 1982, Wilcox, 2005, Ryan, 1997, p.218-223), reweighted least squares, S-estimators (Rousseeuw and Leroy, 1987) and MM-estimators (Yohai, 1987).

For a location model, the breakdown point of an M-estimator with an unbounded objective function is about 0.5 (c.f. Huber, 1984, Zhang and Li, 1998) in the presence of vertical outliers (observations remote in the dependent variables direction). However, the breakdown point for the M-estimator is the same as the LS estimator in the presence of bad leverage points (observations remote in the independent variables direction) (c.f. Montgomery et al., 2001, p.406). So, the M-estimator is not robust against a leverage point (c.f. Hampel, 2001). The leverage points are points with large values on the diagonal of what is usually called the hat matrix (c.f. Kelly, 1992). They could be given full weight by an M-estimator. As a result, the GM-estimator is developed. The GM-estimator tolerates a small positive fraction of the leverage point and has a breakdown point of $1/p$ where p is the number of parameters (c.f. Montgomery et al., 2001, p.406). The LMS, LTS, S- and MM-estimators are estimators with very high breakdown points (c.f Montgomery et al., 2001, p.401-405), however, they need a lot of computing power, so they too can only be used for rather low-dimensional parameter spaces (c.f. Hampel, 2001).

Another important aspect of robustness properties is asymptotic efficiency. Unfortunately, the estimators with a high breakdown point such as the LMS, LTS and S -estimators have very small asymptotic efficiency. They perform poorly relative to the LS under the assumed model (Montgomery et al., 2001, p.404-406). The M-estimator has a very high efficiency under the assumed model relative to the LS estimator (c.f. Goodall, 1983, p.388-395).

Huber (1964) introduced the concept of minimax robust. The Huber M-estimator minimizes the asymptotic variance over some neighborhood of the model (Huber, 1983). The famous Huber-estimator solves a minimax problem for the contaminated normal data, thus being an optimal compromise for a whole neighborhood of the normal model as well as being numerically almost optimal under the normality assumption (c.f. Hampel, 2001). Huber (1983) examined the GM-estimator proposed by Krasker and Welsch (1982) and criticised this estimator in terms of the aspect of minimax properties and developed a minimax approach to handle the influence of leverage points.

In this thesis, the UT, RT and PTT are proposed using tests formulated in the M-estimation methodology. The asymptotic properties of the M-estimator are already established in the literature and have a long track record since they were first introduced more than 40 years ago. Huber (1981, p.132), Jurečková and Sen (1996, p.218-220), Rieder (1994, p.11) and van der Vaart (1998, p.44) are among authors who provide studies using the asymptotic theory of the M-estimators and test statistics in their articles and text books. The regularity conditions for the existence of a consistent M-estimator are also provided in these text books.

For this thesis, the studies on the behaviour of the power functions of the UT, RT and PTT only require the stated regularity conditions. Few of the latest articles use the regularity conditions proposed by Jurečková and Sen (1996) (see Appendix A.1) in their studies. For instance, Ahmed et al. (2006) proposed the PTE and shrinkage estimator under these regularity conditions. In the thesis, we also use the regularity conditions by Jurečková and Sen (1996) to guarantee the consistency of the M-estimators. Since the idea is to use the existing asymptotic results of the M-estimation methodology in the pre-testing framework, we do not intend to update the regularity conditions of a consistent M-estimator, because this is not essential for this study. Readers are referred to a few recent articles, for example, Bantli (2004) and Bachmaier (2007) that provide the nonstandard regularity conditions of a consistent M-estimator. Bantli (2004) proposed the M-estimators that allow for the discontinuity density function while Bachmaier (2007) studied the consistency of the redescending M-estimators.

The asymptotic distributions of the M-estimators are not required to derive the power functions of the UT, RT and PTT. Instead, the asymptotic distributions of the test statistics are required to derive the power functions of UT, RT and PTT under parallel regularity conditions such as those of the M-estimators. In order to derive the asymptotic distribution of the test statistics, results in Lemma 5.5.1 of Jurečková and Sen (1996) are used in this thesis. Therefore, under the standard regularity conditions by Jurečková and Sen (1996) (see Appendix A.1), the UT, RT and PTT are proposed in this thesis.

2.3.1 M-estimation for Location

Let $X_1, \ldots, X_n,\ n \geq 1$ be independently and identically distributed random variables with an unknown distribution G. For a parametric family of distribution function $\{F_\theta(x) : \theta \in \Theta\},\ \ \Theta \subseteq R_p,\ \ p \geq 1$, we wish to estimate θ for which F_θ provides the closest approximation of G.

For the location model, we assume that

$$F_\theta(x) = F_0(x - \theta), \tag{2.3.3}$$

where θ is real and F_0 belongs to a class $\mathfrak{F}_0$ (see Jurečková and Sen, 1996, p.79). We may write (2.3.3) as

$$X_i = \theta + e_i, \ \ i = 1, \ldots, n, \tag{2.3.4}$$

where the error e_i is assumed to be independently and identically distributed with a distribution function, F_0.

The classical LS procedure tries to minimize the sum of the squared of residuals to obtain its estimates. Despite of its mathematical beauty and computational simplicity, there is an increasing criticism of the LS estimator for its dramatic lack of robustness. A single outlier can have an arbitrarily huge effect on the LS estimates (c.f. Rousseeuw, 1984). The LS estimator is sensitive to any observation with a large residual. The M-estimation procedure (Huber, 1973, p.80) is introduced based on the idea of replacing the squared residuals by a function of residuals, say $\rho(\cdot)$, which is known as the objective function. Thus, the M-estimator, say $\hat{\theta}_n$ is the solution (with respect to t) of the minimization of

$$\sum_{i=1}^{n} \rho(X_i - t), \tag{2.3.5}$$

where t is a real number.

If $F_\theta(x)$ has the density $f(x, \theta)$ which is differentiable in θ, choosing $\rho(x, \theta) = -\log f(x, \theta)$, we find an M-estimator is an ML estimator (c.f. Carroll and Ruppert, 1988, p.210, Jurečková and Sen, 1996, p.81). Choosing $\rho(x, t) = \frac{1}{2}(x - t)^2$, for a real $x,\ t$, an M-estimator is an LS estimator (c.f. Goodall, 1983, Montgomery et al., 2001, p.388).

The minimization problem in equation (2.3.5) leads to the solving of equation

$$\sum_{i=1}^{n} \psi(X_i - t) = 0, \tag{2.3.6}$$

with the score function $\psi(x, t) = \frac{\partial}{\partial t}\rho(x, t)$ for all x and t.

The influence function of $\hat{\theta}_n$ is given by

$$IF(x; F, \hat{\theta}_n) = \psi(x)/\gamma(F), \tag{2.3.7}$$

where

$$\gamma(F) = \int \psi'(x) dF(x), \tag{2.3.8}$$

which means the influence function of $\hat{\theta}_n$ is proportional to the score function, ψ. Also, the $IF(x; F, \hat{\theta}_n)$ is bounded whenever ψ is bounded and $\gamma(F) \neq 0$.

Note, for the LS estimator, the score function $\psi(x, t) = x - t, -\infty < x < \infty$. Obviously the influence of a datum on the estimate increases linearly with the magnitude of its error. Thus, the LS method is non-robust since the alteration of a single observation is sufficient to yield any significant offset. As for the Huber estimator (Huber, 1964), the score function is given by

$$\psi(x, t) = \begin{cases} x - t & |x - t| \leq k \\ k \operatorname{sign}(x - t) & |x - t| > k \end{cases} \tag{2.3.9}$$

for $k > 0$. This estimator has a monotone ψ function and does not weight large residuals as heavily as the LS (c.f. Montgomery et al., 2001, p.388). The Huber estimator is asymptotically minimax over a contaminated neighborhood of a standard normal distribution (c.f. Jurečková and Sen, 1996, Hampel, 2001, p.35).

If $\psi(x, t)$ is monotone in t, then the existence of M-estimators can be established under very general regularity conditions (see Huber, 1981, Ch.6) and a consistent sequence of solutions of 2.3.6 can be easily identified under the same regularity conditions. Moreover, for monotone $\psi(x, t)$, the asymptotic normality result is generally obtained. However, the monotonicity of ψ in t is only a sufficient condition and the existence of a consistent M-estimator can also be established for some non-monotone $\psi(x, t)$. Under the general regularity conditions, the asymptotic distribution of a consistent M-estimator is given by

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\psi, F)) \tag{2.3.10}$$

with $\sigma^2(\psi, F) = \gamma^{-2} \int \psi^2(x, F)dF(x)$ (c.f Jurečková and Sen, 1996, p.182-3). The M-estimator for the regression cases also carry the same properties as the location model (c.f Jurečková and Sen, 1996, p.84).

Usually ρ is a convex function and ψ is a bounded variation. For monotone ψ (nondecreasing), the solution to (2.3.6) is expressed as

$$\tilde{\theta}_n = \frac{1}{2}\left[\sup\left\{t : \sum \psi(X_i - t) > 0\right\} + \inf\left\{t : \sum \psi(X_i - t) < 0\right\}\right]. \tag{2.3.11}$$

Note for nondecreasing ψ, $\sum \psi(X_i - t)$ is nonincreasing in $t \in \Re$. Hence $\hat{\theta}_n$ represents the centroid of the set of solutions of (2.3.6) and it removes the possible arbitrariness of such a solution.

2.4 Statistical Tests using Robust Methodology

2.4.1 Hypothesis Testing

A hypothesis is represented by a family of probability distributions of the set of observations X (see Hájek et al., 1999, p.22). Let $H = \{p\}$ or $H = \{P\}$ be the null hypothesis and $K = \{q\}$ or $K = \{Q\}$ be the alternative hypothesis with density p and distribution P are members of H while density q and distribution Q are members of K. The space X is divided into two disjoint parts, the critical region, A_K and the region of acceptance A_H. Whenever the observed value x of X falls into A_K, the null hypothesis H is rejected or else it is accepted. There are two errors one may commit, the Type I error (reject H when it is true) or the Type II error (failure to reject H when it is false). To keep the probability of Type I low, choose a number α, $0 < \alpha < 1$ with condition, $P(X \in A_K) \leq \alpha$ for all $P \in H$. The number α is called the level of significance. The test is based on a statistic $t(x)$, called the test statistic. The correspondence between A_K and $t(x)$ are three types: (i) $\{x \in A_K\} \Longleftrightarrow \{t(x) \geq c_u\}$, (ii) $\{x \in A_K\} \Longleftrightarrow \{t(x) \leq c_l\}$, (iii) $\{x \in A_K\} \Longleftrightarrow \{t(x) \geq c_u \text{ or } t(x) \leq c_l\}$.

The first two cases (i) and (ii) are for the one sided tests based on t and the last case (iii) is for the two sided test based on t. The numbers c_u and c_l are called the upper and the lower critical value, respectively. The test function is defined by

$$\Psi(x) = I\left(t(x) < c_l \ \text{ or } \ t(x) > c_u\right). \tag{2.4.12}$$

According to Hájek et al. (1999, p.23), the size of $\Psi(x)$ is defined as $\int \Psi dP$ or $\sup_{P \in H} \int \Psi dP$ for composite hypotheses. The power of $\Psi(x)$ is defined as $\int \Psi dQ$ or $\sup_{Q \in K} \int \Psi dQ$ for composite hypotheses. The main purpose of the theory of hypothesis is to provide tests with the largest power for a given level of significance (Hájek et al., 1999, p.24).

Often, the level of significance α is chosen and is set as the probability of a Type I error (see Wackerly et al., 2008, p.491, De Veaux et al., 2009, p.544, Hettmansperger and McKean, 1998, p.16). However, when interest is to test the intercept parameter given the NSPI on the slope parameter of the simple regression model, the (actual) size of the test is different from the level of significance. Note, the preassigned level of significance is sometimes known as the nominal size of the test (see Pinto et al., 2003, Carolan and Rayner, 2000). Saleh and Sen (1982, 1983) and Lambert et al. (1985a) show that the (actual) asymptotic size of the RT is larger than the preassigned level of significance. The size and power of the test are obtained from the power function of the test (see Pinto et al., 2003, Saleh and Sen, 1982, 1983).

2.4.2 Robust Test and its Properties

A test is said to be robust if the power of the test is not significantly affected by any departures from the model assumptions (c.f. Burt and Barber, 1996, p.332) and when the nominal and actual sizes are not significantly different under a slight model failure (c.f. Carolan and Rayner, 2000). According to Heritier and Ronchetti (1994), there are two robustness properties a test should achieve. First, the level (size) of a test should be stable under small departures from the null hypothesis (i.e. robustness of validity). Second, the test should still have a good power under small departures from specified alternatives (i.e. robustness efficiency).

One would expect that the sensitivity of an estimator to departures from model assumptions should be inherited by the statistical test which is derived using the same estimation methodology (c.f. Schrader and Hettmansperger, 1980). Using the M-estimation methodology, several robust versions of the Wald, scores and likelihood ratio test have been proposed and investigated in the literature (Schrader and Hettmansperger, 1980, Sen, 1982, Shiraishi, 1990, Silvapulle, 1992, Wu et al., 2007). The text by Jurečková and Sen (1996, Ch.10) provides a note about robust statistical tests for the location model and few other regression models. These robust tests are analogous to the classical counterpart, see for example, Silvey (1975).

The influence with respect to the residuals is bounded for the tests derived using the M-estimation methodology; however, the influence with respect to the regressor is unbounded. As a result, robust tests that are formulated using the GM methodology have been introduced and studied by Markatou and Hettmansperger (1990), Heritier and Ronchetti (1994) and Gagliardini et al. (2005) in the literature. However, the asymptotic distributions of the tests formulated using GM methodology are complicated (c.f. Muller, 1998) and thus may cause difficulties in deriving the power function for the PTT.

The influence function and breakdown point are two important properties in statistics to measure robustness of an estimator; these properties are carried over to the hypothesis testing framework. The influence function approach for parameter estimation developed by Ronchetti and Rousseeuw (see Hampel et al., 1986, Ch.3) is extended to the testing situation (c.f. Heritier and Ronchetti, 1994). The idea is to define a level influence function and power influence function (Hampel et al., 1986, p.198) that describe the influence of contamination on the asymptotic level and power. Further studies of the influence function of the tests are presented by Markatou and Hettmansperger (1990), Markatou and He (1994), Büning (2000), Wang and Qu (2007).

The breakdown point gives the maximum amount of contamination that a test can tolerate (c.f. He et al., 1990, Heritier and Ronchetti, 1994). The breakdown versions for tests were proposed and studied in the literature by few authors, namely Rieder (1982), He et al. (1990), Heritier and Ronchetti (1994), Hettmansperger and McKean (1998, p.31) and Wang and Qu (2007). The power breakdown function gives the amount of contamination of each alternative distribution that can carry the test statistic to a null value. The level breakdown function gives the amount of contamination of a null distribution that can carry the test statistic to each value in the alternative space (c.f. He et al., 1990).

A comprehensive picture of rank tests based methodology is available in Adichie (1967), Hájek et al. (1999), Puri and Sen (1985), Jurečková and Sen (1996, Ch.3 & 6) and Hettmansperger and McKean (1998, p.11). The robustness properties of rank tests are given in Rieder (1982) and Büning (2000). Note that the rank tests have been used to propose the PTE (Saleh and Sen, 1978, 1987, 1985, Sen and Saleh, 1979, Saleh, 2006) and PTT (Saleh and Sen, 1982, 1983) in many published articles for a number of models including the simple regression model, the multivariate simple regression model and the Cox proportional hazard model. However, there is no study found in the literature that uses rank tests to propose the PTT in the parallelism model, multiple linear regression model and multivariate multiple model.

Some robustness properties such as the level and power influence function and breakdown point of a test are not studied in this thesis. The robustness properties of the proposed tests are not studied analytically in this thesis. However, the robustness properties of the proposed tests are investigated on data simulated using the Monte Carlo method. In the simulation, the sensitivity of the tests to aberrant observations are investigated by comparing the size and power of the tests derived using different score functions and under different distributions of the simulated data.

2.4.3 Tests using M-estimation Methodology

This section discusses three robust tests in the M-estimation methodology, namely the robust score type M-test (Sen, 1982), the robust likelihood ratio test (Schrader and Hettmansperger, 1980) and the robust Wald type test for the location model of equation (2.3.4). See Heritier and Ronchetti (1994) for an account of these tests for the multivariate model.

The asymptotic distribution of Wald and score tests are found to be the same as the classical counterpart but the asymptotic distribution for the likelihood ratio test has more complicated asymptotic distribution. The Wald and score type tests are asymptotically equivalent under both the null hypothesis and the alternative (c.f. Heritier and Ronchetti, 1994). Sen (1982) claims that the M-test based on score function is simpler in computation and as efficient as the likelihood ratio test proposed by Schrader and Hettmansperger (1980). The score type test has better global stability than the Wald type M-test for contamination of null data. For local alternatives in the one lay-out, the power breakdown function of an M-test depends on the choice of the score function (c.f. He et al., 1990).

The Score Type M-test

Recall the aforementioned location model of equation (2.3.3). Consider the null hypothesis, $H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1, \theta_1 > (< \text{ or } \neq)\ \theta_0$ for testing the location parameter at the specified value θ_0.

If distribution function F is symmetric and specified, the classical likelihood ratio test (LRT) would be useful for testing H_0. The LRT is optimal when the assumed model holds but it is non-robust even to small departures from assumed F (Jurečková and Sen, 1996, p.408).

Suppose we can identify the least favourable distribution, F_0, where $F_0 \in \{\mathfrak{F}_0\}$ in such a neighborhood, with respect to the two hypotheses. A robust test is obtained by constructing the usual LRT statistic corresponding to the least favourable distribution function, F_0. Huber (1965) claims that this robust test has a maximin power property.

In the same spirit, another robust test statistic, L_n is obtained by constructing the usual (Rao) efficient score statistics corresponding to the least favourable distribution, F_0, thus

$$L_n = \sum_{i=1}^{n} \left(\frac{\partial}{\partial \theta} \right) \log f_0(X_i, \theta)|_{\theta_0}$$

with density function, f_0, corresponding to F_0.

Replacing L_n by an M-statistic, M_n, we obtain

$$M_n = \sum_{i=1}^{n} \psi(X_i - \theta_0),$$

where ψ is the score function. This statistic resembles the Rao efficient score statistic.

Asymptotically

$$n^{-1/2}(M_n/\sigma_\psi) \xrightarrow{d} N(0,1)$$

under H_0 with $\sigma_\psi^2 = \int \psi^2(x)dF(x)$ (Jurečková and Sen, 1996, p.410). The test statistic is given by

$$T_n = n^{-1/2}(M_n/\hat{\sigma}_n) \text{ with } \hat{\sigma}_n^2 = n^{-1}\sum \psi^2(X_i - \hat{\theta}_n),$$

where $\hat{\theta}_n$ is the M-estimator of θ. Under $H_0 : \theta = \theta_0$, T_n is asymptotically normal. For a one-sided alternative hypothesis, $H_1 : \theta > \theta_0$, the (asymptotic) critical level is denoted by τ_α $(0 < \alpha < 1)$. The main justification of this test statistic T_n is the choice of a suitable robust ψ. Now consider a sequence of local alternative $H_n : \theta = \theta_0 + n^{-1/2}\lambda$ for some $\lambda \in \Re$. Using Theorem 5.3.2 of Jurečková and Sen (1996), the asymptotic power function of the test is given by

$$\beta(\lambda) = 1 - \Phi(\tau_\alpha - \gamma\lambda/\sigma_\psi)$$

(c.f. Jurečková and Sen, 1996, p.410), where γ is as defined in equation (2.3.8) and $\lambda \in \Re$. Here, $\Phi(\cdot)$ is the cdf of a normal distribution. Hence, an optimal ψ relates to the maximization of γ/σ_ψ. Note, the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta)$ is σ_ψ^2/γ^2.

As in Huber (1965), we may look for a particular score function $\psi_0 : \Re \to \Re$ such that

$$\sup_{F \in \mathfrak{F}_0} \sigma_\psi^2/\gamma^2$$

is a minimum at $\psi = \psi_0$. We obtain an asymptotically maximin power M-test of the score type with ψ_0 in $\mathfrak{F}_0$. For example, ψ_0 is the Huber function (given in equation (2.3.9)) provided that $\mathfrak{F}_0$ is the family of contaminated normal distributions.

The LRT

Let

$$Z_n = \sum_{i=1}^{n} [\rho(X_i - \theta_0) - \rho(X_i - \hat{\theta}_n)],$$

where $\hat{\theta}_n$ is the M-estimator of θ. Z_n is analogous to the classical LR type statistic. Under H_0, $2\gamma(Z_n/\sigma_\psi^2) \rightarrow \chi_1^2$ (chi-squared distribution function with 1 degree of freedom). The test statistic is $Z_n^\star = 2\hat{\gamma}_n\{Z_n/\hat{\sigma}_n\}$, where $\hat{\gamma}_n$ is an estimated value of γ. $Z_n^\star$ has asymptotically a noncentral chi-squared distribution function with 1 d.f. and noncentrality parameter, $\Delta^\star = \gamma^2\lambda^2\sigma_\psi^{-2}$. Note, for the T_n, only the estimate of σ_ψ^2 is required while for the $Z_n^\star$, both estimates of σ_ψ^2 and γ are required. The LR type statistics for the linear model were proposed by Schrader and McKean (1977) and Schrader and Hettmansperger (1980).

The Wald Test

Another robust test resembles the classical Wald test, the term based on Wald (1943). Under the assumed regularity conditions, we obtain the asymptotic distribution of the M-estimator as has been given in equation (2.3.10). Thus, the robust Wald test statistic is

$$W_n = n(\hat{\theta}_n - \theta_0)'[\hat{\sigma}_n^2/\gamma_n^2]^{-1}(\hat{\theta}_n - \theta_0)$$

and $W_n \rightarrow \chi_1^2$ under H_0. Jurečková and Sen (1996, p.419) and Carroll and Ruppert (1988, p.214) discussed this test in their text books.

In this thesis, the score type M-test is chosen because it is simpler in computation and it is as efficient as the LRT and Wald tests.

Note, the asymptotic properties of the M-statistics by Jurečková (1977, Theorem 4.1) and Jurečková and Sen (1996, p.221) are used to derive the asymptotic distribution of the test statistics under the local alternative hypotheses in this thesis. The contiguity probability measures concept (discussed in the next two Sections) is applied to derive the asymptotic distributions of the test statistics under the sequence of a local alternative hypothesis.

In the next Section, review on the literature of the PTT is given. The studies of the PTT are categorized under the parametric tests and non-parametric tests counterparts. Most of the parametric PTTs assume normal distribution for the underlying distribution of the error variables. Nonparametric PTTs based on rank statistics were also proposed and their performances were investigated by few authors.

2.5 The UT, RT and PTT

Although there is a large number of published articles investigating the properties of the PTE (see Section 2.2), only a limited number of studies found in the literature investigate the PTT.

There are few articles found in the literature that use pre-testing in the analysis of variance framework for some special parametric models (Bechhofer, 1951, Bozivich et al., 1956, Paull, 1950, among others). The pooling procedure for the hypothesis testing using analysis of variance techniques was considered by Bozivich et al. (1956) and Mead et al. (1975). Let V_1, V_2, V_3 be the doubtful error mean square, the error mean square and the treatment mean square respectively, with the corresponding degrees of freedom, n_1, n_2, n_3 and expectations $\sigma_1^2, \sigma_2^2, \sigma_3^2$. There are three possible cases depending on the equality of σ_1^2 and σ_2^2 when the primary interest is to test $H_0^\star : \sigma_3^2 = \sigma_2^2$,

- Case I: When there is no specification or suspicion that $\sigma_1^2 = \sigma_2^2$, the test statistic is $F_1 = V_3/V_2$ (i.e. comparing V_3 with V_2).

- Case II: If $\sigma_1^2 = \sigma_2^2$, the test statistic is $F_2 = V_3/V$ where $V = \frac{n_1V_1+n_2V_2}{n_1+n_2}$. Here, V_1 and V_2 are pooled then V_3 is compared with V to obtain F_2.

- Case III: If $\sigma_1^2 = \sigma_2^2$ is uncertain, then a pre-test of $H_0' : \sigma_1^2 = \sigma_2^2$ with test statistic $F_3 = V_2/V_1$ is performed first following the ultimate test on $H_0^\star$. If H_0' is accepted that is $F_3 < F_{\alpha,n_1,n_2}$, F_2 is used to test $H_0^\star$ otherwise F_1 is used. Here, F_{α,n_1,n_2} is the upper $100\alpha\%$ critical point of a central F distribution with n_1 and n_2 degrees of freedom.

Bozivich et al. (1956) studied the pooling problem for hypothesis testing using analysis of variance for a random model case, while a fixed model case is later studied by Mead et al. (1975). For the two studies of the fixed and random models, the error variables are assumed to follow the normal distribution, so these models are parametric models. In their studies, the size and power of the tests were derived and the choice of significance level for the pre-test was recommended.

Ohtani and his colleges also proposed the PTT for the parametric models. Ohtani and Toyoda (1985) considered the problem of testing the linear hypothesis of regression coefficients after pre-testing the disturbance variance. The problem of testing the equality of regression coefficients after pre-testing the equality of disturbance variances in two linear regression models is later considered by Toyoda and Ohtani (1986). Ohtani (1988) and Ohtani and Giles (1993) extended the ideas related to these parametric problems in their articles.

The PTT has also been proposed and investigated by Tamura (1965) for one and two samples of the nonparametric problems. To propose the PTT, the sign test was used for the final test while the Wilcoxon rank test was used for the pre-test on the uncertain NSPI. The performance (size and power) of the PTT was investigated for small and large sample sizes. The size and power of the PTT were plotted against the preassigned significance level. As the sample size increases, the size of the PTT tends to the value of the preassigned significance level while the power of the test tends to unity.

Consider the simple linear regression in equation (2.2.1) with normal error variables. In Saleh and Sen (1982), when primarily interest is to test $H_0^{\dagger}: \theta = 0$, the UT, RT and PTT are respectively given by

- $A_n^{UT} = \frac{\sqrt{n}(\bar{X}-\tilde{\beta}\bar{c})}{s_n^{\star}\sqrt{1+n\bar{c}^2}}$ when β is unspecified. $H_0^{\dagger}$ is rejected if $A_n^{UT} > t_{n-2,\alpha_2}$.
- $A_n^{RT} = \frac{\sqrt{n}\bar{X}}{s_n}$ when $\beta = 0$. $H_0^{\dagger}$ is rejected if $A_n^{RT} > t_{n-1,\alpha_1}$.
- The PTT is a choice of the UT or RT which depends on the acceptance or rejection of $H_0^{(1)}: \beta = 0$. If $H_0^{(1)}$ with test statistic $A_n^{PT} = \frac{\sqrt{Q_n}\tilde{\beta}}{s_n^{\star}}$ is rejected (i.e. $A_n^{PT} > t_{n-2,\alpha_3}$), the A_n^{UT} is used to test $H_0^{\dagger}$, else A_n^{RT} is used.

Here $t_{m,\alpha}$ is the upper $100\alpha\%$ critical point of the Student's t distribution function with m degrees of freedom, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $s_n^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$, $\tilde{\beta} = \frac{\sum_{i=1}^{n} X_i(c_i-\bar{c}_n)}{Q_n}$, ${s_n^{\star}}^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X}-\tilde{\beta}(c_i-\bar{c}_n))^2}{n-2}$, $\bar{c}_n = \frac{1}{n}\sum_{i=1}^{n} c_i$, $Q_n = \sum_{i=1}^{n}(c_i-\bar{c}_n)^2$.

In the same spirit, the study on the performance of the UT, RT and PTT has been extended to the parallelism model by Lambert et al. (1985b). Instead of assuming normal error variables, tests have been formulated based on the LS estimators which do not rely on the assumption that the observations follow a specified probability distribution. In their paper, the testing on the equality of intercepts is the primary interest and it may depend on the equality of the slopes. However, the LS estimators for the regression coefficients are identical to the maximum likelihood (ML) estimators when the distribution of the error terms is normal. It is suspected that the statistical tests based on LS estimators also deliver the same properties as the tests based on ML estimators, that is, these tests are powerful only if normality holds.

The UT, RT and PTT proposed using the nonparametric rank tests were studied by Saleh and Sen (1982) for the simple regression model. Note, the robust R-estimates are derived from the rank tests (Huber, 1981, p.281). In the paper, the primary interest is to test the significance of the intercept parameter (testing intercept at 0) that obviously depends on the choices of the slope parameter. However, the study of significance testing on the intercept parameter is less realistic compared to the problem of testing any arbitrary values (including 0) on the intercept and slope. The effect of the pre-test (on the slope) on the size and power of the final test (on the intercept) was investigated in Saleh and Sen (1982) paper. In the findings, the PTT is preferable to the RT for the consideration of asymptotic size, though the UT remains as the best choice. For the consideration of asymptotic power, the PTT is preferable to the UT. However, there are some limited discussions of the investigation into the power of the PTT discussed in the paper. Although the analytical analysis is important and was discussed in the paper, the graphical representation was not given, probably due to the limitation in the computation of the bivariate normal integral at the time the paper was published.

Saleh and Sen (1983) also proposed the UT, RT and PTT using nonparametric rank tests for the problem of testing the significance of the intercept vector when there is an uncertain NSPI on the slope vector of the multivariate simple regression model. The effect of the pre-test on the performance of the final test was studied in the paper and the asymptotic size and power of the UT, RT and PTT were derived. The paper does not provide any computational comparisons (table or graph) of the power function for the three tests to support the analytical results discussed. In order to compute the size and power of the PTT, a bivariate noncentral chi-square distribution is required. The nonexistence tools to compute the probability integral of the bivariate noncentral chi-square distribution may be the reason for not pursuing the computational or graphical comparisons. Moreover, in the analytical part, the authors claimed that in terms of the asymptotic power, the PTT *may* have larger power than that of the UT without providing proof. The statement given was not strongly supported, and furthermore it does not mean that the PTT always has larger power than that of the UT. More discussions on the power comparisons are provided in Chapter 4 of this thesis.

To my knowledge, no study on the performance (size and power) of the UT, RT and PTT for tests formulated using the robust M-estimation methodology can be found in the literature.

2.6 Contiguity

An important concept that dominates the asymptotic theory of statistics is the *contiguity* of probability measures (c.f. Jurečková and Sen, 1996, p.61). Contiguity arguments are a technique to obtain the limit distribution of a sequence of statistics under the alternative hypothesis from a limiting distribution under the null hypothesis (van der Vaart, 1998, p.85). A comprehensive account of this concept and its impact on asymptotic is given in Hájek et al. (1999).

Definition Let $\{p_v\}$ be the sequence of simple hypothesis densities and $\{q_v\}$

be the sequence of simple alternative hypothesis. A sequence $\{p_v, q_v\}$ is defined on measure spaces $(X_v, \mathfrak{A}_v)$, $v \geq 1$. If for any sequence of events $\{A_v\}$, $A_v \in \mathfrak{A}_v$,

$$P_v(A_v) \to 0 \text{ implies } Q_v(A_v) \to 0 \tag{2.6.13}$$

holds as $v \to \infty$, we say the densities $\{q_v\}$ are contiguous to the densities $\{p_v\}$, where $\{Q_v\}$ and $\{P_v\}$ are sequences of simple hypothesis probabilities corresponding respectively to $\{q_v\}$ and $\{p_v\}$.

Contiguity implies that any sequence of random variables converging to zero in P_v-probability converges to zero in Q_v-probability, $v \to \infty$. Generally, we are interested in the asymptotic distribution of statistics $\{T_v(X)\}$. Then, convergence of $\{T_v(X)\} \to 0$ under $\{P_v\}$ implies $\{T_v(X)\} \to 0$ under $\{Q_v\}$ if $\{Q_v\}$ is contiguous to $\{P_v\}$ (c.f. Saleh, 2006, p.44). The Le Cam's first, second and third lemmas are provided in the Appendix A.2.

The concept of contiguity is more popular in R-estimation (rank statistic) than in M-estimation. However, Sen (1982) uses the contiguity of probability measures under sequence of alternative hypothesis to those under the null hypothesis to find the asymptotic distribution of test statistics under the alternative hypothesis. For many years, contiguity probability measures have been used to obtain the asymptotic theory of the PTE (see Saleh, 2006, p.201) that is based on rank tests.

2.7 Bivariate Noncentral Chi-square Distribution

The bivariate noncentral chi-square distribution is involved in the power function of the proposed tests for some studies in the pre-testing area (Saleh and Sen, 1983, Lambert et al., 1985a). However, the computational illustrations for the size and power of the tests were not given in these studies presumably due to the difficulties to compute the complicated form of the distribution of the bivariate noncentral chi-square found in the literature.

There have been a number of proposals and investigations into compound distributions that are derived by compounding several probability distributions in the literature (Dubey, 1970, Hutchinson, 1981, Khan, 2000, Gerstenkorn, 2004). Most of the papers are based on developing theoretical ideas about the proposed distribution rather than dealing with the computational aspect. Although there is a suggestion of deriving the bivariate noncentral chi-square distribution by compounding the Poisson probabilities with the bivariate central chi-square distributions in Kotz et al. (2000, p.475), the density function of this bivariate noncentral chi-square distribution is not provided in their comprehensive text.

More or less, the idea of constructing the compound distribution seems parallel to the idea of generating a distribution function from a mixture of distributions. The distribution of the noncentral bivariate chi-square as a mixture of bivariate central chi-square distribution with Poisson probabilities has been proposed by Marshall and Olkin (1990). For various reasons, however their proposed bivariate noncentral chi-square distribution is not practical for computation. By choosing an appropriate central bivariate chi-square distribution to be compounded with the Poisson probabilities, a different density function of the noncentral bivariate chi-square distribution is defined in Yunus and Khan (2009). The proposed distribution is more meaningful than the previous one, from a computational point of view.

There is a small number of proposed bivariate noncentral chi-square distributions in the literature. Some work in this area is theoretical and provides no computational illustration of the distribution. For example, Royen (1995) proposed a distribution which is too complicated for computation. A theoretical paper on bivariate noncentral chi distribution was also proposed by Krishnan (1967) but it is only possible under some strict restrictions. Kocherlakota and Kocherlakota (1999) approximated the distribution of the bivariate noncentral chi-square by using some transformations on the variables of the bivariate central chi-square distribution. Dharmawansa and McKay (2009) derived the joint density of the noncentral bivariate and trivariate chi-square distribution corresponding to the diagonal elements of a complex noncentral Wishart matrix. However, in some cases, we may not have our variables in Wishart matrix form together with its distribution parameters that are required in the computation of the density function of their proposed bivariate noncentral chi-square distribution.

In this thesis, the following bivariate noncentral chi-square distribution suggested by Yunus and Khan (2009) is used: For the random variables (Y_1, Y_2) the density of a bivariate noncentral chi-square distribution is defined as

$$\phi^{\star}(y_1, y_2) = \sum_{j=0}^{\infty} \sum_{\lambda_1=0}^{\infty} \sum_{\lambda_2=0}^{\infty} f_w(y_1, y_2, \rho^2) \frac{e^{-\theta_1/2}(\theta_1/2)^{\lambda_1}}{\lambda_1!} \frac{e^{-\theta_2/2}(\theta_2/2)^{\lambda_2}}{\lambda_2!}, \quad (2.7.14)$$

where

$$f_w(w_1, w_2, \rho^2) = (1-\rho^2)^{\frac{m}{2}} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{m}{2}+j)\rho^{2j}(w_1 w_2)^{\frac{m}{2}+j-1} e^{-\frac{w_1+w_2}{2(1-\rho^2)}}}{\Gamma(\frac{m}{2}) j! \left[2^{\frac{m}{2}+j}\Gamma(\frac{m}{2}+j)(1-\rho^2)^{\frac{m}{2}+j}\right]^2} \quad (2.7.15)$$

is the density of the bivariate central chi-square distribution (Krishnaiah et al., 1963) with $m/2 + j$ degree of freedom, θ_1 and θ_2 are the noncentrality parameters, ρ^2 is the correlation coefficient between (Z_{1j}, Z_{2j}), $j = 1, \ldots, m$ with $Z_{ij} \sim N(0,1)$, $i = 1,2$ and $W_i = \sum_{j=1}^{m} Z_{ij}^2$, $i = 1,2$ is the chi-square random variable.

2.8 Notation

The notations used for one Chapter should not be referred to in the other Chapters.

Chapter 3

Simple Regression Model

3.1 Introduction

In the study of energy usage in a production plant, the relationship between electricity consumption and production output can be modeled by a simple regression model (see Kent, 2008) with electricity consumption as the response and production output as the independent explanatory variables. Obviously the electricity consumption will increase as the production level goes up. The level of electricity consumption when no effective production is taking place is known as the *base load*. The base load is related to the electricity used in lighting, heating, cooling, office equipment, machine repairs and maintenance. Since the base load is unrelated to the production output, reduction in the base load is profitable to the manufacturer (Kent, 2008).

Consider a simple regression model of n observable random variables, $X_i,\ i = 1,\dots,n$,

$$X_i = \theta + \beta c_i + e_i, \tag{3.1.1}$$

where the errors e_i's are from an unspecified symmetric and continuous distribution function, $F_i,\ i = 1,\dots,n$, the c_i's are known real constants of the explanatory variable and θ and β are the unknown intercept and slope parameters respectively.

The management of the production plant may wish to test whether the base load is equal to a specified value while they are not sure about the value of the slope parameter. In this situation, three different scenarios associated with the value of the slope are considered: the slope would either be (i) completely unspecified, or (ii) specific fixed constant, or (iii) uncertain, but suspected to be a fixed quantity from previous knowledge or expert assessment. For the three possible choices of the slope, the three statistical tests are appropriate, namely the (i) unrestricted test (UT), (ii) restricted test (RT) and (iii) pre-test test (PTT) respectively.

Without loss of generosity, we consider the significance testing of the intercept parameter under various conditions on the slope parameter. Testing the intercept of a simple regression model depends on the knowledge of the slope. To simplify,

(i) for the UT, denote ϕ_n^{UT} as the test function for testing $H_0^{(1)}: \theta = 0$ against $H_A^{(1)}: \theta > 0$ when β is unspecified,

(ii) for the RT, denote ϕ_n^{RT} as the test function for testing $H_0^{(1)}: \theta = 0$ against $H_A^{(1)}: \theta > 0$ when β is 0 (specified) and

(iii) for the PTT, denote ϕ_n^{PTT} as the test function for testing $H_0^{(1)}: \theta = 0$ against $H_A^{(1)}: \theta > 0$ following a pre-test (PT) on the slope. As for the PT, let ϕ_n^{PT} be the test function for testing $H_0^{(2)}: \beta = 0$ against $H_A^{(2)}: \beta > 0$, essential for the PTT on θ. Thus, the PTT is a choice between the UT and the RT. If the null hypothesis $H_0^{(2)}$ is rejected in the pre-test (PT), then the UT is used, otherwise the RT is used.

The rest of the chapter is organized as follows. The method of M-estimation is presented in Section 3.2. In Section 3.3, the UT, RT and PTT are defined. The asymptotic distributions of the proposed test statistics are derived in Section 3.4. In Section 3.5, the asymptotic power functions of the tests are provided. The analytical comparisons of the power functions of the UT, RT and PTT are given in Section 3.6 and comparisons of the power function of the UT, RT and PTT through simulation examples are provided in Section 3.7. The first seven Sections of this Chapter are devoted to testing the significance of the intercept and slope.

In Section 3.8, the UT, RT and PTT for testing the intercept at any arbitrary value (including zero) are proposed. Similarly, the asymptotic distributions of the test statistics and their power functions are derived. The analytical and graphical comparisons of the UT, RT and PTT are provided in the same Section. In Section 3.9, the use of the UT, RT and PT is demonstrated on a real data set. Discussion and conclusions are presented in the final Section of this Chapter.

3.2 The M-estimation

Given an absolutely continuous function $\rho : \Re \rightarrow \Re$, M-estimator of θ and β is defined as the values of θ and β that minimize the objective function

$$\sum_{i=1}^{n} \rho\left(\frac{X_i - \theta - \beta c_i}{S_n}\right). \tag{3.2.1}$$

Here S_n is an appropriate scale statistic for some functional $S = S(F) > 0$. If F is $N(0, \sigma^2)$, $S_n = MAD/0.6745$ is an estimate of $S = \sigma$, where MAD is the mean absolute deviation (Wilcox, 2005, p.78, Montgomery et al., 2001, p.387). The M-estimator of θ and β can also be defined as the solution of the system of equations,

$$\begin{array}{rcccl} \sum_{i=1}^{n} \psi_\theta(X_i) & = & \sum_{i=1}^{n} \psi\left(\frac{X_i-\theta-\beta c_i}{S_n}\right) & = & 0, \\ \sum_{i=1}^{n} \psi_\beta(X_i) & = & \sum_{i=1}^{n} c_i \psi\left(\frac{X_i-\theta-\beta c_i}{S_n}\right) & = & 0, \end{array} \tag{3.2.2}$$

where $\psi(\cdot)$ is known as the score function. If ρ is differentiable with partial derivatives $\psi_\theta = \partial\rho/\partial\theta$ and $\psi_\beta = \partial\rho/\partial\beta$, then the M-estimators that minimize the function in (3.2.1) are the solutions to the system (3.2.2). By contrast, the M-estimators obtained from solving system (3.2.2) may not minimize equation (3.2.1) (c.f Carroll and Ruppert, 1988, p.210). The system of equations (3.2.2) may have more roots, while only one of them leads to a global minimum of (3.2.1).

Consider that

(a) ψ is nondecreasing and skew symmetric that is $\psi(-x) = -\psi(x)$,

(b) the error, e_i has distribution F which is continuous and symmetric about 0.

The distribution F has finite Fisher information,

$$I(f) = \int_{-\infty}^{\infty} \{f'(x)/f(x)\}^2 dF(x), \tag{3.2.3}$$

where $f'(x) = (d/dx)f(x) = (d^2/dx^2)F(x)$. Assume that

(i) finite constants $\bar{c}$ and $C^{\star}(> 0)$ exist such that

$$\lim_{n\to\infty} \bar{c}_n = \bar{c} \quad \text{and} \quad \lim_{n\to\infty} n^{-1} C_n^{\star 2} = C^{\star 2}, \tag{3.2.4}$$

where $\bar{c}_n = n^{-1} \sum_{i=1}^n c_i$ and $C_n^{\star 2} = \sum_{i=1}^n c_i^2 - n\bar{c}_n^2$ both exist.

(ii) the c_i's are all bounded, so that by (i),

$$\max_{1\le i\le n} (c_i - \bar{c}_n)^2 / C_n^{\star 2} \to 0, \quad \text{as } n \to \infty. \tag{3.2.5}$$

We may write (3.1.1) as

$$\boldsymbol{X} = \boldsymbol{C}_n \begin{pmatrix} \theta \\ \beta \end{pmatrix} + \boldsymbol{e}, \tag{3.2.6}$$

where $\boldsymbol{X} = (X_1 \dots X_n)'$ and

$$\boldsymbol{C}'_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ c_1 & c_2 & \cdots & c_n \end{pmatrix}.$$

Note,

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{C}'_n \boldsymbol{C}_n = \lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} n & \sum_{i=1}^n c_i \\ \sum_{i=1}^n c_i & \sum_{i=1}^n c_i^2 \end{pmatrix} = \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} = \boldsymbol{C}'\boldsymbol{C}.$$

Some articles found in the literature, for example Goodall (1983) investigated how to choose a ψ-function for good resistant and robustness of efficiency. It was found that the influence function (IF) of the M-estimator was proportional to the ψ-function (Huber, 1981, p.45) or in other words the IF and ψ have the same shape. The influence function was introduced by Hampel (1968, 1974), indicating the effect on an estimate of adding or deleting an observation in large sample. The mathematical definition is given in many text books including Hoaglin et al. (1983, p.354), Wilcox (2005, p.25) and Huber (1981, p.13). Adapting Corollary 2 of Goodall (1983, p.354) that is used for a location model with a fixed scale to a simple regression model together with assumptions (a) and (b) given above, we find

$$IF(X;F;S) = \frac{\psi\left(\frac{X-\theta-\beta c}{S}\right)}{\gamma}, \tag{3.2.7}$$

where

$$\begin{aligned}\gamma &= E\left[\psi'\left(\frac{X-\theta-\beta c}{S}\right)\right] \\ &= \frac{1}{S}\int_{-\infty}^{\infty}\psi'\left(\frac{X-\theta-\beta c}{S}\right)dF(X-\theta-\beta c).\end{aligned} \tag{3.2.8}$$

Note, for a large sample size, we drop "i" in X_i and "n" in S_n.

A bounded, skew symmetric and continuous ψ-function are among the criteria for choosing a robust M-estimator (see Goodall, 1983, p.365). The ψ-function of the ML (maximum likelihood) method, $\psi_{ML}(U_i) = U_i$, where $U_i = \frac{X_i-\theta-\beta c_i}{S_n}$ is a straight unbounded line, as illustrated in Figure 3.1, so that the MLE is sensitive to any observations and in particular, is adversely affected by outliers. The ψ-function of Huber's method is denoted by $\psi_H(U_i) = U_i$ if $|U_i| \leq k_1$, otherwise it is k_1 sign(U_i), where k_1 is the tuning constant that fine-tunes the robustness of Huber's method. The $\psi_H(U_i)$ is a bounded function (see Figure 3.1), so outliers do not adversely affect the parameter estimation. Note, γ is a positive constant in equation (3.2.7), which means the size but not the shape of the influence function depends on the underlying distribution. Note, if the ML ψ-function is used, then $\gamma = 1/S$. If the Huber ψ-function is used, then $\gamma = F(k_1)/S - F(-k_1)/S$, when $|U| < k_1$, or 0 when $|U| > k_1$.

For a large sample, assume

$$E\left[\psi\left(\frac{X-\theta-\beta c}{S}\right)\right] = \int_{-\infty}^{\infty}\psi\left(\frac{X-\theta-\beta c}{S}\right)dF(X-\theta-\beta c) = 0. \tag{3.2.9}$$

To achieve a zero mean of the $\psi(\cdot)$ function, we may need assumptions (a) and (b). Then, let

$$\sigma_0^2 = \int \psi^2 \left(\frac{X - \theta - \beta c}{S} \right) dF(X - \theta - \beta c), \tag{3.2.10}$$

which means σ_0^2 is the second moment of $\psi(\cdot)$ or $\sigma_0^2 = E\left[\psi^2\left(\frac{X-\theta-\beta c}{S}\right)\right]$. Obviously for the ψ_{ML}, $\sigma_0^2 = \int \left(\frac{X-\theta-\beta c}{S}\right)^2 dF(X-\theta-\beta c)$.

In a large sample, another important property of an estimator is its asymptotic variance. For the M-estimation, under appropriate regularity conditions (see Huber, 1964, Lemma 4 or Huber, 1981, equation (6.3)), there is a special link between the asymptotic covariance and the IF which is given by

$$\begin{aligned} AC(X;F) &= \frac{E\left[\psi^2\left(\frac{X-\theta-\beta c}{S}\right)\right]}{\left\{E\left[\psi'\left(\frac{X-\theta-\beta c}{S}\right)\right]\right\}^2}\left[\boldsymbol{C}'\boldsymbol{C}\right]^{-1} \\ &= \int_{-\infty}^{\infty} IF(X;F)^2 dF(X-\theta-\beta c)\left[\boldsymbol{C}'\boldsymbol{C}\right]^{-1} \\ &= \frac{\sigma_0^2}{\gamma^2}\left[\boldsymbol{C}'\boldsymbol{C}\right]^{-1}. \end{aligned} \tag{3.2.11}$$

The MLE is an unbiased estimator of θ and β when F is $N(0,\sigma^2)$. If the ψ_{ML} is used,

$$AC_{ML} = \int_{-\infty}^{\infty} (X-\theta-\beta c)^2 dF(X-\theta-\beta c) = \sigma^2\left[\boldsymbol{C}'\boldsymbol{C}\right]^{-1}. \tag{3.2.12}$$

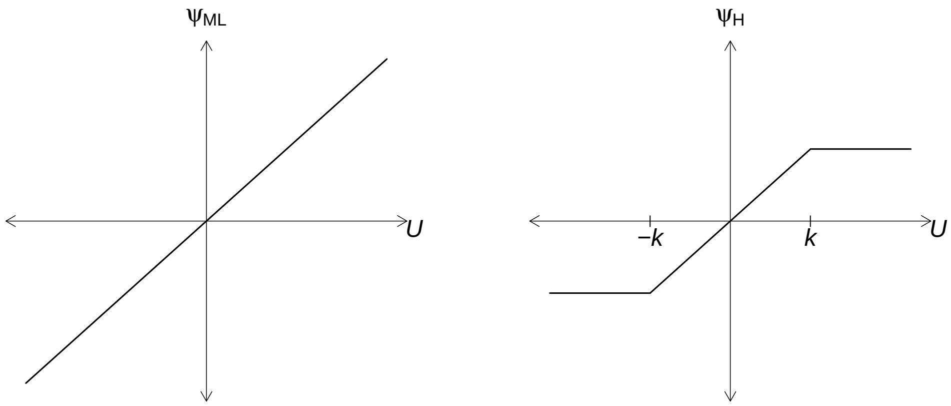
For any real numbers a and b, consider the following statistics,

$$M_{n_1}(a,b) = \sum_{i=1}^{n} \psi\left(\frac{X_i - a - bc_i}{S_n}\right), \quad M_{n_2}(a,b) = \sum_{i=1}^{n} c_i \psi\left(\frac{X_i - a - bc_i}{S_n}\right).$$

Let $\tilde{\theta}$ be the constrained M-estimator of θ when $\beta = 0$, that is $\tilde{\theta}$ is the solution of $M_{n_1}(a,0) = 0$ with respect to a. Similarly, let $\tilde{\beta}$ be the constrained M-estimator of β when $\theta = 0$, that is $\tilde{\beta}$ is the solution of $M_{n_2}(0,b) = 0$ with respect to b. In other words,

$$M_{n_1}(\tilde{\theta},0) = 0 \tag{3.2.13}$$

$$M_{n_2}(0,\tilde{\beta}) = 0. \tag{3.2.14}$$



(a) Maximum Likelihood ψ-function (b) Huber ψ-function

Figure 3.1: Graphs of ψ-functions

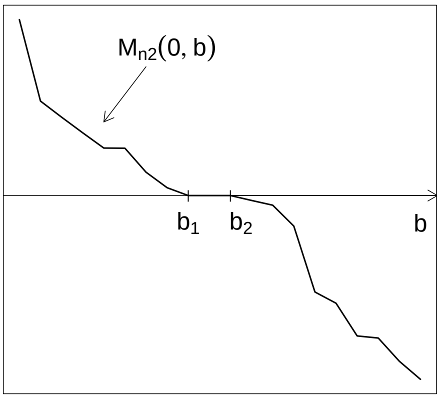
For a nondecreasing monotone $\psi : \Re \rightarrow \Re$ function, $M_{n_1}(a, 0)$ is nonincreasing in a and $M_{n_2}(0, b)$ is nonincreasing in b (c.f. Jurečková and Sen, 1981, p.85). Thus, it may be convenient to write

$$\tilde{\theta} = [\sup\{a : M_{n_1}(a, 0) > 0\} \; + \; \inf\{a : M_{n_1}(a, 0) < 0\}]/2 \tag{3.2.15}$$

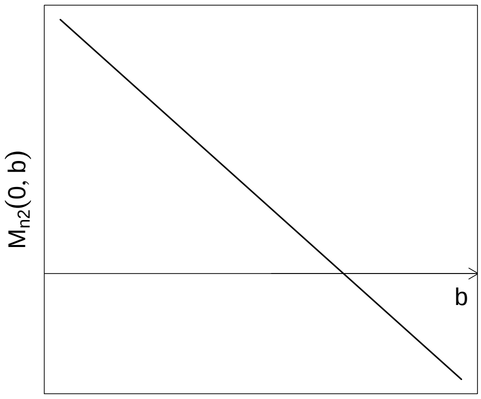
and

$$\tilde{\beta} = [\underbrace{\sup\{b : M_{n_2}(0, b) > 0\}}_{b_1} \; + \; \underbrace{\inf\{b : M_{n_2}(0, b) < 0\}}_{b_2}]/2. \tag{3.2.16}$$

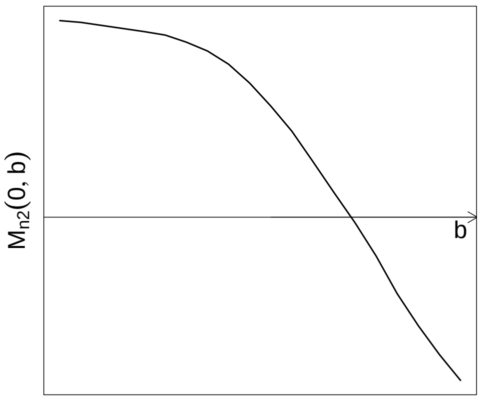
Any value $b_1 < b < b_2$ can serve as the estimate of $M_{n_2}(0, b)$. Figure 3.2(a) exhibits the interpretation of b_1 and b_2. In Figures 3.2(b)-(d), statistic $M_{n_2}(0, b)$ is plotted with respect to b for three different ψ-functions. For the nondecreasing (monotone) ML and Huber ψ-functions (see Figures 3.2(b) and 3.2(c)), statistic $M_{n_2}(0, b)$ is nonincreasing in b. The non monotone Tukey ψ-function is given by $\psi_T(U_i) = (U_i/k_2)[1 - (U_i/k_2)^2]^2$ if $U_i/k_2 \leq 1$ otherwise it is 0. Here, k_2 is the tuning constant for the Tukey ψ-function. The statistic $M_{n_2}(0, b)$ is non monotone in b for the Tukey ψ-function (see Figure 3.2(d)).



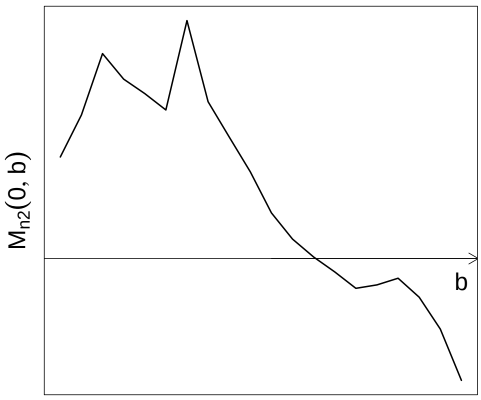
(a) Illustration 1



(b) $M_{n_2}(0, b) = \sum \psi_{ML}(0, b)$



(c) $M_{n_2}(0, b) = \sum \psi_H(0, b)$



(d) $M_{n_2}(0, b) = \sum \psi_T(0, b)$

Figure 3.2: Graphs of $M_{n_2}(0, b)$ functions

Since S_n is an estimate of S, following Jurečková and Sen (1996, p.217), we write

$$n^{\frac{1}{2}}(S_n - S) = O_p(1). \tag{3.2.17}$$

If S is known or if we consider the nonstudentized M-estimator, we may omit this condition.

In this thesis, the asymptotic results of Sen (1982) and Jurečková and Sen (1996, p.221) (see equations B.1.3, B.1.4 and B.1.5 in Appendix B.1) are used to derive the distributions of the proposed tests. For simplicity, we omit condition (3.2.17) and let $S_n = S$ in equation (5.5.29) of Jurečková and Sen (1996, p.221).

Sen (1982) shows that the asymptotic distribution of

$$n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta}, 0) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2}) \tag{3.2.18}$$

under $H_0^{(2)} : \beta = 0$ when the nonstudentized M-estimator is considered. The consistency of ${S_n^{(3)}}^2 = n^{-1} \sum_{i=1}^n \psi^2 \left(\frac{X_i - \tilde{\theta}}{S_n} \right)$ as an estimator of σ_0^2 follows from Jurečková and Sen (1981). Hence, a test statistic $A_n = M_{n_2}(\tilde{\theta}, 0)[C_n^\star \, S_n^{(3)}]^{-1}$ is proposed by Sen (1982). The advantage of this test statistic (score-type M-test) is that it does not require the computation of the unrestricted M-estimates or the estimation of functional γ.

In the same way, it is easy to show that the asymptotic distribution of

$$n^{-\frac{1}{2}} M_{n_1}(0, \tilde{\beta}) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2} / \{C^{\star 2} + \bar{c}^2\}) \tag{3.2.19}$$

under $H_0^{(1)} : \theta = 0$. By the same token, the consistency of ${S_n^{(1)}}^2 = n^{-1} \sum_{i=1}^n \psi^2 \left(\frac{X_i - \tilde{\beta} c_i}{S_n} \right)$ as an estimator of σ_0^2 follows.

3.3 The UT, RT and PTT

3.3.1 The Unrestricted Test (UT)

If β is unspecified, the designated test function is ϕ_n^{UT} with the null hypothesis $H_0^{(1)} : \theta = 0$ against the alternative hypothesis $H_A^{(1)} : \theta > 0$. The testing for

θ involves the elimination of the nuisance parameter β. We consider the test statistic $T_n^{UT} = M_{n_1}(0, \tilde{\beta})$, where $\tilde{\beta}$ is a constrained M-estimator defined in equation (3.2.16). It follows from equation (3.2.19) that under $H_0^{(1)}$,

$$T_n^{UT} / \sqrt{C_n^{(1)} {S_n^{(1)}}^2} \xrightarrow{d} N(0,1) \tag{3.3.1}$$

as $n \to \infty$, with $C_n^{(1)} = n - n^2 \bar{c}_n^2 / \sum c_i^2 = n C_n^{\star 2} / (C_n^{\star 2} + n \bar{c}_n^2)$. We choose α_1 $(0 < \alpha_1 < 1)$ such that for large n,

$$P[T_n^{UT} > \ell_{n,\alpha_1}^{UT} | H_0^{(1)} : \theta = 0] = \alpha_1, \tag{3.3.2}$$

where ℓ_{n,α_1}^{UT} is the critical value of T_n^{UT} at the α_1 level of significance. Let τ_{α_i} be the upper $100\alpha_i$th percentile and $\Phi(\cdot)$ be the cumulative distribution function of the standard normal distribution. Then

$$\Phi(\tau_{\alpha_i}) = 1 - \alpha_i, \quad \text{for} \;\; 0 < \alpha_i < 1, \quad i = 1, 2, 3. \tag{3.3.3}$$

Using (3.3.1), (3.3.2) and (3.3.3),

$$\begin{aligned} 1 - \alpha_1 &= P[T_n^{UT} \le \ell_{n,\alpha_1}^{UT}] = P\left[\frac{n^{-1/2} T_n^{UT}}{\sqrt{\frac{1}{n} {S_n^{(1)}}^2 \frac{n C_n^{\star 2}}{C_n^{\star 2} + n \bar{c}_n^2}}} \le \frac{n^{-1/2} \ell_{n,\alpha_1}^{UT}}{\sqrt{\frac{1}{n} {S_n^{(1)}}^2 \frac{n C_n^{\star 2}}{C_n^{\star 2} + n \bar{c}_n^2}}} \right] \\ &\xrightarrow{p} P\left[\frac{n^{-1/2} T_n^{UT}}{\sqrt{\sigma_0^2 \frac{C^{\star 2}}{C^{\star 2} + \bar{c}^2}}} \le \frac{n^{-1/2} \ell_{n,\alpha_2}^{UT}}{\sqrt{\sigma_0^2 \frac{C^{\star 2}}{C^{\star 2} + \bar{c}^2}}} \right] = \Phi\left(\frac{n^{-1/2} \ell_{n,\alpha_1}^{UT}}{\sqrt{\sigma_0^2 \frac{C^{\star 2}}{C^{\star 2} + \bar{c}^2}}} \right). \end{aligned} \tag{3.3.4}$$

We observe that as $n \to \infty$,

$$n^{-1/2} \ell_{n,\alpha_1}^{UT} / \sqrt{{S_n^{(1)}}^2 C_n^{(1)} / n} \xrightarrow{p} \tau_{\alpha_1} = n^{-1/2} \ell_{n,\alpha_1}^{UT} / \sqrt{\sigma_0^2 C^{\star 2} / (C^{\star 2} + \bar{c}^2)} \;\; \text{(say)}. \tag{3.3.5}$$

So, for the test function $\phi_n^{UT} = I(T_n^{UT} > \ell_{n,\alpha_1}^{UT})$, the power function of the UT becomes $\Pi_n^{UT}(\theta) = E(\phi_n^{UT} | \theta) = P(T_n^{UT} > \ell_{n,\alpha_1}^{UT} | \theta)$, where $I(A)$ stands for the indicator function of the set A. It takes value 1 if A occurs, otherwise it is 0.

3.3.2 The Restricted Test (RT)

If $\beta = 0$, the designated test function is ϕ_n^{RT} for testing the null hypothesis $H_0^{(1)} : \theta = 0$ against the alternative hypothesis $H_A^{(1)} : \theta > 0$. The proposed test statistic is $T_n^{RT} = M_{n_1}(0,0)$. Note that for large n, under $H_0 : \theta = 0, \beta = 0$,

$$n^{-\frac{1}{2}} T_n^{RT} / \sqrt{S_n^{(2)^2}} \xrightarrow{d} N(0,1), \tag{3.3.6}$$

where $S_n^{(2)^2} = n^{-1} \sum \psi^2(X_i/S_n)$. For a large sample size, we find

$$P[T_n^{RT} > \ell_{n,\alpha_2}^{RT} | H_0 : \theta = 0, \beta = 0] = \alpha_2, \tag{3.3.7}$$

where ℓ_{n,α_2}^{RT} is the critical value of T_n^{RT} at the α_2 level of significance. Using equations (3.3.3), (3.3.6) and (3.3.7), we obtain

$$\begin{aligned} 1 - \alpha_2 &= P[T_n^{RT} \leq \ell_{n,\alpha_2}^{RT}] = P\left[n^{-1/2} T_n^{RT} / \sqrt{S_n^{(2)^2}} \leq n^{-1/2} \ell_{n,\alpha_2}^{RT} / \sqrt{S_n^{(2)^2}} \right] \\ &\xrightarrow{p} P\left[n^{-1/2} T_n^{RT} / \sqrt{\sigma_0^2} \leq n^{-1/2} \ell_{n,\alpha_2}^{RT} / \sqrt{\sigma_0^2} \right] = \Phi(n^{-1/2} \ell_{n,\alpha_2}^{RT} / \sigma_0). \end{aligned}$$

Thus as $n \to \infty$ we have

$$n^{-1/2} \ell_{n,\alpha_2}^{RT} / \sqrt{S_n^{(2)^2}} \xrightarrow{p} \tau_{\alpha_2} = n^{-1/2} \ell_{n,\alpha_2}^{RT} / \sqrt{\sigma_0^2} \text{ (say).} \tag{3.3.8}$$

Then, for the test function $\phi_n^{RT} = I(T_n^{RT} > \ell_{n,\alpha_2}^{RT})$, the power of the RT becomes $\Pi_n^{RT}(\theta) = E(\phi_n^{RT} | \theta) = P(T_n^{RT} > \ell_{n,\alpha_2}^{RT} | \theta)$.

3.3.3 The Pre-test (PT)

For the pre-test on the slope, the test function, ϕ_n^{PT} is designed to test the null hypothesis $H_0^{(2)} : \beta = 0$ against the alternative hypothesis $H_A^{(2)} : \beta > 0$. The proposed test statistic is $T_n^{PT} = M_{n_2}(\tilde{\theta}, 0)$, where $\tilde{\theta}$ is a constrained M-estimator (given in equation (3.2.15)). Under $H_0^{(2)}$, it follows from equation (3.2.18) that

$$T_n^{PT} / \sqrt{C_n^{(3)} S_n^{(3)^2}} \xrightarrow{d} N(0,1) \tag{3.3.9}$$

as $n \to \infty$, with $C_n^{(3)} = \sum c_i^2 - n\bar{c}_n^2 = C_n^{\star 2}$.

The consistency of ${S_n^{(1)}}^2$, ${S_n^{(2)}}^2$ and ${S_n^{(3)}}^2$ as estimators of σ_0^2 follows from the law of large numbers (c.f. Jurečková and Sen, 1981).

Similarly, as $n \to \infty$,

$$P[T_n^{PT} > \ell_{n,\alpha_3}^{PT} | H_0^{(2)} : \beta = 0] = \alpha_3. \tag{3.3.10}$$

In the same manner, using (3.3.3), (3.3.9) and (3.3.10), we find that as $n \to \infty$,

$$n^{-1/2}\ell_{n,\alpha_3}^{PT} / \sqrt{{S_n^{(3)}}^2 {C_n^{\star}}^2 / n} \xrightarrow{p} \tau_{\alpha_3} = n^{-1/2}\ell_{n,\alpha_3}^{PT} / \sqrt{\sigma_0^2 C^{\star 2}} \text{ (say)}, \tag{3.3.11}$$

where ℓ_{n,α_3}^{PT} is the critical value of T_n^{PT} at the α_3 level of significance.

3.3.4 The Pre-test Test (PTT)

Now, we are in a position to formulate a test function ϕ_n^{PTT} to test $H_0^{(1)} : \theta = 0$ following a pre-test on β. We write

$$\phi_n^{PTT} = I\left[(T_n^{PT} \leq \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{UT})\right] \tag{3.3.12}$$

as the test function for testing $H_0^{(1)} : \theta = 0$ after a pre-test on β. The function enables us to define the power of the test ϕ_n^{PTT}, that is given by

$$\begin{aligned} & \Pi_n^{PTT}(\theta) = E(\phi_n^{PTT} | \theta) \\ = \; & P[T_n^{PT} \leq \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT} | \theta] + P[T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{UT} | \theta]. \end{aligned} \tag{3.3.13}$$

In general, the power function of the PTT depends on $\alpha_1, \alpha_2, \alpha_3, \theta, n$ as well as β. Note, the size of the PTT is a special case of the power function of the PTT when $\theta = 0$. Since the nuisance parameter β is unknown, but, suspected to be close to 0, it is of interest to study the dependence of both α_n^{PTT} and $\Pi_n^{PTT}(\theta)$ on β (close to 0).

3.4 Asymptotic Distributions under Local Alternatives

The contiguity concept is utilized to derive the asymptotic distributions of statistics $n^{-\frac{1}{2}}[\,T_n^{RT}, T_n^{PT}]$ and $n^{-\frac{1}{2}}[\,T_n^{UT}, T_n^{PT}]$ under K_n (defined in Theorem 3.4.1). The concept of contiguity is more popular in R-estimation (rank statistic) than in M-estimation. However, Sen (1982) used the contiguity of probability measures under $H_n : \beta = n^{-\frac{1}{2}}\lambda$ to those under $H_0' : \beta = 0$ to find the asymptotic distribution of $n^{-\frac{1}{2}}\,[M_{n_1}(\theta,0), M_{n_2}(\theta,0)]$ under H_n.

Theorem 3.4.1 *Let $\{K_n\}$ be a sequence of local alternative hypotheses, where*

$$K_n : (\theta,\beta) = (n^{-\frac{1}{2}}\lambda_1, n^{-\frac{1}{2}}\lambda_2), \tag{3.4.1}$$

with $\lambda_1 = \sqrt{n}\theta \geq 0, \lambda_2 = \sqrt{n}\beta \geq 0$ are fixed real numbers. Under $\{K_n\}$, for large sample,

(i)

$$n^{-1/2}\begin{bmatrix} T_n^{RT} \\ T_n^{PT} \end{bmatrix} \xrightarrow{d} N_2\left[\begin{pmatrix} \gamma(\lambda_1+\lambda_2\bar{c}) \\ \gamma\lambda_2 C^{\star 2} \end{pmatrix}, \sigma_0^2\begin{pmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{pmatrix}\right], \tag{3.4.2}$$

(ii)

$$n^{-1/2}\begin{bmatrix} T_n^{UT} \\ T_n^{PT} \end{bmatrix} \xrightarrow{d} N_2\left[\begin{pmatrix} \frac{\gamma\lambda_1 C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ \gamma\lambda_2 C^{\star 2} \end{pmatrix}, \sigma_0^2\begin{pmatrix} \frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2} & -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} & C^{\star 2} \end{pmatrix}\right]. \tag{3.4.3}$$

Proof of part (i) of Theorem 3.4.1: Following Jurečková and Sen (1996, p.259), let $\{P_n\}$ and $\{Q_n\}$ denote the probability distributions with the densities $p_n = \prod_{i=1}^n f(X_i)$ and $q_n = \prod_{i=1}^n f(X_i - n^{-\frac{1}{2}}\lambda_1 - n^{-\frac{1}{2}}\lambda_2 c_i)$ of the null hypothesis H_0 and the alternative hypothesis K_n, respectively. Note that under (3.1.1), (3.2.4), (3.2.5) and (3.4.1), the contiguity of the sequence of probability measures under $\{K_n\}$ to those under H_0 follows from Le Cam's first and second lemmas (Hájek et al., 1999, Ch.7). We are interested in the asymptotic distribution of the joint statistics $\left[n^{-\frac{1}{2}}\ T_n^{RT},\ n^{-\frac{1}{2}}\ T_n^{PT}\right]$. Here, convergence of $\left[n^{-\frac{1}{2}}T_n^{RT}, n^{-\frac{1}{2}}T_n^{PT}\right] + \Upsilon \to [0,0]$ under H_0 implies $\left[n^{-\frac{1}{2}}T_n^{RT}, n^{-\frac{1}{2}}T_n^{PT}\right] + \Upsilon \to [0,0]$ under $\{K_n\}$ since the probability measures under $\{K_n\}$ are contiguous to those under H_0 (c.f. Saleh, 2006, p.44). Here, Υ is a known vector.

Under $H_0 : \theta = 0,\ \beta = 0$, with relation to (B.1.4) and (B.1.5),

$$n^{-1/2}M_{n_2}(\tilde{\theta}, 0) = n^{-1/2}M_{n_2}(0,0) - n^{\frac{1}{2}}\gamma\tilde{\theta}\bar{c} + o_p(1) \text{ and} \tag{3.4.4}$$

$$n^{-1/2}M_{n_1}(\tilde{\theta}, 0) = n^{-1/2}M_{n_1}(0,0) - n^{\frac{1}{2}}\gamma\tilde{\theta} + o_p(1). \tag{3.4.5}$$

Recalling definition (3.2.15), the equation (3.4.5) reduces to

$$n^{-1/2}M_{n_1}(0,0) = n^{\frac{1}{2}}\gamma\tilde{\theta} + o_p(1), \tag{3.4.6}$$

and hence the equation (3.4.4) becomes

$$n^{-1/2}M_{n_2}(\tilde{\theta}, 0) = n^{-1/2}M_{n_2}(0,0) - n^{-1/2}M_{n_1}(0,0)\bar{c} + o_p(1). \tag{3.4.7}$$

Therefore, under H_0, we find

$$\begin{bmatrix} n^{-1/2}M_{n_1}(0,0) \\ n^{-1/2}M_{n_2}(\tilde{\theta},0) \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-1/2}M_{n_1}(0,0) \\ n^{-1/2}M_{n_2}(0,0) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.4.8}$$

Now utilizing the contiguity of probability measures under $\{K_n\}$ to those under H_0, the equation (3.4.8) implies that $[n^{-1/2}M_{n_1}(0,0), n^{-1/2}M_{n_2}(\tilde{\theta},0)]'$ under $\{K_n\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(0,0) \\ n^{-\frac{1}{2}}M_{n_2}(0,0) \end{bmatrix}$$

under H_0. But the asymptotic distribution of the above random vector under $\{K_n\}$ is the same as

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(-n^{-\frac{1}{2}}\lambda_1, -n^{-\frac{1}{2}}\lambda_2) \\ n^{-\frac{1}{2}}M_{n_2}(-n^{-\frac{1}{2}}\lambda_1, -n^{-\frac{1}{2}}\lambda_2) \end{bmatrix}$$

under H_0 due to the fact that the distribution of $M_{n_1}(a,b)$ under $\theta=a,\beta=b$ is the same as that of $M_{n_1}(\theta-a,\beta-b)$ under $\theta=0,\beta=0$, and similar to $M_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332)

Note that under $H_0:\theta=0,\beta=0$, with relation to (B.1.4) and (B.1.5),

$$\begin{aligned} n^{-\frac{1}{2}}M_{n_1}(-n^{-\frac{1}{2}}\lambda_1,-n^{-\frac{1}{2}}\lambda_2) &= n^{-\frac{1}{2}}M_{n_1}(0,0)-\gamma(\lambda_1+\lambda_2\bar{c})+o_p(1) \text{ and} \\ n^{-\frac{1}{2}}M_{n_2}(-n^{-\frac{1}{2}}\lambda_1,-n^{-\frac{1}{2}}\lambda_2) &= n^{-\frac{1}{2}}M_{n_2}(0,0)-\gamma\{\lambda_1\bar{c}+\lambda_2(C^{\star 2}+\bar{c}^2)\}+o_p(1). \end{aligned}$$

Hence, under H_0, $n^{-\frac{1}{2}}[M_{n_1}(-n^{-\frac{1}{2}}\lambda_1,-n^{-\frac{1}{2}}\lambda_2),M_{n_2}(-n^{-\frac{1}{2}}\lambda_1,-n^{-\frac{1}{2}}\lambda_2)]'$

$$\xrightarrow{d} N_2\left[\begin{pmatrix}\gamma(\lambda_1+\lambda_2\bar{c}) \\ \gamma\{\lambda_1\bar{c}+\lambda_2(C^{\star 2}+\bar{c}^2)\}\end{pmatrix},\ \sigma_0^2\begin{pmatrix}1 & \bar{c} \\ \bar{c} & C^{\star 2}+\bar{c}^2\end{pmatrix}\right] \tag{3.4.9}$$

by equation (B.1.3).

Thus, the distribution of $n^{-\frac{1}{2}}[T_n^{RT},T_n^{PT}]'=n^{-\frac{1}{2}}[M_{n_1}(0,0),\,M_{n_2}(\tilde{\theta},0)]'$ under $\{K_n\}$ is bivariate normal with mean vector

$$\begin{bmatrix}1 & 0 \\ -\bar{c} & 1\end{bmatrix}\begin{bmatrix}\gamma(\lambda_1+\lambda_2\bar{c}) \\ \gamma\{\lambda_1\bar{c}+\lambda_2(C^{\star 2}+\bar{c}^2)\}\end{bmatrix}=\begin{bmatrix}\gamma(\lambda_1+\lambda_2\bar{c}) \\ \gamma\lambda_2 C^{\star 2}\end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix}1 & 0 \\ -\bar{c} & 1\end{bmatrix}\sigma_0^2\begin{pmatrix}1 & \bar{c} \\ \bar{c} & C^{\star 2}+\bar{c}^2\end{pmatrix}\begin{bmatrix}1 & 0 \\ -\bar{c} & 1\end{bmatrix}'=\sigma_0^2\begin{bmatrix}1 & 0 \\ 0 & C^{\star 2}\end{bmatrix}. \tag{3.4.10}$$

Since the two statistics $n^{-\frac{1}{2}}T_n^{RT}$ and $n^{-\frac{1}{2}}T_n^{PT}$ are uncorrelated, asymptotically, they are independently distributed normal variables.

Proof of part (ii) of Theorem 3.4.1: Under $H_0:\theta=0,\beta=0$, using equations (3.2.16), (3.4.7), (B.1.4) and (B.1.5), as $n\to\infty$,

$$\begin{bmatrix}n^{-\frac{1}{2}}M_{n_1}(0,\tilde{\beta}) \\ n^{-\frac{1}{2}}M_{n_2}(\tilde{\theta},0)\end{bmatrix}-\begin{bmatrix}1 & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2} \\ -\bar{c} & 1\end{bmatrix}\begin{bmatrix}n^{-\frac{1}{2}}M_{n_1}(0,0) \\ n^{-\frac{1}{2}}M_{n_2}(0,0)\end{bmatrix}\xrightarrow{p}\begin{bmatrix}0 \\ 0\end{bmatrix}. \tag{3.4.11}$$

Now by using the contiguity of probability measures under $\{K_n\}$ to those under H_0, the equation (3.4.11) implies that $[n^{-\frac{1}{2}}M_{n_1}(0,\tilde{\beta}),n^{-\frac{1}{2}}M_{n_2}(\tilde{\theta},0)]'$ under

$\{K_n\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} 1 & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2} \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(0,0) \\ n^{-\frac{1}{2}} M_{n_2}(0,0) \end{bmatrix}.$$

But the asymptotic distribution of the above random vector under $\{K_n\}$ is the same as

$$\begin{bmatrix} 1 & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2} \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(-n^{-\frac{1}{2}}\lambda_1, -n^{-\frac{1}{2}}\lambda_2) \\ n^{-\frac{1}{2}} M_{n_2}(-n^{-\frac{1}{2}}\lambda_1, -n^{-\frac{1}{2}}\lambda_2) \end{bmatrix}$$

under H_0. Then it follows that by equation (3.4.9), $n^{-\frac{1}{2}}[T_n^{UT}, T_n^{PT}]' = n^{-\frac{1}{2}}\ [M_{n_1}(0,\tilde{\beta}),\ M_{n_2}(\tilde{\theta},0)]'$ is bivariate normal with mean vector

$$\begin{bmatrix} 1 & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2} \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} \gamma(\lambda_1+\lambda_2\bar{c}) \\ \gamma\{\lambda_1\bar{c}+\lambda_2(C^{\star 2}+\bar{c}^2)\} \end{bmatrix} = \begin{bmatrix} \gamma\lambda_1 C^{\star 2}/(C^{\star 2}+\bar{c}^2) \\ \gamma\lambda_2 C^{\star 2} \end{bmatrix}$$

and covariance matrix

$$\begin{aligned} &\begin{bmatrix} 1 & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2} \\ -\bar{c} & 1 \end{bmatrix} \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2}+\bar{c}^2 \end{pmatrix} \begin{bmatrix} 1 & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2} \\ -\bar{c} & 1 \end{bmatrix}' \\ = \ &\sigma_0^2 \begin{bmatrix} C^{\star 2}/(C^{\star 2}+\bar{c}^2) & -\bar{c}\,C^{\star 2}/(C^{\star 2}+\bar{c}^2) \\ -\bar{c}\,C^{\star 2}/(C^{\star 2}+\bar{c}^2) & C^{\star 2} \end{bmatrix}. \end{aligned} \tag{3.4.12}$$

Clearly, the two test statistics $n^{-\frac{1}{2}}T_n^{UT}$ and $n^{-\frac{1}{2}}T_n^{PT}$ are not independent, but rather correlated.

3.5 Local Asymptotic Power Functions

In this Section, the asymptotic power functions of the UT, RT and PTT are derived by using the results obtained in the previous Sections of this Chapter.

Under $\{K_n\}$, the asymptotic power function for the PTT is given by

$$\begin{aligned} \Pi_n^{PTT}(\lambda_1,\lambda_2) &= E(\phi_n^{PTT}|K_n) \\ &= P[T_n^{PT} \le \ell_{n,\alpha_3}^{PT}, T_n^{RT} \ge \ell_{n,\alpha_2}^{RT}|K_n] + \\ &\quad\ P[T_n^{PT} \ge \ell_{n,\alpha_3}^{PT}, T_n^{UT} \ge \ell_{n,\alpha_1}^{UT}|K_n]. \end{aligned} \tag{3.5.1}$$

Note that

$$
\begin{aligned}
& P[T_n^{PT} \leq \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT} | K_n] \\
= \;& P\left[\frac{n^{-\frac{1}{2}} T_n^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{{S_n^{(3)}}^2 C_n^{\star 2}/n}} \leq \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_3}^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{{S_n^{(3)}}^2 C_n^{\star 2}/n}}, \right. \\
& \left. \frac{n^{-\frac{1}{2}} T_n^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{{S_n^{(2)}}^2}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_2}^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{{S_n^{(2)}}^2}} \right] \\
\xrightarrow{p} \;& P\left[\frac{n^{-\frac{1}{2}} T_n^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{\sigma_0^2 C^{\star 2}}} \leq \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_3}^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{\sigma_0^2 C^{\star 2}}}, \right. \\
& \left. \frac{n^{-\frac{1}{2}} T_n^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{\sigma_0^2}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_2}^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{\sigma_0^2}} \right], \qquad (3.5.2)
\end{aligned}
$$

as $n \to \infty$ because the limit of ${S_n^{(2)}}^2$ and ${S_n^{(3)}}^2$ are σ_0^2 and $C_n^{\star 2}/n \xrightarrow{p} C^\star$ as $n \to \infty$.

From equations (3.3.3), (3.3.8) and (3.3.11) and (3.4.2), the probability statement in (3.5.2) becomes

$$
\Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0)[1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2 \bar{c})/\sigma_0)].
$$

Note that T_n^{RT} and T_n^{PT} are independent by equation (3.4.2).

Define $d(q_1, q_2 : \rho)$ to be the bivariate normal probability integral for random variables x and y,

$$
d(q_1, q_2; \rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \int_{q_1}^{\infty} \int_{q_2}^{\infty} \exp\left\{ \frac{-(x^2 + y^2 - 2\rho xy)}{2(1-\rho^2)} \right\} dxdy, \quad (3.5.3)
$$

where q_1, q_2 are real numbers and $-1 < \rho < 1$.

Since ${S_n^{(1)}}^2$ and ${S_n^{(3)}}^2$ both converge to σ_0^2, and $C_n^{(1)}/n \xrightarrow{p} C^{\star 2}/(C^{\star 2} + \bar{c}^2)$ as

$n \to \infty$, we observe that

$$
\begin{aligned}
& P[T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{UT} | K_n] \\
= \; & P\left[\frac{n^{-\frac{1}{2}} T_n^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{{S_n^{(3)}}^2 C_n^{\star 2}/n}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_3}^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{{S_n^{(3)}}^2 C_n^{\star 2}/n}}, \right. \\
& \left. \frac{n^{-\frac{1}{2}} T_n^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2} + \bar{c}^2)}{\sqrt{{S_n^{(1)}}^2 C_n^{(1)}/n}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_1}^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2} + \bar{c}^2)}{\sqrt{{S_n^{(1)}}^2 C_n^{(1)}/n}} \right] \\
\xrightarrow{p} \; & P\left[\frac{n^{-\frac{1}{2}} T_n^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{\sigma_0^2 C^{\star 2}}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_3}^{PT} - \gamma\lambda_2 C^{\star 2}}{\sqrt{\sigma_0^2 C^{\star 2}}}, \right. \\
& \left. \frac{n^{-\frac{1}{2}} T_n^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2} + \bar{c}^2)}{\sqrt{\sigma_0^2 C^{\star 2}/(C^{\star 2} + \bar{c}^2)}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_1}^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2} + \bar{c}^2)}{\sqrt{\sigma_0^2 C^{\star 2}/(C^{\star 2} + \bar{c}^2)}} \right]
\end{aligned}
\tag{3.5.4}
$$

as $n \to \infty$. Further, we write equation (3.5.4) as

$$
d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_{\alpha_1} - \gamma\lambda_1 \sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0; -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}\ )
$$

by using equations (3.3.5), (3.3.11), (3.4.3) and (3.5.3). Note that T_n^{UT} and T_n^{PT} are not independent because of (3.4.3).

Hence, the asymptotic power function for the PTT becomes

$$
\begin{aligned}
& \Pi_n^{PTT}(\lambda_1, \lambda_2) = E(\phi_n^{PTT} | K_n) \to \Pi^{PTT}(\lambda_1, \lambda_2) \\
& = \Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0)[1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2 \bar{c})/\sigma_0)] \; + \\
& \quad d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_{\alpha_1} - \gamma\lambda_1 \sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0; -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}\ ).
\end{aligned}
\tag{3.5.5}
$$

Similarly, the asymptotic power function for the RT is given by

$$
\begin{aligned}
\Pi_n^{RT}(\lambda_1, \lambda_2) &= E(\phi_n^{RT} | K_n) = P[T_n^{RT} > \ell_{n,\alpha_2}^{RT} | K_n] \\
&= P\left[\frac{n^{-\frac{1}{2}} T_n^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{{S_n^{(2)}}^2}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_2}^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{{S_n^{(2)}}^2}} \right] \\
&\xrightarrow{p} P\left[\frac{n^{-\frac{1}{2}} T_n^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{\sigma_0^2}} > \frac{n^{-\frac{1}{2}} \ell_{n,\alpha_2}^{RT} - \gamma(\lambda_1 + \lambda_2 \bar{c})}{\sqrt{\sigma_0^2}} \right]
\end{aligned}
\tag{3.5.6}
$$

since ${S_n^{(2)}}^2 \xrightarrow{p} \sigma_0^2$. Combining equations (3.3.3) and (3.3.8), the asymptotic power function for the RT becomes

$$\Pi^{RT}(\lambda_1, \lambda_2) = 1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2 \bar{c})/\sigma_0). \tag{3.5.7}$$

Finally, the asymptotic power function for the UT is obtained as

$$\begin{aligned}
\Pi_n^{UT}(\lambda_1, \lambda_2) &= E(\phi_n^{UT}|K_n) = P[T_n^{UT} > \ell_{n,\alpha_1}^{UT}|K_n] \\
= \quad & P\left[\frac{n^{-\frac{1}{2}}T_n^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2}+\bar{c}^2)}{\sqrt{{S_n^{(1)}}^2 C_n^{(1)}/n}} > \frac{n^{-\frac{1}{2}}\ell_{n,\alpha_1}^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2}+\bar{c}^2)}{\sqrt{{S_n^{(1)}}^2 C_n^{(1)}/n}}\right] \\
\xrightarrow{p} \quad & P\left[\frac{n^{-\frac{1}{2}}T_n^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2}+\bar{c}^2)}{\sqrt{\sigma_0^2 C^{\star 2}/(C^{\star 2}+\bar{c}^2)}} > \frac{n^{-\frac{1}{2}}\ell_{n,\alpha_1}^{UT} - \gamma\lambda_1 C^{\star 2}/(C^{\star 2}+\bar{c}^2)}{\sqrt{\sigma_0^2 C^{\star 2}/(C^{\star 2}+\bar{c}^2)}}\right]
\end{aligned} \tag{3.5.8}$$

since ${S_n^{(1)}}^2 \xrightarrow{p} \sigma_0^2$. Further, the asymptotic power function for the UT is written as

$$\Pi^{UT}(\lambda_1, \lambda_2) = 1 - \Phi(\tau_{\alpha_1} - \gamma\lambda_1 \sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}\,/\sigma_0) \tag{3.5.9}$$

using equations (3.3.3) and (3.3.5).

3.6 Analytical Comparison

This Section provides an analytic comparison of the asymptotic power functions of the UT, RT and PTT.

If we consider $\bar{c} = 0$ in equation (3.5.5),

$$\begin{aligned}
\Pi^{PTT}(\lambda_1, \lambda_2) = \Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0)[1 - \Phi(\tau_{\alpha_2} - \gamma\lambda_1/\sigma_0)] + \\
[1 - \Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0)][1 - \Phi(\tau_{\alpha_1} - \gamma\lambda_1/\sigma_0)].
\end{aligned} \tag{3.6.1}$$

Letting $\alpha_1 = \alpha_2 = \alpha$ and from equations (3.5.7), (3.5.9) and (3.6.1), we observe that the power functions for the UT, RT and PTT are the same, i.e.

$$\begin{aligned}
\Pi^{UT}(\lambda_1, \lambda_2) &= \Pi^{RT}(\lambda_1, \lambda_2) = \Pi^{PTT}(\lambda_1, \lambda_2) \\
&= 1 - \Phi(\tau_\alpha - \gamma(\lambda_1 + \lambda_2 \bar{c})/\sigma_0).
\end{aligned} \tag{3.6.2}$$

From equations (3.5.5) and (3.5.7),

$$
\begin{aligned}
& \Pi^{RT}(\lambda_1,\lambda_2)-\Pi^{PTT}(\lambda_1,\lambda_2) \\
= \;& 1-\Phi(\tau_{\alpha_2}-\gamma(\lambda_1+\lambda_2\bar{c})/\sigma_0) \\
- \;& \Phi(\tau_{\alpha_3}-\gamma\lambda_2 C^\star/\sigma_0)[1-\Phi(\tau_{\alpha_2}-\gamma(\lambda_1+\lambda_2\bar{c})/\sigma_0)] \\
& -d(\tau_{\alpha_3}-\gamma\lambda_2 C^\star/\sigma_0,\tau_{\alpha_1}-\gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}/\sigma_0;-\bar{c}/\sqrt{C^{\star 2}+\bar{c}^2}\;) \\
= \;& d(\tau_{\alpha_3}-\gamma\lambda_2 C^\star/\sigma_0,\tau_{\alpha_2}-\gamma(\lambda_1+\lambda_2\bar{c})/\sigma_0;0) \\
& -d(\tau_{\alpha_3}-\gamma\lambda_2 C^\star/\sigma_0,\tau_{\alpha_1}-\gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}/\sigma_0;-\bar{c}/\sqrt{C^{\star 2}+\bar{c}^2}\;).
\end{aligned}
\tag{3.6.3}
$$

Letting $\alpha_1=\alpha_2=\alpha$, $\bar{c}>0$, $\lambda_2>0$ and $\lambda_1+\lambda_2\bar{c}>\lambda_1\sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}$,

Result (i): $\Pi^{RT}(\lambda_1,\lambda_2)>\Pi^{PTT}(\lambda_1,\lambda_2)$ from equation (3.6.3) and

Result (ii): $\Pi^{RT}(\lambda_1,\lambda_2)>\Pi^{UT}(\lambda_1,\lambda_2)$ from equations (3.5.7) and (3.5.9).

On the contrary, taking $\alpha_1=\alpha_2=\alpha$, $\bar{c}<0$, $\lambda_2>0$ and $\lambda_1+\lambda_2\bar{c}<\lambda_1\sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}$,

Result (iii): $\Pi^{RT}(\lambda_1,\lambda_2)<\Pi^{PTT}(\lambda_1,\lambda_2)$ from equation (3.6.3) and

Result (iv): $\Pi^{RT}(\lambda_1,\lambda_2)<\Pi^{UT}(\lambda_1,\lambda_2)$ from equations (3.5.7) and (3.5.9).

From equations (3.5.7) and (3.5.9), when $\lambda_1=\lambda_2=0$ and $\alpha_1=\alpha_2=\alpha$, we find $\Pi^{RT}=\Pi^{UT}=\alpha$. Failure to satisfy the conditions does not mean *Result* (i) and *Result* (iii) could not be obtained. However if $\lambda_1=0$, these conditions are always met. Hence, under $H_0^{(1)}:\theta=0$, $\alpha^{RT}>\alpha^{PTT}$ and $\alpha^{RT}>\alpha^{UT}=\alpha$ when $\bar{c}>0$ and $\lambda_2>0$. Letting $\alpha_1=\alpha_2=\alpha$, we write

$$\Pi^{UT}(\lambda_1,\lambda_2)-\Pi^{PTT}(\lambda_1,\lambda_2)=A+B,$$

where $A=[1-\Phi(\tau_\alpha-\gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}/\sigma_0)]-[1-\Phi(\tau_\alpha-\gamma(\lambda_1+\lambda_2\bar{c})/\sigma_0)]$ and $B=d(\tau_{\alpha_3}-\gamma\lambda_2 C^\star/\sigma_0,\tau_\alpha-\gamma(\lambda_1+\lambda_2\bar{c})/\sigma_0;0)-d(\tau_{\alpha_3}-\gamma\lambda_2 C^\star/\sigma_0,\tau_\alpha-\gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2}+\bar{c}^2)}/\;\sigma_0;\;-\bar{c}/\sqrt{C^{\star 2}+\bar{c}^2}\;)$.

For $\bar{c} > 0$, then $\lambda_1 + \lambda_2\bar{c} \geq \lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}$ and $\tau_\alpha - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}$ $\geq \tau_\alpha - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0$. Thus, $A = [1 - \Phi_2] - [1 - \Phi_1] \leq 0$ because $\Phi_1 \leq \Phi_2$ where $\Phi_1 = \Phi(\tau_\alpha - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0)$ and $\Phi_2 = \Phi(\tau_\alpha - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)})$. We observe three cases

$$\Pi^{UT}(\lambda_1, \lambda_2) - \Pi^{PTT}(\lambda_1, \lambda_2) \lesseqqgtr 0 \quad \text{if} \quad B \lesseqqgtr |A|,$$

In a special case, $\lambda_1 = 0 = \lambda_2$, $A = 0$ and $B > 0$, thus, $\Pi^{UT}(0, 0) > \Pi^{PTT}(0, 0)$.

When $\bar{c} > 0$ and $\lambda_2 > 0$, the asymptotic size of the RT is larger than both UT and PTT. For $\bar{c} > 0$ and $\lambda_1 = 0$, the size of the PTT may also be smaller than that of UT (when λ_2 is small). Similarly, for $\bar{c} < 0$, $\alpha^{RT} < \alpha$ and $\alpha^{RT} < \alpha^{PTT}$ while α^{PTT} is closer to α.

Refer to equation (3.5.5), as $\alpha_3 \to 0$ and $\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0 \to \infty$, $\Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0) \to 1$ and $d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0;\ -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}) \to 0$ because one of the lower limits is approaching infinity. Thus, we observe that

$$\Pi^{PTT}(\lambda_1, \lambda_2) \to 1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0) = \Pi^{RT}(\lambda_1, \lambda_2) \text{ as } \alpha_3 \to 0 \tag{3.6.4}$$

Whereas as $\alpha_3 \to 1$ and $\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0 \to -\infty$, $\Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0) \to 0$ and

$$\begin{aligned} d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0\ ; -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}) \\ \to 1 - \Phi(\tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0) \end{aligned}$$

because one of the lower limits is approaching negative infinity. Thus, we observe that

$$\Pi^{PTT}(\lambda_1, \lambda_2) \to 1 - \Phi(\tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0) = \Pi^{UT}(\lambda_1, \lambda_2) \tag{3.6.5}$$

as $\alpha_3 \to 1$.

Now fix the values of λ_1 and λ_2 and let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. For any $\gamma_v/\sigma_{0v} < \gamma_w/\sigma_{0w}$, then

(v) $\Pi^{UT}(\gamma_v, \sigma_{0v}) < \Pi^{UT}(\gamma_w, \sigma_{0w})$, (vi) $\Pi^{RT}(\gamma_v, \sigma_{0v}) < \Pi^{RT}(\gamma_w, \sigma_{0w})$,

(vii) Let $V = \Phi(\tau_\alpha - \gamma_v \lambda_2 C^\star/\sigma_{0v})[1 - \Phi(\tau_\alpha - \gamma_v(\lambda_1 + \lambda_2 \bar{c})/\sigma_{0v})]$ and $W = \Phi(\tau_\alpha - \gamma_w \lambda_2 C^\star/\sigma_{0w})[1 - \Phi(\tau_\alpha - \gamma_w(\lambda_1 + \lambda_2 \bar{c})/\sigma_{0w})]$, we obtain

$$\Pi^{PTT}(\gamma_v, \sigma_{0v}) < \Pi^{PTT}(\gamma_w, \sigma_{0w}) \text{ if } V < W$$
$$\Pi^{PTT}(\gamma_v, \sigma_{0v}) > \Pi^{PTT}(\gamma_w, \sigma_{0w}) \text{ if } W < V.$$

Huber (1964) showed that an estimator generated from the Huber ψ-function is minimaximally robust (i.e. minimax asymptotic bias and minimax asymptotic variance) for a contaminated normal distribution (c.f. Jurečková and Picek, 2006, p.50). Huber (1965) then proposed a robust probability ratio test by using the concept of maximin power of a test. Given any contaminated normal data, the Huber ψ-function that maximizes γ/σ_0 yields an asymptotically maximin power M-test (c.f. Jurečková and Sen, 1996, p.409-410). From results (v) and (vi), the power of the UT and RT is maximum when γ/σ_0 attains its maximum. However, the PTT may not enjoy this maximin power property under certain circumstances (see result (vii)).

The analytical results in this Section are accompanied by an illustrative example in investigating the comparison of the power of the tests discussed in the next Section. The behavior of the power functions corresponding to the probabilities of Type I and Type II errors is also studied. The study of the relationship between the level of significance for the PTT and the nominal size of the PT, as well as the nominal sizes of the UT and RT, is explored.

3.7 Numerical Examples and Simulation Studies

In this Section, numerical examples and results from a simulation study are used to investigate some of the properties of the proposed tests.

The Monte Carlo technique is used in this study to simulate various situations. The sample size considered is $n = 100$ and the model is of the form

$$X_i = 2 + 3c_i + e_i, \quad i = 1, 2, ..., n. \tag{3.7.6}$$

Samples are classified into two cases and data in each case are generated as follows:

Case 1: No contaminant

Generate $e_i \sim N(0, 1)$ for $i = 1, 2, ..., n$.

Case 2: 10 % Contaminant

(i) Generate $e_i \sim N(0, 1)$ for $i = 1, \ldots, 0.9(n)$.

(ii) Generate $e_i \sim U(3, 5), U(-5, -3)$ with 50% for each, $i = 0.9(n+1), \ldots, n$.

Then c_i, $i = 1, \ldots, n$ are generated and we consider three cases of c_i:

(a) $c_i = 0, 1$ with 50% for each, $i = 1, 2, ..., n$. So, $\bar{c} > 0$.

(b) $c_i = -1, 0$ with 50% for each, $i = 1, 2, ..., n$. So, $\bar{c} < 0$.

(c) $c_i = -1, 1$ with 50% for each, $i = 1, 2, ..., n$. So, $\bar{c} = 0$.

The e_i and c_i are then used to calculate X_i in equation (3.7.6). Then, the unrestricted M-estimates for θ and β denoted respectively by $\hat{\theta}$ and $\hat{\beta}$ are computed using R-codes. These estimates are then used to compute the residual, $r_i = X_i - \hat{\theta} - \hat{\beta} c_i$ which is required for the computation of $\hat{\sigma}_0$ and $\hat{\gamma}$. Finally, the power functions of the UT, RT and PTT for the data set are obtained using equations (3.5.5), (3.5.7) and (3.5.9). The simulation is run 3000 times. The average values of the power functions for 3000 data sets are plotted. The R-package (mvtnorm) is used to evaluate the bivariate normal probability integral for the power function of the PTT.

The estimate of σ_0 is obtained using,

$$\hat{\sigma}_0^2 = n^{-1} \sum_{i=1}^{n} \psi^2 \left(\frac{X_i - \hat{\theta} - \hat{\beta} c_i}{S_n} \right) \tag{3.7.7}$$

with $S_n = MAD/0.6745$. For the estimation of γ, an R-estimate from the Wilcoxon sign rank statistics is used. The estimate of γ is the value of t such that $S(V_1, \ldots, V_n, t) = \sum_{i=1}^{n} \text{sign}(V_i - t) a_n(R_{n_i}^+(t)) = 0$, where $R_{n_i}^+(t)$ is the rank of $V_i - t$ and $a_n(k) = k/(n+1), k = 1, \ldots, n$.

In this simulation, we consider two types of ψ-functions in the M-estimation, namely:

(i) The ψ-function for the maximum likelihood (ML) estimation, $\psi_{ML}(U_i) = U_i$ with derivative $\psi'_{ML}(U_i) = 1$ for any U_i.

(ii) The Huber ψ-function $\psi_H(\cdot)$, defined by equation (A.1.1). Three values of tuning constant for the Huber ψ-function are selected, namely $k = 1.04, 1.28$ and 1.64. The value of $k = 1.28$ is the 90th quantile of a standard normal distribution, so there is a 0.8 probability that a randomly sampled observation will have a value between $-k$ and k (see Wilcox, 2005, p.76) while $k = 1.04$ (and 1.64) means there is 0.7 (and 0.90) probability that a randomly sample observation will have a value in the range of $(-1.04, 1.04)$ (and $(-1.64, 1.64)$). When the Huber ψ-function is used, the estimate for σ_0^2 is taken to be $\sum \psi_H^2(U_i)/n$. For the estimation of γ, let $V_i = \psi'_H(U_i)/S_n$, where $\psi'_H(U_i)$ is just the derivative of the Huber ψ-function.

3.7.1 Power Comparison of Huber M-test and ML-based test

In this Section, the power functions of the tests obtained using the ML ψ-function are compared to those of the Huber ψ-function for both the uncontaminated and contaminated responses (see Case 1 and Case 2). The minimax property of the Huber function is observed and the role of the tuning constant of the Huber function is studied in this section. In practice, often the normality assumptions are not met due to the presence of contaminants in the collected data. It is suspected that the power functions of the tests using the ML method are sensitive to departures from normality.

The comparison of the Huber M-test and the ML-based test is carried out for the UT, RT and PTT and is represented through graphs in this Section. Figure 3.3 shows the power curves of the tests against λ_2 ($= \sqrt{n}\beta$) at two values of λ_1 ($= \sqrt{n}\theta$) for both ML and Huber methods in the uncontaminated and contaminated cases. Here, $\lambda_1 = 0$ is chosen to study the asymptotic sizes of the tests and we desire the size of a particular test to be small so that the probability of a Type I error is small. Since we also expect to get a small value of probability of a Type II error, the power of the test at $\lambda_1 = 2$ is considered. An acceptable power function of the test is the one that has smaller values when the null hypothesis is true and larger values when λ_1 differs much from $\theta = 0$. The power curves of the UT, RT and PTT are plotted in separate graphs in Figure 3.3 for the selected value of λ_1.

The role of tuning constant of the Huber ψ-function as a key parameter that control the efficiency and robustness of the procedure is studied (see Figure 3.3). Figure 3.3 displays the power curves obtained using the Huber ψ-function for three different values of tuning constant when there is 10% contamination in the data. The asymptotic size and power obtained using the ML ψ-function for the contaminated and uncontaminated data are also displayed in the same

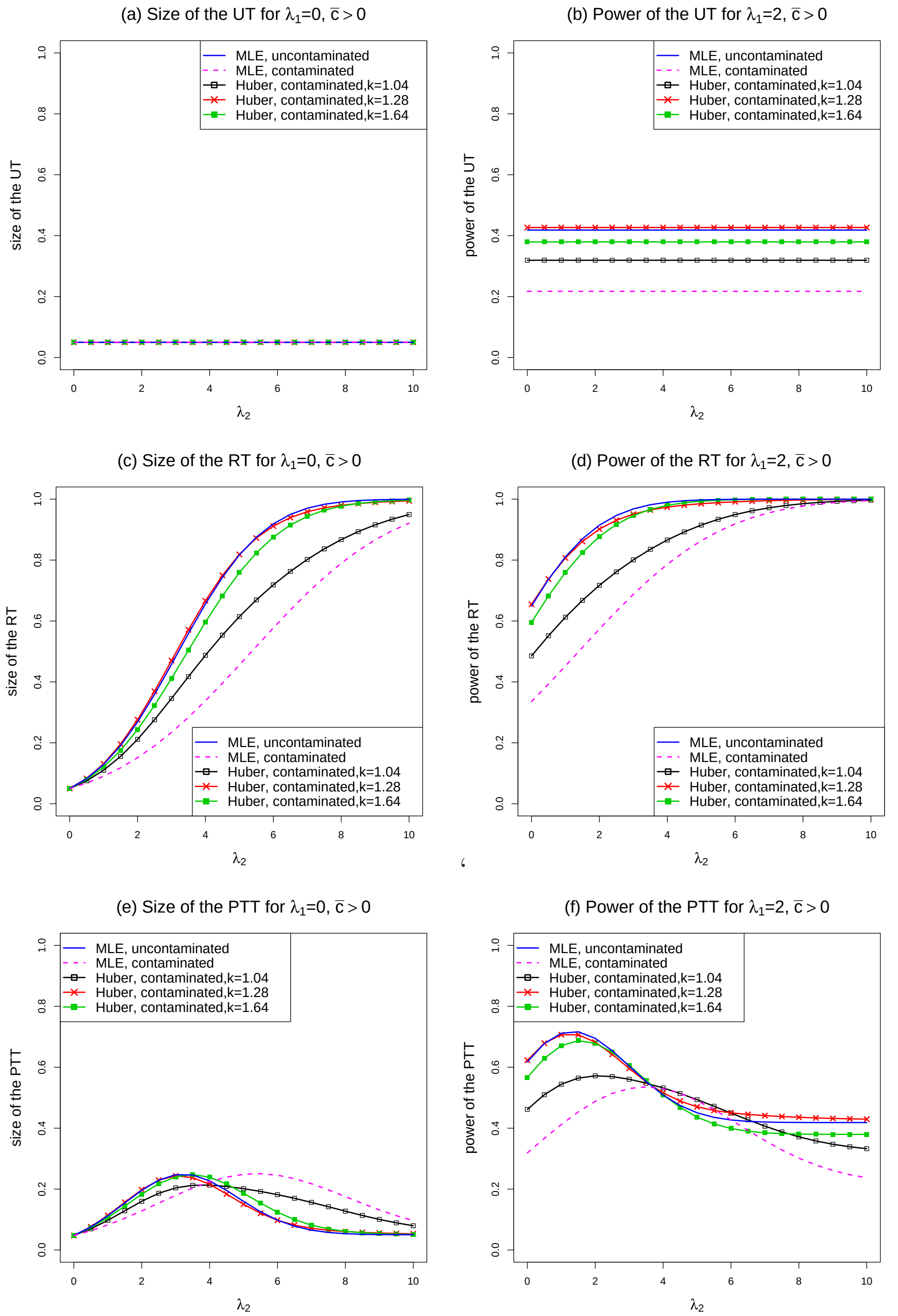


Figure 3.3: Graphs of power functions as a function of λ_2 for selected values of λ_1 and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$.

graphs.

Under the normality assumption, we know that the MLE of θ and β are unbiased estimators. The power function of ML-based test inherits the same good property. The ML-based test becomes the most powerful test when the normality assumptions are met. However, this normality assumption may not be satisfied in practical situations. Studies show that the ML estimator is non-robust when there are departures from the model assumptions or when outliers or contaminants occur in the data. Figure 3.3(a) depicts that the size of the ML-based UT is constant regardless of the value of λ_2 for both contaminated and uncontaminated responses cases. However, the ML-based UT for the contaminated normal data has smaller power than that of the uncontaminated data (see Figure 3.3(b)). Figures 3.3(c)-3.3(f) show that the power curves of the RT and PTT obtained using the ML ψ-function for the contaminated data is far from those of the uncontaminated data. The large distance between the two curves suggests that the ML-based test is not robust when there is contamination in the data.

On the other hand, there is a tuning constant that fine-tunes the robustness of the Huber ψ-function based procedure. The power curves obtained using the Huber ψ-function, with appropriate selection of tuning constant, are closer to the power curves obtained using the ML ψ-function for the uncontaminated data. In the presence of contamination, the power curves obtained using the Huber ψ-function with tuning constant $k = 1.28$ is closer to that of the uncontaminated ML procedure (see Figures 3.3(b)-3.3(f)). This small distance between two curves means that even if there is 10% contamination in the data, the Huber procedure with tuning constant $k = 1.28$ is not affected by these contaminants. Thus the power curves obtained using the Huber ψ-function with $k = 1.28$ represents the majority of the data and the procedure is robust against some departures from the model assumptions.

Obviously for a contaminated case, the power of the Huber M-test for the

UT and RT is at least as large or larger than those of the ML-based tests for any λ_2 (see Figures 3.3(b) and 3.3(d)). The Huber M-test for the UT and RT enjoys the minimax property for any values of λ_1 and λ_2. For a range of values of λ_1 and λ_2, the Huber M-test for the PTT also has larger power compared to the ML-based PTT for the contaminated normal data (see Figure 3.3(f)). However, for some λ_1 and λ_2, it has less power. For a smaller value of λ_2, ($\lambda_2 < 3.2$) the Huber M-test for the PTT has higher power than that of the ML-based test, but for larger values of λ_2, the power of the Huber M-test for the PTT may be found to be smaller than the power of the ML-based PTT when λ_1 is smaller.

3.7.2 Power Comparison of UT, RT and PTT

The asymptotic power functions of the UT, RT and PTT are compared in this Section and are supported by the analytical results given in Section 3.6 for the three cases ((a), (b) and (c)) of c_i.

In Figure 3.4, the power functions for the UT, RT and PTT are plotted against λ_2 at two values of λ_1. The first set of regressors (when $\bar{c} > 0$) is used to plot Figures 3.4(a) and 3.4(d). As λ_2 grows larger, $\Pi^{RT}(0, \lambda_2)$ approaches 1. Hence, the RT is not a valid test because it does not satisfy the asymptotic level constraint. The $\Pi^{PTT}(0, \lambda_2)$, after an initial increase, drops and converges to the nominal size $\alpha = 0.05$ as λ_2 grows larger. Thus, the asymptotic size (with very small λ_1) of the PTT is close to α for small λ_2 and large λ_2, while for moderate values of λ_2 it is somehow larger than α but less than that of $\Pi^{RT}(0, \lambda_2)$. The $\Pi^{UT}(0, \lambda_2)$ is constant and does not depend on λ_2. The same pattern occurs in Figure 3.4(d) but the power functions are always significantly larger than α, in this case larger than 0.4. If one only considers the size of the test, the PTT is preferred to the RT, though the UT remains the best choice. Although the RT has the largest power among the tests, it is not a valid test. Thus, in terms of power, the PTT is preferred to the UT.

It is impossible to obtain a test that uniformly minimizes the size and maximizes the power at the same time. We are looking for a test that is a compromise between minimizing the size and maximizing the power (small probabilities of Type I and Type II errors). The RT is the worst choice for its largest size that reaches 1 as λ_2 grows larger, so it is not a valid test. On the contrary, the UT is the best choice for its smallest size but the worst choice for its smallest power. Both RT and UT uniformly minimize or maximize the size and power at the same time. The PTT has larger power than the UT for small and moderate values of λ_2 and it has significantly smaller size than that of the RT for moderate and large λ_2. Therefore, if our objective is to obtain a test that has better probabilities for both Type I and Type II errors, the PTT is suggested as the best option. The PTT is a compromise between minimizing the size and maximizing the power between the RT and UT.

The cases for $\bar{c} = 0$ and $\bar{c} < 0$ are also considered in this Section, though $\bar{c} > 0$ is more important than the other two because it is more realistic. Setting $\bar{c} = 0$ in Figures 3.4(b) and 3.4(e) implies all power functions remain the same regardless of the value of λ_2 and these constant power functions increase as λ_1 increases. Figures 3.4(c) and 3.4(f) illustrate the case when $\bar{c} < 0$. The graphs show that $\Pi^{RT} < \Pi^{PTT}$ for any λ_2 and $\Pi^{PTT} \leq \Pi^{UT}$ for any $\lambda_2 > \lambda_0$, where λ_0 is a small positive real number. The probability of a Type I error for all test functions is fairly small. The size and power of the RT is decreasing to 0 as λ_2 grows larger (see Figures 3.4(c) and 3.4(f)) suggests the RT as the best choice for size but the worst choice for power. Since $\Pi^{PTT}(2, \lambda_2) \geq \Pi^{RT}(2, \lambda_2)$ for all λ_2, the PTT is preferred over the RT . Also, $\Pi^{PTT}(2, \lambda_2) \geq \Pi^{UT}(2, \lambda_2)$ except for some moderate values of λ_2, but the difference is relatively small. From the examination of all the graphs in Figure 3.4, the PTT is suggested as the best choice when both probabilities of Type I and Type II errors are considered.

The relation between power functions and λ_1 is shown in Figure 3.5. All power functions are approaching 1 as λ_1 grows larger, regardless of the value of λ_2. This is because the probability of rejecting $H_0^{(1)} : \theta = 0$ increases as λ_1 increases. When $\bar{c} > 0$, the PTT is preferable than the UT for having smaller probability of a Type II error for all values of λ_1 while the RT is not a valid test for having a size that reaches 1 as λ_2 increases. When $\bar{c} < 0$, the PTT is preferable for its comparatively smaller probability of a Type II error than the other two tests. When $\bar{c} = 0$, all tests have the same probability of a Type II error regardless of the value of λ_1 (refer to the equation (3.6.2) for analytical results).

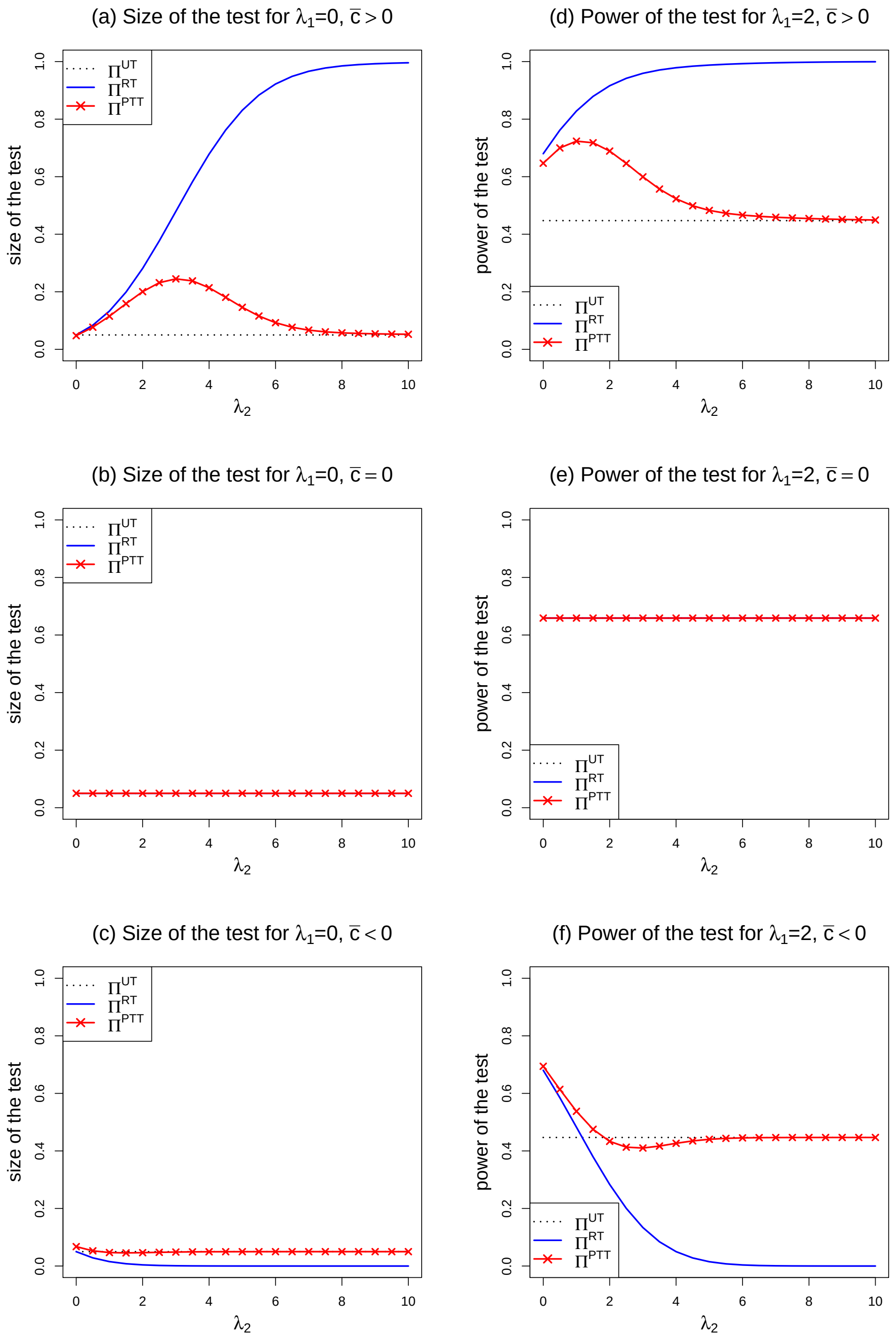


Figure 3.4: Graphs of power functions as a function of λ_2 for selected values of λ_1 and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$. Power functions using Huber score function with $k = 1.28$ in the presence of 10% contamination.

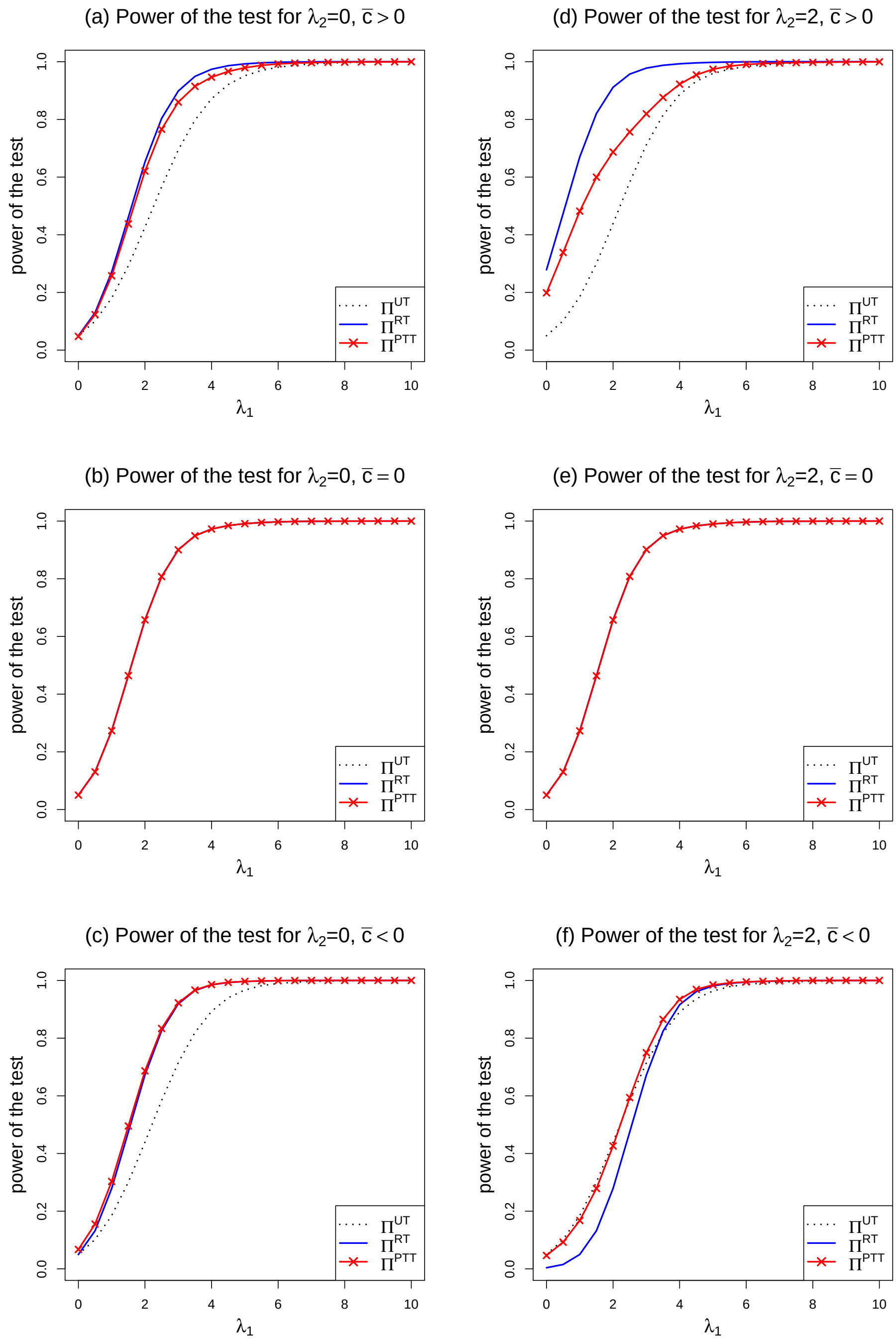


Figure 3.5: Graphs of power functions as a function of λ_1 for selected values of λ_2 and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$. Power functions using Huber score function with $k = 1.28$ in the presence of 10% contamination.

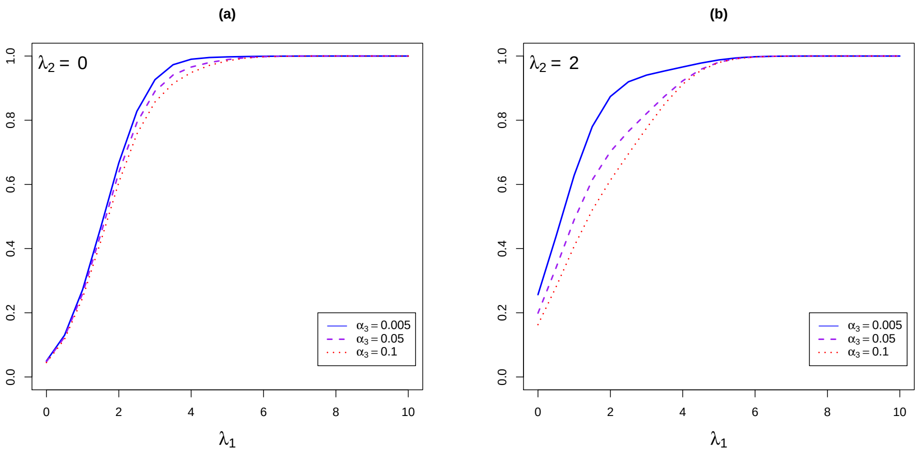


Figure 3.6: Graphs of power function $\Pi^{PTT}(\lambda_1, \lambda_2)$ for nominal sizes $\alpha_3 = 0.005$, 0.05 and 0.1. The $\bar{c} > 0$ and $\alpha_2 = \alpha_1 = \alpha = 0.05$ for all graphs.

3.7.3 Investigation on the PTT

This Section investigates the relationship between the power function of the PTT with its arguments, namely slope and nominal sizes α_1, α_2 and α_3 respectively.

Figure 3.6 illustrates the behavior of the power function of the PTT at three different values of the nominal size of the PT (α_3) that is 0.005, 0.05 and 0.1. The graphs show that a PTT with a smaller nominal size (significance level) has a greater power than that of a larger nominal size. The smaller nominal size however increases the probability of a Type I error as λ_2 moves away from zero. This is illustrated in Figure 3.6(b), $\Pi^{PTT}(\lambda_1, 2)$ at $\alpha_3 = 0.005$, 0.05 and 0.1, starting at different values before growing larger and converging to 1.

We usually assign a small nominal size (significance level) to the test, so the test will have a small probability of a Type I error. In the investigation, we concentrate on small nominal sizes, α_1, α_2 and α_3 with a view to achieve a small probability of a Type I error for the PTT. First, the relationship between the size of the PTT, that is, $\alpha^{PTT} = \Pi^{PTT}(0, \lambda_2)$ and the nominal size of the

PT, α_3 is studied. Figure 3.7 shows the graphs of $\Pi^{PTT}(0, \lambda_2)$ against α_3 for different values of λ_2 with $\alpha_1 = \alpha_2 = 0.05$ and $\bar{c} > 0$. For smaller values of λ_2, as α_3 increases, the size of the PTT decreases and reaches its minimum at the value of $\alpha_3 = \alpha_3'$ (say), before growing larger and converging to $\alpha = 0.05$. Let the value of α_3 be α_3'' when the size of the PTT is 0.05, while the value of α_3'' increases as λ_2 increases. As we consider larger values of λ_2, the size of the PTT decreases dramatically, then slowly converges (appears as flat in the graph) to α at some positive value α_3'''.

From Figure 3.7, we observe that for smaller values of λ_2 (≤ 0.6), the size of the PTT is reasonably small when the nominal size α_3 is small. However, for larger values of λ_2 (≥ 1), the size of the PTT is small when the nominal size α_3 is large and it is large when α_3 is small.

Equations (3.6.4) and (3.6.5) show that the size and power of the PTT is approaching the size and power of the RT as the nominal size α_3 is closer to 0, but is approaching the size and power of the UT as the nominal size α_3 is closer to 1. From equation (3.6.4), setting the nominal size of the PT, $\alpha_3 = 0$ implies the size and power of the PTT is entirely contributed by the size and power of the RT with none from the UT. The contribution of the size and power of the UT to the size and power of the PTT is not substantial when the nominal size α_3 is small.

Figure 3.8 shows the graphs of $\alpha^{PTT} = \Pi^{PTT}(0, \lambda_2)$ for $0 \leq \alpha_3 \leq 1$ at selected values of λ_2, α_1 and α_2 when $\bar{c} > 0$. From the graphs, the decrease in the contribution of the size of the RT reduces the size of the PTT as α_3 differs from zero. By contrast, setting the nominal size $\alpha_3 = 1$ causes the size of the PTT to be totally contributed by the size of the UT (see equation (3.6.5)). The contribution of the size of the RT is not significant when the nominal size α_3 is large. As the value of α_3 differs from 1, a lesser contribution from the size of the UT imposes a smaller size of the PTT. The size of the PTT decreases from both ends and the minimum of the size of the PTT is achieved at a particular value of α_3 (see Figure 3.8).

Further, analysis is carried out to investigate the dependence of the size of the PTT to the changes in the nominal sizes α_1, α_2 and α_3. From observation of Figures 3.8(a)-3.8(f), there is an increase in the percentage of α^{PTT} in [0,0.1] for α_3 in [0,0.2] when set at a smaller nominal size of α_2 for a larger value of λ_2. For example, there is 47.62% of $\Pi^{PTT}(0,1)$ in [0,0.10] for α_3 in [0,0.2] at a nominal size of $\alpha_2 = 0.05$ (see Figure 3.8(a)) but there is 100% of $\Pi^{PTT}(0,1)$ when $\alpha_2 = 0.03$ (see Figure 3.8(b)).

For some moderate values of λ_2, there is an increment in the percentage of $\Pi^{PTT}(0,\ \lambda_2)$ when we choose a smaller nominal size of α_2. However, only a small increment is observed for a larger value of λ_2. For example, there is no $\Pi^{PTT}(0,3)$ in [0,0.10] for α_3 in [0,0.2] when we set the nominal size $\alpha_2 = 0.05$ (see Figure 3.8(a)) but there is only 4.76% of $\Pi^{PTT}(0,3)$ when $\alpha_2 = 0.03$ (see Figure 3.8(b)). The small increment suggests setting a much smaller value of nominal size α_2 maybe necessary to achieve a small size of PTT with a small nominal size of PT α_3 for moderate values of the slope. However, this rule fails for a large value of λ_2.

We wish to have a small size of the PTT by setting small nominal sizes of α_1, α_2 and α_3. Figure 3.8 shows that this could not be achieved when λ_2 is large and α_3 is very small (close to zero) even if we set a very small value of α_2. For instance, there is less than 100% (i.e. 80.95%) of $\Pi^{PTT}(0,6)$ in [0,0.10] as α_3 in [0,0.2] (see Figures 3.8(a) and 3.8(b)) for both nominal sizes $\alpha_2 = 0.03$ and $\alpha_2 = 0.05$. The percentage does not reach 100% eventhough $0 < \alpha_2 < 0.03$ is chosen.

Since the size of the PTT behaves like the size of the RT when the nominal size α_3 is small, the null hypothesis $H_0^{(1)} : \theta = 0$ is rejected more often for a small nominal size of α_3 when λ_2 is large because the nominal size α_2 is smaller than the actual size of the RT. The null hypothesis $H_0^{(1)} : \theta = 0$ should not be rejected if the true value of $\theta = 0$. In this case however, the possibility of

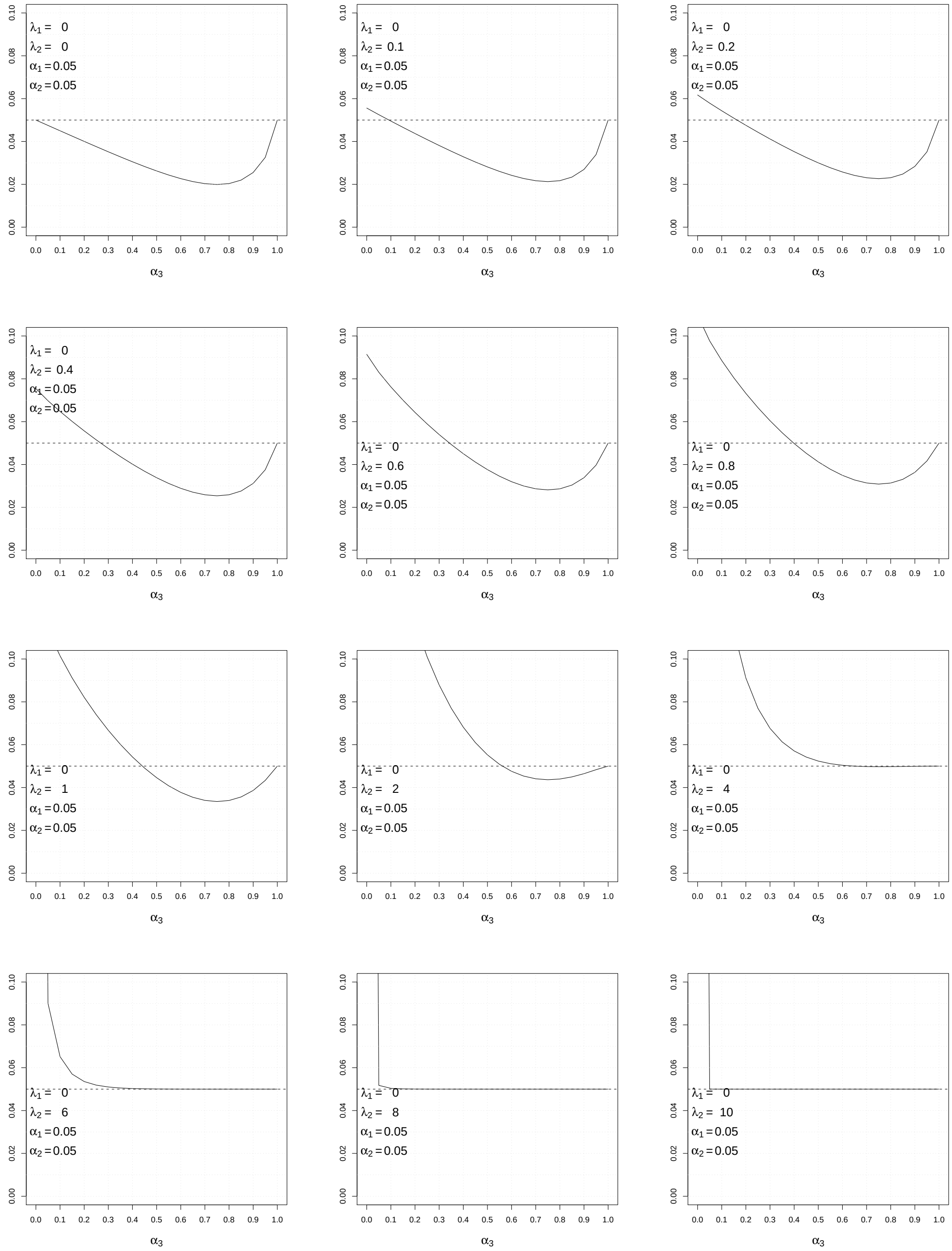


Figure 3.7: Graphs of size of the PTT $\alpha^{PTT} = \Pi^{PTT}(0, \lambda_2)$ as α_3 and λ_2 increasing when $\bar{c} > 0$ and $\alpha_1 = \alpha_2 = 0.05$ for all graphs.

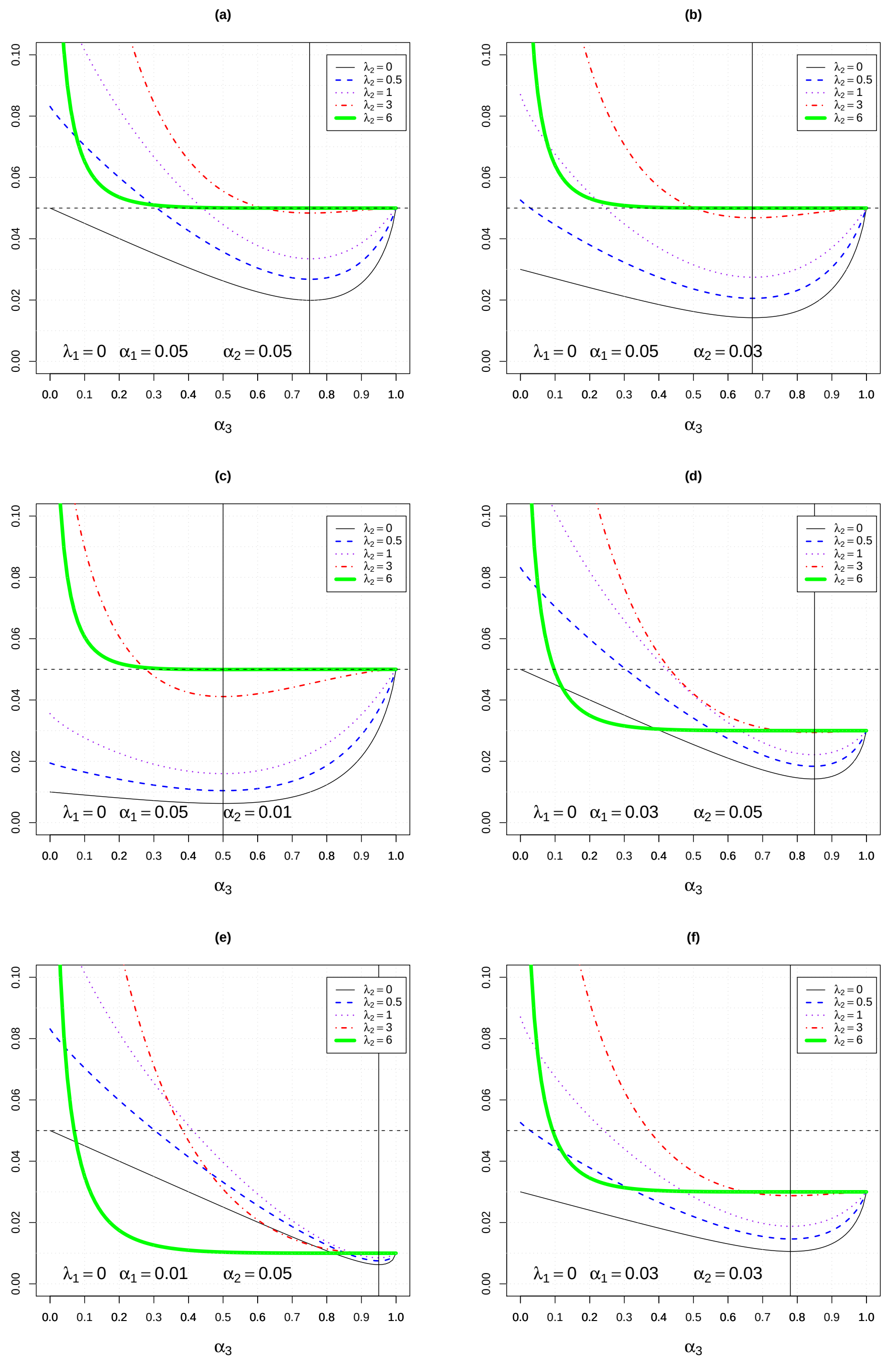


Figure 3.8: Graphs of the size of the PTT for increasing α_3, selected at different values of nominal sizes of α_1 and α_2 with $\bar{c} > 0$. The intersection with the vertical line represents the minimum.

rejection is large when λ_2 differs much from 0 because $\beta = 0$ is assumed in the test statistic T_n^{RT}. This is the reason why a very small nominal size of the PT α_3 (close to zero) implies a very large size of the PTT when λ_2 is large.

Table 3.1 shows the size of the PTT as a function of the nominal size of the PT (α_3) for selected values of λ_2 and α_2 with $\alpha_1 = 0.05$ and $\bar{c} > 0$. The values of α_1, α_2 and α_3 for the size of the PTT near point 0.05 when $\lambda_2 = 0.5$ and 1, and near point 0.10 with $\lambda_2 = 3$ and 6 are given in the table. The table enables us to observe the changes in the values of the nominal size of the PT (α_3) as the nominal size α_2 changes and the significance level of the PTT is around the same value. We wish to have a small nominal size of the PT α_3 that allows us to get a 5 or 10% significance level of the PTT. Within the table, this is achieved by selecting a smaller nominal size of α_2 for moderate and small values of λ_2. When $\lambda_2 = 3$ (moderate value), selecting nominal size α_2 as small as 0.01, we have as much as 8% of the nominal size of the PT α_3 to get below a 10% significance level of the PTT (see Table 3.1, row:1, col:7-9). In column 1-3 of the table, for $\lambda_2 = 0.5$ (small), an approximately 5% level of significance of the PTT is obtained by setting a nominal size of $\alpha_2 = 0.05$ and a nominal size of the PT $\alpha_3 = 0.3$ or by setting both nominal sizes of α_2 and α_3 equal to 0.03 but the latter with smaller nominal sizes of α_2 and α_3 is more preferable. For a larger value of the slope, as the nominal size α_3 is closer to 0, the size of the PTT is growing too large. When $\lambda_2 = 6$ (large), to obtain at most a 10% of significance level of the PTT, the least nominal size for the PT α_3 that we should set is 5% (see Table 3.1, row:3, col:10-12) when the nominal size α_2 is set from 0.05 to 0.10.

Table 3.1: Size of the PTT (α^{PTT}) as a function of nominal size of the PT, (α_3) at selected values of α_2 and λ_2 with $\alpha_1 = 0.05$.

$\lambda_2 = 0.5$			$\lambda_2 = 1$			$\lambda_2 = 3$			$\lambda_2 = 6$		
α_2	α_3	α^{PTT}	α_2	α_3	α^{PTT}	α_2	α_3	α^{PTT}	α_2	α_3	α^{PTT}
0.03	0.03	0.0498	0.01	0.00	0.0355	0.01	0.08	0.0994	0.03	0.04	0.0983
	0.02	0.0507		0.01	0.0343		0.07	0.1052		0.03	0.1145
0.04	0.19	0.0499	0.02	0.00	0.0623	0.02	0.15	0.0987	0.04	0.05	0.0891
	0.18	0.0508		0.01	0.0605		0.14	0.1029		0.04	0.1002
0.05	0.31	0.0497	0.03	0.00	0.0870	0.03	0.20	0.0965	0.05	0.05	0.0901
	0.30	0.0507		0.01	0.0839		0.19	0.1005		0.04	0.1014
0.06	0.34	0.0508	0.04	0.02	0.1026	0.04	0.22	0.1007	0.06	0.04	0.1024
	0.04	0.0499		0.03	0.0999		0.23	0.0972		0.05	0.0909
0.07	0.48	0.0491	0.05	0.11	0.0993	0.05	0.25	0.0987	0.10	0.05	0.0093
	0.47	0.0500		0.10	0.1015		0.24	0.1021		0.04	0.1045

The α^{PTT} is the actual achievable significance level and α_3 is the nominal PT significance level.

3.8 Testing Intercept and Slope at Any Value

Consider the simple regression model in Section 3.1; we want to test the intercept parameter θ, at a fixed value, under various conditions on the slope parameter β. Again, we have three conditions on the slope,

(i) *Unspecified* β : Denote the unrestricted test (UT) by the test function $\phi_n^{\star UT}$ for testing $H_0^{\star(1)} : \theta = \theta_0$ against $H_A^{\star(1)} : \theta > \theta_0$ when β is unspecified;

(ii) *Specified* β : Denote the restricted test (RT) by the test function $\phi_n^{\star RT}$ for testing $H_0^{\star(1)} : \theta = \theta_0$ against $H_A^{\star(1)} : \theta > \theta_0$ when $\beta = \beta_0$ (fixed and specified); and

(iii) *Uncertain suspected* β : Denote the pre-test test (PTT) by the test function $\phi_n^{\star PTT}$ for testing $H_0^{\star(1)} : \theta = \theta_0$ against $H_A^{\star(1)} : \theta > \theta_0$ (when it is suspected that $\beta = \beta_0$, but not sure) following a pre-test on the slope, $H_0^{\star(2)} : \beta = \beta_0$ against $H_A^{\star(2)} : \beta > \beta_0$. The last test on β, $H_0^{\star(2)} : \beta = \beta_0$, with the test function $\phi_n^{\star PT}$, is a pre-test (PT) essential for the PTT on θ.

Hence, significance testing on the intercept and slope, considered in Sections 3.1-3.7, is a special case of testing any arbitrary values discussed in this Section.

Let $\check{\beta}$ be the constrained M-estimator of β when $\theta = \theta_0$, that is, $\check{\beta}$ is the solution of $M_{n_2}(\theta_0, b) = 0$ and it may be conveniently expressed as

$$\check{\beta} = [\sup\{b : M_{n_2}(\theta_0, b) > 0\} + \inf\{b : M_{n_2}(\theta_0, b) < 0\}]/2. \tag{3.8.1}$$

Similarly, let $\check{\theta}$ be the constrained M-estimator of θ when $\beta = \beta_0$, that is, $\check{\theta}$ is the solution of $M_{n_1}(a, \beta_0) = 0$ and is conveniently be expressed as

$$\check{\theta} = [\sup\{a : M_{n_1}(a, \beta_0) > 0\} + \inf\{a : M_{n_1}(a, \beta_0) < 0\}]/2. \tag{3.8.2}$$

3.8.1 Asymptotic Distributions of Test Statistics

Theorem 3.8.1 *Given the results in (i)-(iii) in Appendix B.1, as $n \to \infty$,*

(a) under $H_0^{\star(2)} : \beta = \beta_0$,

$$n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2}), \tag{3.8.3}$$

(b) under $H_0^{\star(1)} : \theta = \theta_0$,

$$n^{-\frac{1}{2}} M_{n_1}(\theta_0, \check{\beta}) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2}/\{C^{\star 2} + \bar{c}^2\}), \tag{3.8.4}$$

(c) under $H_0^{\star} : \theta = \theta_0, \beta = \beta_0$,

$$n^{-\frac{1}{2}} \begin{pmatrix} M_{n_1}(\theta_0, \beta_0) \\ \\ M_{n_2}(\theta_0, \beta_0) \end{pmatrix} \xrightarrow{d} N_2 \left[\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \right], \tag{3.8.5}$$

where $N_2(\cdot\ ,\cdot\ )$ represents a bivariate normal distribution with appropriate parameters and $\sigma_0^2 = \int_{-\infty}^{\infty} \psi^2(x/S) dF(x)$.

The proof of this Theorem is written in Appendix B.1. These three asymptotic distributions results of Theorem (3.8.1) are useful to construct UT, RT and PTT in the next Section.

3.8.2 The Proposed Tests

3.8.2.1 The Unrestricted Test (UT)

If β is unspecified, let $\phi_n^{\star UT}$ denote the test function to test $H_0^{\star(1)} : \theta = \theta_0$ against $H_A^{\star(1)} : \theta > \theta_0$. We consider the test statistic $T_n^{\star UT} = M_{n_1}(\theta_0, \check{\beta})$, where $\check{\beta}$ is a constrained M-estimate defined in equation (3.8.1). Under $H_0^{\star(1)}$, it follows from equation (3.8.4) that

$$n^{-\frac{1}{2}} T_n^{\star UT} / \sqrt{C_n^{(1)} {S_n^{\star(1)}}^2 / n} = n^{-\frac{1}{2}} M_{n_1}(\theta_0, \check{\beta}) / \sqrt{C_n^{(1)} {S_n^{\star(1)}}^2 / n} \xrightarrow{d} N(0, 1) \tag{3.8.6}$$

as $n \to \infty$, with $C_n^{(1)} = n C_n^{\star 2} / (C_n^{\star 2} + n\bar{c}_n^2)$ and ${S_n^{\star(1)}}^2 = n^{-1} \sum \psi^2 \left(\frac{x_i - \theta_0 - \check{\beta} c_i}{S_n} \right)$.

3.8.2.2 The Restricted Test (RT)

If $\beta = \beta_0$, let $\phi_n^{\star RT}$ be the test function for testing $H_0^{\star(1)} : \theta = \theta_0$ against $H_A^{\star(1)} : \theta > \theta_0$. The proposed test statistic is $T_n^{\star RT} = M_{n_1}(\theta_0, \beta_0)$. Note that for large n, it follows from equation (3.8.5) that

$$n^{-\frac{1}{2}} T_n^{\star RT} / \sqrt{S_n^{\star(2)^2}} = n^{-\frac{1}{2}} M_{n_1}(\theta_0, \beta_0) / \sqrt{S_n^{\star(2)^2}} \xrightarrow{d} N(0, 1) \tag{3.8.7}$$

under $H_0^\star : \theta = \theta_0, \beta = \beta_0$, where $S_n^{\star(2)^2} = n^{-1} \sum \psi^2 \left(\frac{x_i - \theta_0 - \beta_0 c_i}{S_n} \right)$.

3.8.2.3 The Pre-test (PT)

For the pre-test (PT) on the slope, let $\phi_n^{\star PT}$ be the test function to test $H_0^{\star(2)} : \beta = \beta_0$ against $H_A^{\star(2)} : \beta > \beta_0$. The proposed test statistic is $T_n^{\star PT} = M_{n_2}(\check{\theta}, \beta_0)$, where $\check{\theta}$ is the constrained M-estimate defined in equation (3.8.2). It follows from equation (3.8.3) that under $H_0^{\star(2)}$, as $n \to \infty$,

$$n^{-\frac{1}{2}} T_n^{\star PT} / \sqrt{C_n^{(3)} S_n^{\star(3)^2} / n} = n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) / \sqrt{C_n^{(3)} S_n^{\star(3)^2} / n} \xrightarrow{d} N(0, 1), \tag{3.8.8}$$

where $C_n^{(3)} = \sum c_i^2 - n\bar{c}_n^2 = C_n^{\star 2}$ and $S_n^{\star(3)^2} = n^{-1} \sum \psi^2 \left(\frac{x_i - \check{\theta} - \beta_0 c_i}{S_n} \right)$. The consistency of $S_n^{\star(1)^2}$, $S_n^{\star(2)^2}$ and $S_n^{\star(3)^2}$ as estimators of σ_0^2 may follow from the law of large numbers (Jurečková and Sen, 1981).

3.8.2.4 The Pre-test Test (PTT)

Now, let $\phi_n^{\star PTT}$ be the test function to test $H_0^{\star(1)} : \theta = \theta_0$ against $H_A^{\star(1)} : \theta > \theta_0$ following a pre-test on β. We write the test function for the PTT as

$$\phi_n^{\star PTT} = I \left[(T_n^{\star PT} \le \iota_{n,\alpha_3}^{PT}, T_n^{\star RT} > \iota_{n,\alpha_2}^{RT}) \text{ or } (T_n^{\star PT} > \iota_{n,\alpha_3}^{PT}, T_n^{\star UT} > \iota_{n,\alpha_1}^{UT}) \right], \tag{3.8.9}$$

where ι_{n,α_j}^h is the critical value of $T_n^{\star h}$ at the α_j, $j = 1, 2, 3$ level of significance and h is any of the UT, RT and PT. Then we define the power of the test from the test function $\phi_n^{\star PTT}$.

3.8.3 Properties of Tests under Local Alternatives

The contiguity concept is utilized to find the asymptotic joint distributions of statistics $n^{-\frac{1}{2}}[\,T_n^{\star RT}, T_n^{\star PT}]$ and $n^{-\frac{1}{2}}[\,T_n^{\star UT}, T_n^{\star PT}]$ under $K_n^\star$ given below.

Theorem 3.8.2 *Let $\{K_n^\star\}$ be a sequence of alternative hypotheses, where*

$$K_n^\star : (\theta, \beta) = (\theta_0 + n^{-\frac{1}{2}}\lambda_1, \beta_0 + n^{-\frac{1}{2}}\lambda_2), \tag{3.8.10}$$

with $\lambda_1 = \sqrt{n}(\theta - \theta_0)$ and $\lambda_2 = \sqrt{n}(\beta - \beta_0)$ being fixed real numbers. Under $\{K_n^\star\}$, for large sample,

$$\text{(i) } n^{-1/2}\begin{bmatrix} T_n^{\star UT} \\ T_n^{\star PT} \end{bmatrix} \xrightarrow{d} N_2\left[\begin{pmatrix} \frac{\gamma\lambda_1 C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ \gamma\lambda_2 C^{\star 2} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2} & -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} & C^{\star 2} \end{pmatrix}\right], \tag{3.8.11}$$

$$\text{(ii) } n^{-1/2}\begin{bmatrix} T_n^{\star RT} \\ T_n^{\star PT} \end{bmatrix} \xrightarrow{d} N_2\left[\begin{pmatrix} \gamma(\lambda_1 + \lambda_2\bar{c}) \\ \gamma\lambda_2 C^{\star 2} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{pmatrix}\right]. \tag{3.8.12}$$

See Appendix B.1 for the proof of Theorem 3.8.2.

The joint distribution of $n^{-\frac{1}{2}}[T_n^{\star RT}, T_n^{\star PT}]$ and $n^{-\frac{1}{2}}[T_n^{\star UT}, T_n^{\star PT}]$ under $K_n^\star$ is the same as the joint distribution of $n^{-\frac{1}{2}}[T_n{}^{RT}, T_n{}^{PT}]$ and $n^{-\frac{1}{2}}[T_n{}^{UT}, T_n{}^{PT}]$ under K_n. Thus, the asymptotic power functions for $\phi_n^{\star UT}$, $\phi_n^{\star RT}$ and $\phi_n^{\star PTT}$ are obtained in the same manner as those of ϕ_n^{UT}, ϕ_n^{RT} and ϕ_n^{PTT} (defined in Section 3.5).

We find the asymptotic power functions for the UT, RT and PTT under $\{K_n^\star\}$ are respectively

$$\begin{aligned}
\Pi^{\star UT}(\lambda_1, \lambda_2) &= 1 - \Phi(\tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}\,/\sigma_0) \text{ and } && (3.8.13)\\
\Pi^{\star RT}(\lambda_1, \lambda_2) &= 1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0). && (3.8.14)\\
\Pi^{\star PTT}(\lambda_1, \lambda_2) &= \Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0)[1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0)] \; + \\
&\quad d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0; \rho), \\
& && (3.8.15)
\end{aligned}$$

where $\rho = -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}$. Note that the asymptotic power functions, $\Pi^{\star UT}$ $(\lambda_1,$ $\lambda_2)$, $\Pi^{\star RT}$ $(\lambda_1,$ $\lambda_2)$ and $\Pi^{\star PTT}(\lambda_1, \lambda_2)$ are respectively equal to the asymptotic power functions, $\Pi^{UT}(\lambda_1, \lambda_2)$, $\Pi^{RT}(\lambda_1, \lambda_2)$ and $\Pi^{PTT}(\lambda_1, \lambda_2)$ when $\theta_0 = 0$ and $\beta_0 = 0$. This makes sense because testing the significance of θ and β is a special case of testing θ and β at any arbitrary values.

3.8.4 Analytical Comparison

This Section provides an analytical comparison of asymptotic relative efficiency for power functions of the UT, RT and PTT. Define the relative efficiency for T_1 with respect to T_2 as

$$RE(\Pi^{\star T_1} : \Pi^{\star T_2}) = \Pi^{\star T_1} \div \Pi^{\star T_2}, \qquad (3.8.1)$$

where T_1 and T_2 are any of the UT, RT and PTT.

In the same manner as in Section (3.6), we arrive at the following results:

For $\bar{c} > 0$, $\alpha_1 = \alpha_2 = \alpha$ and $\lambda_2 \geq 0$, it is easy to show that $\lambda_1 + \lambda_2\bar{c} \geq \lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}$. Thus,

(a) $\Pi^{\star RT}(\lambda_1, \lambda_2) \geq \Pi^{\star PTT}(\lambda_1, \lambda_2)$, (b) $\Pi^{\star RT}(\lambda_1, \lambda_2) \geq \Pi^{\star UT}(\lambda_1, \lambda_2)$,

and

(c) $\Pi^{\star UT}(\lambda_1, \lambda_2) - \Pi^{\star PTT}(\lambda_1, \lambda_2) \lesseqqgtr 0$ if $B^\star \lesseqqgtr |A^\star|$, where

$$\begin{aligned} A^\star &= \Phi(\tau_\alpha - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0) - \Phi(\tau_\alpha - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0) \text{ and} \\ B^\star &= d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_\alpha - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0; 0) - \\ &\quad d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_\alpha - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0; -\bar{c}/\rho). \end{aligned}$$

Thus, it is straightforward to see that

(d) $RE(\Pi^{\star PTT} : \Pi^{\star RT}) \leq 1$, (e) $RE(\Pi^{\star UT} : \Pi^{\star RT}) \leq 1$,

(f) $RE(\Pi^{\star UT} : \Pi^{\star RT}) \gtreqqless RE(\Pi^{\star PTT} : \Pi^{\star RT})$ if $B^\star \lesseqqgtr |A^\star|$,

(g) $RE(\Pi^{\star PTT} : \Pi^{\star UT}) \leq RE(\Pi^{\star RT} : \Pi^{\star UT})$ and

(h) $RE(\Pi^{\star PTT} : \Pi^{\star UT}) \gtreqqless 1$ if $B^\star \lesseqqgtr |A^\star|$.

3.8.5 Computational Comparison

The comparison of asymptotic relative efficiency in the presence of contaminations for power functions of the UT, RT and PTT is discussed in this Section. The simulated data generated using the Monte Carlo method given in Section 3.7 is used.

In Section 3.8.3, when $\lambda_1 = 0$, then $\alpha^{UT} = \Pi^{\star UT}(\theta = \theta_0)$ is the size of the test. The same applies to the RT and PTT test functions. We define relative efficiency with respect to size as a ratio of size of two tests. As for λ_1 other than 0, a ratio of the power of two tests is the relative efficiency of one test relative to the other.

Note that λ_1 is a function of $\theta - \theta_0$ while λ_2 is a function of $\beta - \beta_0$. Since we are concerned about testing $H_0^{\star(1)} : \theta = \theta_0$, the relative efficiency with respect to the UT and RT is plotted against λ_1 in Figure 3.9 in the presence of 10% contaminations in the responses using the Huber ψ-function with $k = 1.28$. As λ_1 grows larger, it is suspected that the probability of rejecting $H_0^{\star(1)}$ when it

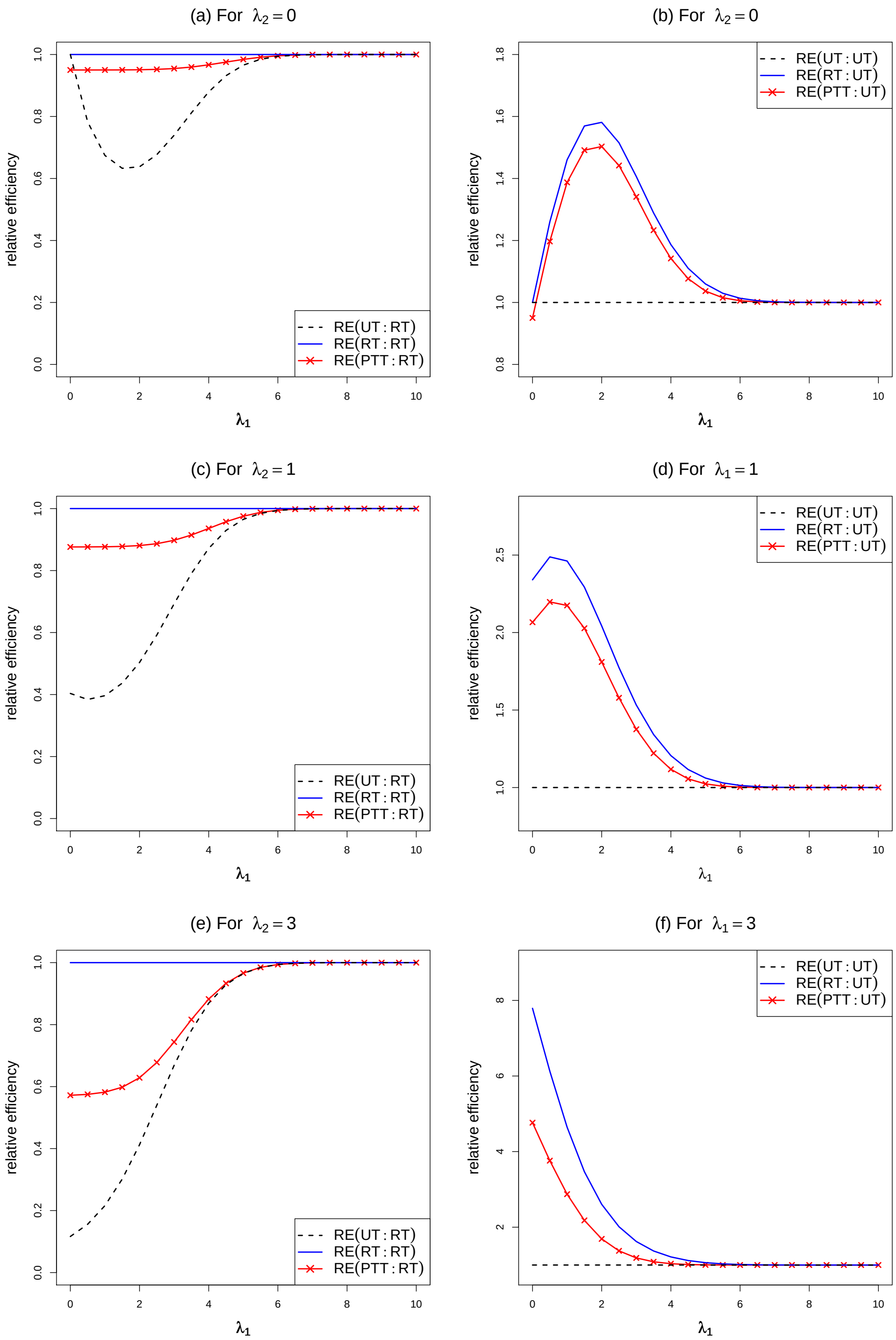


Figure 3.9: Graphs of relative efficiency functions in terms of the power of the tests as a function of λ_1 for selected values of λ_2, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$ and $\bar{c} > 0$. $RE(T_1 : T_2)$ represents $RE(\Pi^{\star T_1} : \Pi^{\star T_2})$ where T_1 and T_2 are any of the UT, RT and PTT.

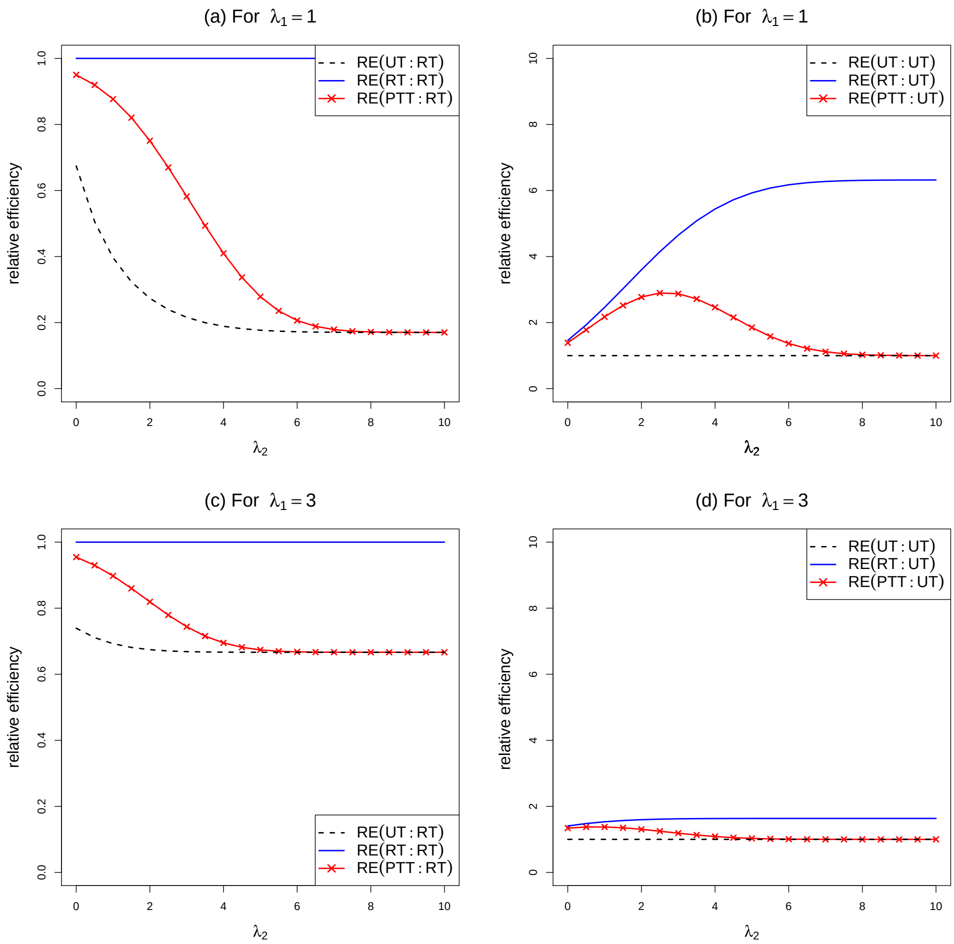


Figure 3.10: Graphs of relative efficiency functions in terms of the power of the tests as a function of λ_2 for selected values of λ_1, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$ and $\bar{c} > 0$. $RE(T_1 : T_2)$ represents $RE(\Pi^{\star T_1} : \Pi^{\star T_2})$ where T_1 and T_2 are any of the UT, RT and PTT.

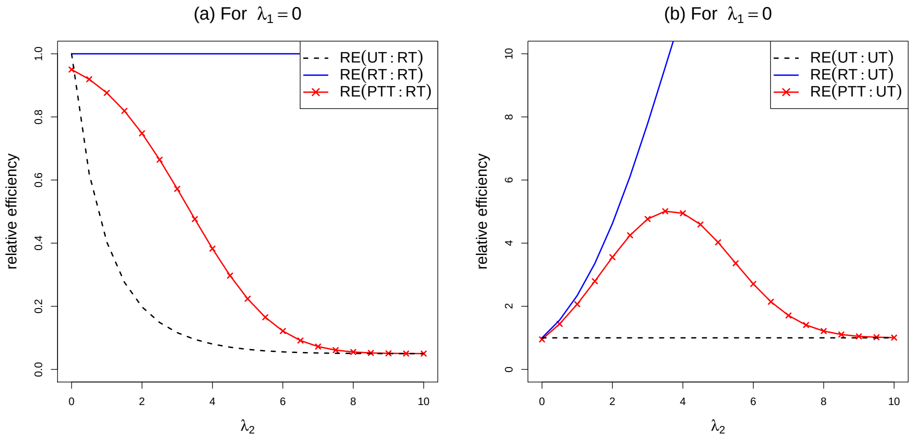


Figure 3.11: Graphs of relative efficiency functions in terms of the size of the tests as a function of λ_2, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$ and $\bar{c} > 0$. $RE(T_1 : T_2)$ represents $RE(\Pi^{\star T_1} : \Pi^{\star T_2})$ where T_1 and T_2 are any of the UT, RT and PTT.

is true (probability of a Type I error) grows large and the probability of failing to reject $H_0^{\star(1)}$ when it is false (probability of a Type II error) decreases for all tests. In Figure 3.9, the efficiency of the UT, RT and PTT are almost the same for large λ_1. For smaller λ_1, the RT is the most efficient in terms of the power because the relative efficiency of the UT and the PTT with respect to the RT is at most 1 (see (d) and (e) of Section 3.8.4). We also find in Figure 3.9 that the PTT is more efficient in terms of the power than the UT when λ_1 is small. Although the PTT has the smallest power when λ_1 and λ_2 are very close to 0 (see Figure 3.9, when $\lambda_2 = 0$), the difference between the true intercept and the suspected value is very small.

The relative efficiency with respect to the UT and RT is plotted against λ_2 in Figure 3.10 and Figure 3.11 because the three tests are defined according to the knowledge about β. The relative efficiency in terms of the power for the UT and PTT is almost the same when λ_2 is large, while the RT always has the largest power (see Figure 3.10). It is also observed that for small λ_2, the PTT is more efficient in terms of the power than the UT.

The relative efficiency for the size of the tests is plotted in Figure 3.11. For small λ_2, the UT is more efficient in terms of the size than the PTT while the RT has the largest size regardless of the values of λ_1 or λ_2 and this is supported by the analytical results in (a) and (b) of Section 3.8.4. Both UT and PTT have almost the same efficiency when λ_2 is large. The PTT has the smallest relative efficiency in terms of the size when λ_1 and λ_2 are very close to 0 (see Figure 3.11).

3.9 Application to Data

This example relates to the study of the relationship between the distance by road and the linear distance. Twenty different pairs of points of the values of the two variables in Sheffield is reported by Gilchrist (1984) (c.f Abraham and Ledolter, 2006, p.63). To check the robustness of the test, one data point $(5.0, 6.5)$ is changed to $(5.0, 46.5)$ to create the modified data set. The one sided t-test is applied to the original and modified data sets and the summary statistics are presented in Table 3.2. The scatter plot and the fitted regression lines for the original and modified data sets are given in Figure 3.12. For both original and modified data sets, the slope is significantly different from zero. For the original data set, the intercept is not significantly different from zero. However, the intercept is significantly different from zero for the modified data set.

In this Section, the main objective is to test the significance of the intercept parameter when it is suspected that the slope parameter may be zero. The summary statistics for the UT, RT and PT on the intercept are given in Table 3.3 for the original data. The intercept is not significantly different from zero from the UT whereas the intercept is significantly different from zero under the RT. The PT (on the slope) indicates a significant linear relationship between the two variables. Obviously the RT (on the intercept) is not an appropriate

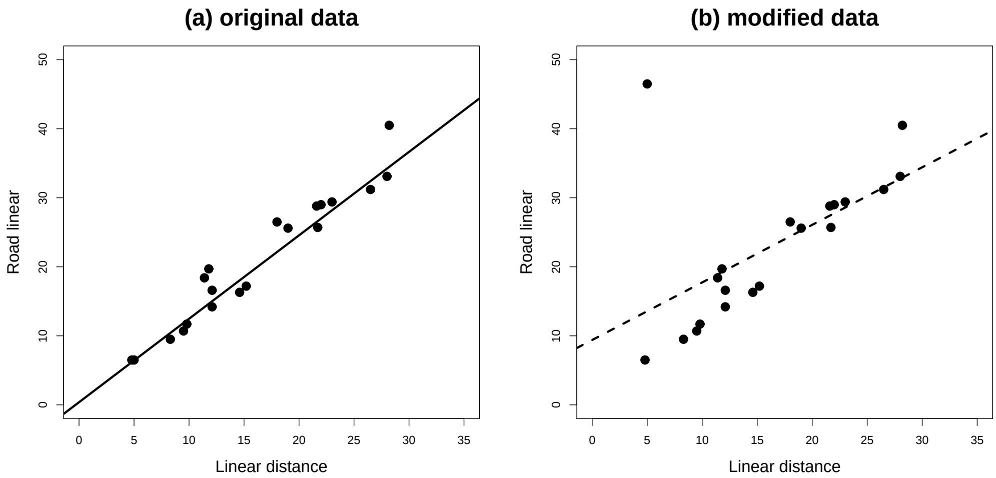


Figure 3.12: Graphs of the fitted regression line for the original and modified data

Table 3.2: Summary statistics of a one sided t-test on the distance data.

	Original data			Modified data		
	coefficient	t-statistic	p-value	coefficient	t-statistic	p-value
Intercept	0.379	0.282	0.3905	9.400	1.922	0.0355
Slope	1.209	16.665	0.0000	0.834	3.009	0.0040

test because the hypothesis of suspected zero slope is rejected. In the analysis, the intercept is significantly different from zero when using the RT. The UT is more appropriate than the RT since the UT does not depend on the prior information. In general, if the prior information is available, the uncertainty in the value of the slope is removed using the PT before testing on the intercept.

The sensitivity of a robust test using the Huber ψ-function to an aberrant observation is studied by introducing a modification in one of the data points. For the modified data, the original data point $(5.0, 6.5)$ is replaced by a new

Table 3.3: Summary statistics of a one sided test for the UT, RT and PT using the ML ψ-function on the original data.

	UT	RT	PT
Null hypothesis	$H_0^\star : \theta = 0$	$H_0^\star : \theta = 0$	$H_0^{(1)} : \beta = 0$
Model under null	$X_i = \beta c_i + e_i$	$X_i = e_i$	$X_i = \theta + e_i$
Coefficient	$\tilde{\beta} = 1.289$	None	$\tilde{\theta} = 20.855$
z-statistic	$\frac{T_n^{UT}}{\sqrt{C_n^{(1)} S_n^{(1)2}}} = 0.2967$	$\frac{T_n^{RT}}{\sqrt{S_n^{(2)2}}} = 4.0795$	$\frac{T_n^{PT}}{\sqrt{C_n^{(3)} S_n^{(3)2}}} = 2.19 \times 10^{16}$
p-value	0.3834	2.26×10^{-5}	0

Table 3.4: Summary statistics of a one sided test for the UT, RT and PTT using the ML and Huber ψ-functions on the modified data.

(a) ML ψ-function

	UT	RT	PT
Null hypothesis	$H_0^\star : \theta = 0$	$H_0^\star : \theta = 0$	$H_0^{(1)} : \beta = 0$
Model under null	$X_i' = \beta c_i + e_i$	$X_i' = e_i$	$X_i' = \theta + e_i$
Coefficient	$\tilde{\beta} = 1.321$	None	$\tilde{\theta} = 22.855$
z-statistic	$\frac{T_n^{UT}}{\sqrt{C_n^{(1)} S_n^{(1)2}}} = 1.8452$	$\frac{T_n^{RT}}{\sqrt{S_n^{(2)2}}} = 4.0764$	$\frac{T_n^{PT}}{\sqrt{C_n^{(3)} S_n^{(3)2}}} = 1.16 \times 10^{16}$
p-value	0.0325	2.28×10^{-5}	0

(b) Huber ψ-function

	UT	RT	PT
Null hypothesis	$H_0^\star : \theta = 0$	$H_0^\star : \theta = 0$	$H_0^{(1)} : \beta = 0$
Model under null	$X_i' = \beta c_i + e_i$	$X_i' = e_i$	$X_i' = \theta + e_i$
Coefficient	$\tilde{\beta} = 1.289$	None	$\tilde{\theta} = 22.208$
z-statistic	$\frac{T_n^{UT}}{\sqrt{C_n^{(1)} S_n^{(1)2}}} = 0.9394$	$\frac{T_n^{RT}}{\sqrt{S_n^{(2)2}}} = 4.4335$	$\frac{T_n^{PT}}{\sqrt{C_n^{(3)} S_n^{(3)2}}} = 3.17 \times 10^{16}$
p-value	0.1738	4.64×10^{-6}	7.61×10^{-4}

(arbitrary) data point $(5.0, 46.5)$. This replacement causes a significant change in the values of the coefficients and the outcomes of the t-test. The summary statistics for the UT, RT and PT using both the ML and Huber ψ-functions for the modified data are displayed in Table 3.4. It is found that the UT using the Huber ψ-function is not much affected by the aberrant point, compared to that of the ML ψ- function. From the UT based on ML ψ- function, the intercept is significantly different from zero. However, it is not significantly different from zero under the UT that is based on the Huber ψ-function. The outcomes for the other two tests for the modified data are not much different from those of the original data.

3.10 Discussion and Conclusion

The UT, RT and PTT, defined using the M-test for testing the intercept θ under three different scenarios of the slope β, are provided in this Chapter. The asymptotic power functions of the tests are derived by using the results from the asymptotic sampling distribution of the test statistics. Under a sequence of local alternative hypotheses, the sampling distributions when the sample size is large for the UT, RT and PT follow a normal distribution with appropriate mean and variance. However, that of the PTT is a bivariate normal distribution. There is a correlation between the UT and PT but there is no such correlation between the RT and PT.

In the estimation regime, it is well known that the RE has the smallest MSE if distance parameter (a function of $\beta - \beta_0$) is 0 or close to 0, but its MSE is unbounded for larger values of the distance parameter. The UE has a constant MSE that does not depend on the distance parameter. The PTE has a smaller MSE than that of the RE for moderate and larger values of the distance parameter. The PTE has a smaller MSE than that of the UE if the value of the distance parameter is close or equal to 0. In the testing context, the power functions of the UT, RT and PTT demonstrate a similar behavior as the MSE of the UE, RE and PTE.

For a set of realistic values of the regressor, with a mean value larger than 0, the size of the RT is small when $\beta = \beta_0$ or λ_2 is close to 0, but the size grows large and converges to 1 for larger values of λ_2. Hence, the RT does not satisfy the asymptotic level constraint. Therefore it is not a valid test. The UT has a constant size regardless of the value of λ_2. The PTT has a smaller size than that of the RT when λ_2 is 0 and very close to 0, and significantly smaller than that of the RT for moderate and large values of λ_2. The PTT has a smaller size than the UT when λ_2 is 0 or very close to 0.

Usually, the power of a test is used to make comparisons between different statistical tests of the same hypothesis while the level of significance α is set as the size of the test. However, in this study, the (actual) size of the RT is growing larger and differs from the nominal significance level α as λ_2 grows larger. The size of a test also depends on the hypothesis we wish to test. If the significance of the intercept and slope is tested using the RT in a study, the size of the RT is equal to the level of significance α. However, in this study, we wish to test the significance of the intercept for the ultimate test using the UT, RT and PTT. Since the UT, RT and PTT are defined based on the knowledge of the slope, the performance of the three tests varies with respect to the value of λ_2. Testing the significance of the intercept causes the actual size of the RT differs from its level of significance α as λ_2 grows large. Since the UT does not depend on the slope parameter, its size is equal to the level of significance α. The PTT is a choice between the UT and the RT, so, its size also differs from the level of significance α. However, the difference between the actual size and the nominal size for the PTT is far less than that of the RT as λ_2 grows larger.

Again for a set of realistic values of the regressor, with a mean larger than 0, although the RT is the best choice for having the largest power, the RT is not a valid test because its size is too large and does not satisfy the asymptotic level constraint. The size of the UT is constant regardless of the value of λ_2. The UT is the best choice for having the smallest size but the worst choice for having the lowest power. The PTT has smaller size than the RT for moderate and larger values of λ_2 and has larger power than the UT for smaller and moderate values of λ_2. Therefore, the power function of the PTT is found to behave similar to the MSE of the PTE, in the sense that although it is not uniformly the best statistical test with the smallest size and the largest power, it does protect from the risk of too large size and power being too small. Thus, the power function of the PTT is a compromise between that of the UT and RT. In the face of uncertainty on the value of the slope, if the objective of a researcher is to minimize the size and maximize the power of the test, the PTT is the best choice.

The analysis is further developed by investigating the relationship between the power function of the PTT and its arguments, namely the slope and the nominal sizes, of the UT, RT and PT. The chosen values of the nominal sizes that are set before testing affect the actual size of the PTT.

In order to get small probability of a Type I error for the PTT, our investigations concentrate on small nominal sizes of the UT, RT and PT with a view to achieving small (actual) significance level of the PTT. The study revealed that for small and moderate values of slope, the smaller the nominal size of the RT, the smaller the size of the PTT when other nominal sizes are kept fixed and small. For moderate and large values of the slope, a large size of the PTT is observed when nominal size of PT is set close to 0. The size of the PTT behaves much like that of the RT when the nominal size of PT is small, but it behaves more like that of the UT when the nominal size of the PT is large.

The power of the PTT is larger for moderate values of λ_2 than for smaller and larger values of λ_2. It is shown analytically that the power of the PTT approaches the power of the RT when the nominal size of PT is closer to 0 but approaches the power of the UT when the nominal size of the PT is closer to

1. In practical applications, the size of the PT should be small (ideally close to 0), and in such cases the power of the PTT is close to that of the RT (which is much higher than that of the UT).

To avoid the larger size of the RT, practitioners are recommended to use the PTT as it achieves smaller size (than the RT) and higher power (than the UT) when the value of λ_2 is small or moderate. Even for large values of λ_2 the PTT has at least as much power as the UT.

Chapter 4

Multivariate Simple Regression Model

4.1 Introduction

The multivariate simple regression model is a generalisation of the commonly used simple regression model. In a multivariate simple regression model, there is a set of response variables corresponding to a non-zero single value of the explanatory variable. For example, in the study of energy usage in a production plant, let the usage of solar, electrical and gas energies be the level of responses corresponding to the level of production output. The level of each response variable, when no effective production is taking place, is known as the *base load*. The base load is related to the energy used in lighting, heating, cooling, office equipment, machine repairs and maintenance. Since the base load is unrelated to the production output, reduction in the base load is profitable to the manufacturer (Kent, 2008). Obviously the usage of energy will increase as the production level goes up. Then, the production level is the explanatory variable that affects the usage of different types of energy. Thus, the usage of energy can be modeled by the multivariate simple regression model.

Consider the multivariate simple regression model,

$$\boldsymbol{X}_i = \boldsymbol{\theta} + \boldsymbol{\beta} c_i + \boldsymbol{e}_i,\ i = 1, \ldots, n, \tag{4.1.1}$$

where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})'$ is the p dimensional response vector, c_i is a non-zero scalar value of the explanatory variable, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ are unknown intercept and slope vectors, and $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{ip})'$ is the p dimensional vector of errors. Assume that each element of $\boldsymbol{e}_i,\ i = 1, \ldots, n$ is not independent but $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are mutually independent with distribution F.

The management of the production plant may wish to test that the base load is equal to a specified vector ($\boldsymbol{\theta}_0$) while they are not sure about the values of the slope parameters. In this situation, they face the following three different scenarios. The management may consider the value of the slope vector either to be (i) completely unspecified, or (ii) specified to a fixed quantity, or (iii) uncertain, but suspected to be a given value from previous knowledge or expert assessment. Three statistical tests are proposed, namely, the unrestricted test (UT), the restricted test (RT) and the pre-test test (PTT) appropriate to the above three possible choices of the slope vector respectively. Thus, the UT is to test $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta}$ is unspecified, the RT is to test $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ (or specified and fixed) and the PTT is to test $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ after pre-testing $H_0^{(3)} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against $H_A^{(3)} : \boldsymbol{\beta} > \boldsymbol{\beta}_0$ (to remove the uncertainty). The PTT is a choice between the UT and the RT. If the null hypothesis $H_0^{(3)}$ is rejected in the pre-test (PT), then the UT is used, otherwise the RT is used.

4.2 The M-estimation

Let F_p be the class of all p-variate absolutely continuous distribution functions which are diagonally symmetric about $\mathbf{0}$ and have a finite Fisher information matrix,

$$\boldsymbol{I} = ((I_{jk})),\ I_{jk} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f'_{[j]}(x)}{f_{[j]}(x)}\frac{f'_{[k]}(y)}{f_{[k]}(y)} dF_{[jk]}(x,y),\ \text{for } j,k = 1,\ldots,p,$$

where $F_{[j]}(x)$ and $F_{[jk]}(x,y)$ are one and two-dimensional marginals of F and $f_{[j]}(x)$ and $f'_{[j]}(x)$ are the first two derivatives of $F_{[j]}(x)$. For every $n(\geq 1)$, define

$$\bar{c}_n = n^{-1}\sum_{i=1}^{n} c_i \ \text{ and } \ C_n^{\star 2} = \sum_{i=1}^{n} c_i^2 - n\bar{c}_n^2. \tag{4.2.1}$$

Also assume that constants $\bar{c}$ and $C^{\star 2}$ exist such that $\bar{c} = \lim_{n\to\infty}\bar{c}_n$, $C^{\star 2} = \lim_{n\to\infty} n^{-1}\, C_n^{\star 2}$ and c_i's are bounded so that

$$\max_{1\leq i\leq n}(c_i - \bar{c}_n)^2/C_n^{\star 2} \to 0 \ \text{ as } \ n\to\infty. \tag{4.2.2}$$

Let $\rho : \Re \to \Re$ be an absolutely continuous function. Thus, the M-estimators of θ_j and β_j are defined as the values of θ_j and β_j that minimize the objective function

$$\sum_{i=1}^{n} \rho\left(\frac{X_{ij} - \theta_j - \beta_j c_i}{S_n}\right), \quad j = 1,2,\ldots,p. \tag{4.2.3}$$

Here, S_n is an appropriate scale statistic for some functional $S = S(F) > 0$. If F is $N(0, \sigma^2)$, $S_n = MAD/0.6745$ is an estimate of $S = \sigma$, where MAD is the mean absolute deviation (Wilcox, 2005, p.78, Montgomery et al., 2001, p.387). The M-estimators based on the componentwise estimating equations for the multivariate model (see Koenker and Portnoy, 1990) are considered in this Chapter. However, the assumption of strong correlation between elements of $\boldsymbol{e}_i$ is questionable because the M-estimates obtained using the method of componentwise equations are more appropriate when there is small dependence between the equations (i.e. weak correlation among the elements of $\boldsymbol{e}_i$) (c.f. Koenker and Portnoy, 1990).

If $\psi = \rho'$ is continuous, then the M-estimators of θ_j and β_j are the solutions of the system of equations (Carroll and Ruppert, 1988, p.210),

$$\sum_{i=1}^{n} \psi\left(\frac{X_{ij} - \theta_j - \beta_j c_i}{S_n}\right) = 0 \text{ and } \sum_{i=1}^{n} c_i \psi\left(\frac{X_{ij} - \theta_j - \beta_j c_i}{S_n}\right) = 0. \quad (4.2.4)$$

Let $\psi : \Re \to \Re$ be a nondecreasing and skew symmetric score function. For any real numbers a_j and b_j, consider

$$\begin{aligned} \boldsymbol{M}_{n_1}(\boldsymbol{a}, \boldsymbol{b}) &= \left[M_{n_1 1}(a_1, b_1), \ldots, M_{n_1 p}(a_p, b_p)\right]' \text{ and} \\ \boldsymbol{M}_{n_2}(\boldsymbol{a}, \boldsymbol{b}) &= \left[M_{n_2 1}(a_1, b_1), \ldots, M_{n_2 p}(a_p, b_p)\right]', \end{aligned}$$

where $M_{n_1 j}(a_j, b_j) = \sum_{i=1}^{n} \psi\left(\frac{X_{ij} - a_j - b_j c_i}{S_n}\right)$ and $M_{n_2 j}(a_j, b_j) = \sum_{i=1}^{n} c_i \psi\left(\frac{X_{ij} - a_j - b_j c_i}{S_n}\right)$ with $\boldsymbol{a} = (a_1, \ldots, a_p)'$ and $\boldsymbol{b} = (b_1, \ldots, b_p)'$.

Let $\tilde{\boldsymbol{\beta}}$ be the constrained M-estimator of $\boldsymbol{\beta}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0 = (\theta_{0_1}, \ldots, \theta_{0_p})'$ is a vector of fixed real numbers, that is, $\tilde{\boldsymbol{\beta}}$ is the solution of $\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{t}_2) = \boldsymbol{0}$, and it may be conveniently expressed as

$$\tilde{\boldsymbol{\beta}} = [\sup\{\boldsymbol{t}_2 : \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{t}_2) > \boldsymbol{0}\} + \inf\{\boldsymbol{t}_2 : \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{t}_2) < \boldsymbol{0}\}]/2. \quad (4.2.5)$$

Similarly, let $\tilde{\boldsymbol{\theta}}$ be the constrained M-estimator of $\boldsymbol{\theta}$ when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0 = (\beta_{0_1}, \ldots, \beta_{0_p})'$ is a vector of fixed real numbers, that is, $\tilde{\boldsymbol{\theta}}$ is the solution of $\boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \boldsymbol{\beta}_0) = \boldsymbol{0}$ and may conveniently be expressed as

$$\tilde{\boldsymbol{\theta}} = [\sup\{\boldsymbol{t}_1 : \boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \boldsymbol{\beta}_0) > \boldsymbol{0}\} + \inf\{\boldsymbol{t}_1 : \boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \boldsymbol{\beta}_0) < \boldsymbol{0}\}]/2. \quad (4.2.6)$$

The constrained M-estimators of intercept and slope parameters are defined by Sen (1982) to test the significance of the slope parameter of the simple regression model. Further note that

$$\int \psi(x/S)dF_{[j]}(x) = 0, \ \ 1 \le j \le p. \tag{4.2.7}$$

Denote

$$\boldsymbol{\Lambda} = ((\lambda_{jk})), \ \ \text{where } \ \lambda_{jk} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x/S)\psi(y/S)dF_{[jk]}(x,y), \tag{4.2.8}$$

$j, k = 1, \ldots, p$ and define

$$\begin{aligned}
\boldsymbol{\Delta}^{(1)} &= ((\delta_{jk}^{(1)})), \ \ \boldsymbol{\Delta}^{(2)} = ((\delta_{jk}^{(2)})) \text{ and } \boldsymbol{\Delta}^{(3)} = ((\delta_{jk}^{(3)})), \ \text{where} && (4.2.9)\\
\delta_{jk}^{(1)} &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{x_{ij}-\theta_{0_j}-\tilde{\beta}_j c_i}{S_n}\right)\psi\left(\frac{x_{ik}-\theta_{0_k}-\tilde{\beta}_k c_i}{S_n}\right), && (4.2.10)\\
\delta_{jk}^{(2)} &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{x_{ij}-\theta_{0_j}-\beta_{0_j} c_i}{S_n}\right)\psi\left(\frac{x_{ik}-\theta_{0_k}-\beta_{0_k} c_i}{S_n}\right) \text{ and} && (4.2.11)\\
\delta_{jk}^{(3)} &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{x_{ij}-\tilde{\theta}_j-\beta_{0_j} c_i}{S_n}\right)\psi\left(\frac{x_{ik}-\tilde{\theta}_k-\beta_{0_k} c_i}{S_n}\right). && (4.2.12)
\end{aligned}$$

Also, define

$$\boldsymbol{T} = ((\tau_{jk})), \ \ \text{with } \tau_{jk} = \lambda_{jk}/(\gamma_j\gamma_k) \tag{4.2.13}$$

and

$$\gamma_j = \frac{1}{S}\int_{-\infty}^{\infty}\psi'(x/S)dF_{[j]}(x).$$

The asymptotic distributions of statistics given in the following Theorem are useful to derive the UT, RT and PTT in the next Section.

Theorem 4.2.1 *Given the asymptotic properties results (i) and (ii) in Ap-*

pendix B.2, as $n \to \infty$,

(i) under $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \;\; n^{-\frac{1}{2}}$

$$\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Lambda} C^{\star 2}/(C^{\star 2} + \bar{c}^2)), \tag{4.2.14}$$

(ii) under $H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\beta} = \boldsymbol{\beta}_0$,

$$n^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{pmatrix} \xrightarrow{d} N_{2p} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \otimes \boldsymbol{\Lambda} \right], \tag{4.2.15}$$

(iii) under $H_0^{(3)} : \boldsymbol{\beta} = \boldsymbol{\beta}_0, \;\; n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Lambda} C^{\star 2}),$ (4.2.16)

where $N_p(\cdot\,,\cdot\,)$ *represents a p-variate normal distribution with appropriate parameters. Here,* $\boldsymbol{A} \otimes \boldsymbol{B}$ *denotes the Kronecker product between matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$.

The proof of Theorem 4.2.1 is given in Appendix B.2.

4.3 The UT, RT and PTT

This Section provides the statistical tests for the UT, RT and PT. These tests are defined using the score function in the M-estimation methodology. The test function of the PTT is proposed in this Section. The asymptotic distributions used to derive the UT, RT and PT are given in Theorem 4.2.1.

4.3.1 The Unrestricted Test (UT)

Let ϕ_n^{UT} be the test function of $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta}$ is unspecified. The proposed test statistic is

$$L_n^{UT} = n^{-1} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}})' \boldsymbol{\Delta}^{(1)^{-1}} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) / (C_n^{\star 2}/(C_n^{\star 2} + n\bar{c}_n^2)),$$

where $\tilde{\boldsymbol{\beta}}$ is the constrained M-estimator defined by equation (4.2.5). It follows from equation (4.2.14) that for large n, $L_n^{UT} \xrightarrow{d} \chi_p^2$ (chi-square distribution with p degrees of freedom) under $H_0^{(1)}$.

Let ℓ_{n,α_1}^{UT} be the critical value of L_n^{UT} at the α_1 level of significance. So, for the test function $\phi_n^{UT} = I(L_n^{UT} > \ell_{n,\alpha_1}^{UT})$, the power function of the UT becomes $\Pi_n^{UT}(\boldsymbol{\theta}) = E(\phi_n^{UT}|\boldsymbol{\theta}) = P(L_n^{UT} > \ell_{n,\alpha_1}^{UT}|\boldsymbol{\theta})$, where $I(A)$ is an indicator function of the set A. It takes value 1 if A occurs, otherwise it is 0.

4.3.2 The Restricted Test (RT)

Let ϕ_n^{RT} be the test function of $H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, the proposed test statistic is

$$L_n^{RT} = n^{-1}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)'\boldsymbol{\Delta}^{(2)^{-1}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0).$$

Note that for large n, it follows from equation (4.2.15) that L_n^{RT} is χ_p^2 under $H_0^{(2)}$. Again, let ℓ_{n,α_2}^{RT} be the critical value of L_n^{RT} at the α_2 level of significance. Thus, for the test function $\phi_n^{RT} = I(L_n^{RT} > \ell_{n,\alpha_2}^{RT})$, the power function of the RT becomes $\Pi_n^{RT}(\boldsymbol{\theta}) = E(\phi_n^{RT}|\boldsymbol{\theta}) = P(L_n^{RT} > \ell_{n,\alpha_2}^{RT}|\boldsymbol{\theta})$.

4.3.3 The Pre-test (PT)

For the pre-test on the slope vector, the test function of $H_0^{(3)} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against $H_A^{(3)} : \boldsymbol{\beta} > \boldsymbol{\beta}_0$ is ϕ_n^{PT}. The proposed test statistic is

$$L_n^{PT} = n^{-1}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)'\boldsymbol{\Delta}^{(3)^{-1}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)/(C_n^{\star 2}/n),$$

where $\tilde{\boldsymbol{\theta}}$ is the constrained M-estimator as defined in equation (4.2.6). Under $H_0^{(3)}$, as $n \to \infty$, it follows from equation (4.2.16) that $L_n^{PT} \xrightarrow{d} \chi_p^2$.

4.3.4 The Pre-test Test (PTT)

Let ϕ_n^{PTT} be the test function for testing $H_0^{(1)}$ against $H_A^{(1)}$ following a pre-test on $\boldsymbol{\beta}$. Since the PTT is a choice between the RT and the UT, define,

$$\phi_n^{PTT} = I[(L_n^{PT} < \ell_{n,\alpha_3}^{PT}, L_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (L_n^{PT} > \ell_{n,\alpha_3}^{PT}, L_n^{UT} > \ell_{n,\alpha_1}^{UT})], \quad (4.3.1)$$

where ℓ_{n,α_3}^{PT} is the critical value of L_n^{PT} at the α_3 level of significance. The power function of the PTT is given by

$$\Pi_n^{PTT}(\boldsymbol{\theta}) = E(\phi_n^{PTT}|\boldsymbol{\theta}) \quad (4.3.2)$$

and the size of the PTT is obtained by substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ in equation (4.3.2).

4.4 Asymptotic Distributions under Local Alternatives

Theorem 4.4.1 *Let $\{K_n\}$ be a sequence of local alternative hypotheses, where*

$$K_n : (\boldsymbol{\theta}, \boldsymbol{\beta}) = (\boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 + n^{-\frac{1}{2}}\boldsymbol{\varrho}_2), \quad (4.4.1)$$

with $\boldsymbol{\varrho}_1 = n^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) > \mathbf{0}$ and $\boldsymbol{\varrho}_2 = n^{\frac{1}{2}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) > \mathbf{0}$. Here, $\boldsymbol{\varrho}_1 = (\varrho_{1_1}, \ldots, \varrho_{1_p})'$ and $\boldsymbol{\varrho}_2 = (\varrho_{2_1}, \ldots, \varrho_{2_p})'$ are vectors of fixed real numbers. Under $\{K_n\}$, for large sample,

$$\text{(i)} \quad n^{-\frac{1}{2}} \begin{bmatrix} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) \end{bmatrix} \xrightarrow{d} N_{2p}\left[\begin{pmatrix} \boldsymbol{\gamma}(\boldsymbol{\varrho}_1 + \boldsymbol{\varrho}_2\bar{c}) \\ \boldsymbol{\gamma}\boldsymbol{\varrho}_2 C^{\star 2} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{pmatrix} \otimes \boldsymbol{\Lambda} \right], \quad (4.4.2)$$

$$
(ii) \qquad n^{-\frac{1}{2}} \begin{bmatrix} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \\ \\ \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) \end{bmatrix} \xrightarrow{d} N_{2p}
$$

$$
\left[\begin{pmatrix} \frac{\boldsymbol{\gamma}\boldsymbol{\varrho}_1 C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ \\ \boldsymbol{\gamma}\boldsymbol{\varrho}_2 C^{\star 2} \end{pmatrix}, \begin{pmatrix} \frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2} & -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ \\ -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} & C^{\star 2} \end{pmatrix} \otimes \boldsymbol{\Lambda} \right], \quad (4.4.3)
$$

where $\boldsymbol{\gamma} = diag(\gamma_1 \ldots \gamma_p)$.

The proof of Theorem 4.4.1 is given in Appendix B.2.

Theorem 4.4.2 *Under* $\{K_n\}$*, asymptotically* (L_n^{RT}, L_n^{PT}) *are independently distributed as bivariate noncentral chi-square with* p *degrees of freedom and* (L_n^{UT}, L_n^{PT}) *are distributed as correlated bivariate noncentral chi-square with* p *degrees of freedom and noncentrality parameters,*

$$
\begin{aligned}
\theta^{UT} &= (\boldsymbol{\varrho}_1' \boldsymbol{T}^{-1} \boldsymbol{\varrho}_1)(C^{\star 2}/(C^{\star 2} + \bar{c}^2)), && (4.4.4) \\
\theta^{RT} &= (\boldsymbol{\varrho}_1 + \bar{c}\boldsymbol{\varrho}_2)' \boldsymbol{T}^{-1} (\boldsymbol{\varrho}_1 + \bar{c}\boldsymbol{\varrho}_2), && (4.4.5) \\
\theta^{PT} &= (\boldsymbol{\varrho}_2' \boldsymbol{T}^{-1} \boldsymbol{\varrho}_2) C^{\star 2}, && (4.4.6)
\end{aligned}
$$

where $\boldsymbol{T}$ *is defined in equation (4.2.13).*

Proof The proof of this theorem is directly obtained using Theorem 4.4.1 and Theorem 1.4.1 of Muirhead (1982, p.26).

4.5 Asymptotic Performance of the Tests

Using results in Section 4.4, under $\{K_n\}$, the asymptotic power functions for the UT, RT and PT which are denoted by $\Pi^h(\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2)$, $h = UT, RT, PT$, are defined as

$$\Pi^h(\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2) = \lim_{n\to\infty} \Pi_n^h(\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2) = \lim_{n\to\infty} P(L_n^h > \ell_{n,\alpha_\nu}^h | K_n) = 1 - G_p(\chi^2_{p,\alpha_\nu}; \theta^h), \tag{4.5.1}$$

where $G_p(\chi^2_{p,\alpha_\nu}, \theta^h)$ is the cdf of a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter θ^h. The level of significance, $\alpha_\nu, \nu = 1, 2, 3$ is chosen together with the critical values ℓ^h_{n,α_ν} for the UT, RT and PT. Here, $\chi^2_{p,\alpha}$ is the upper $100\alpha\%$ critical value of a central chi-square distribution and $\ell^{UT}_{n,\alpha_1} \to \chi^2_{p,\alpha_1}$ under $H_0^{(1)}$, $\ell^{RT}_{n,\alpha_2} \to \chi^2_{p,\alpha_2}$ under $H_0^{(2)}$ and $\ell^{PT}_{n,\alpha_3} \to \chi^2_{p,\alpha_3}$under $H_0^{(3)}$.

Obviously the power function of the UT, RT and PTT which relies on the noncentrality parameters θ^{UT}, θ^{RT} and θ^{PT}, depends on the sample size n. The noncentrality parameters are a function of $\boldsymbol{\varrho}_1$, $\boldsymbol{\varrho}_2$ and $\boldsymbol{T}$, which depend on the sample size n (see equations (4.4.4), (4.4.5) and (4.4.6)). Both $\boldsymbol{\varrho}_1$ and $\boldsymbol{\varrho}_2$ are decreasing as n increases (see equation (4.4.1)). However, $\boldsymbol{T}$ depends on the set of random variables, $X_i, \quad i = 1, \ldots, n$ (see equation (4.2.13)). A set of observations with a larger sample size does not imply its estimate of $\boldsymbol{T}$ always smaller (or always larger) than that of a set of observations with a smaller sample size. Thus, the power function of the UT, RT and PTT is not monotone increasing or monotone decreasing as n increases when arguments $\boldsymbol{\varrho}_1$, $\boldsymbol{\varrho}_2$ and $\alpha_i, \quad i = 1, 2, 3$ are fixed. For example, Π^{RT} increases as $\boldsymbol{\varrho}_2$ increases when the other arguments, including the sample size n, are fixed. However, due to the randomness of observations, θ^{RT} does not increase monotonely (or decrease monotonely) as the sample size n increases, when the other arguments are fixed. Thus, there is no explicit relationship between the sample size and the power function of the UT, RT or the PTT.

For a large n (fixed), since $\theta^{RT} \geq \theta^{UT}$ when $\bar{c} \geq 0$ or $\boldsymbol{\varrho}_2 = \mathbf{0}$, this means the asymptotic power function of the RT is greater than that of the UT if $\alpha_1 = \alpha_2$. We conclude that the asymptotic size of the RT is larger than that of the UT but the asymptotic power of the UT is smaller than that of the RT. For testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}$, using equation (4.3.1) and Theorem 4.4.2, the asymptotic power function for the PTT under $\{K_n\}$ is given as

$$\begin{aligned}
& \Pi^{PTT}(\boldsymbol{\varrho}_1, \boldsymbol{\varrho}_2) \\
= \; & \lim_{n\to\infty} P(L_n^{PT} \leq \ell_{n,\alpha_3}^{PT}, L_n^{RT} > \ell_{n,\alpha_2}^{RT} | K_n) + \\
& \lim_{n\to\infty} P(L_n^{PT} > \ell_{n,\alpha_3}^{PT}, L_n^{UT} > \ell_{n,\alpha_1}^{UT} | K_n) \\
= \; & G_p(\chi^2_{p,\alpha_3}; \theta^{PT})\{1 - G_p(\chi^2_{p,\alpha_2}; \theta^{RT})\} + \int_{\chi^2_{p,\alpha_1}} \int_{\chi^2_{p,\alpha_3}} \tilde{\phi}(w_1, w_2) dw_1 dw_2,
\end{aligned} \tag{4.5.2}$$

where $\tilde{\phi}(\cdot)$ is the density function of the bivariate noncentral chi-square distribution with p degrees of freedom, noncentrality parameters, θ^{UT} and θ^{PT} and correlation coefficient, $\rho = -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}$. It is observed that $G_p(\chi^2_{p,\alpha_3}; \theta^{PT})$ is decreasing as θ^{PT} increases and $1 - G_p(\chi^2_{p,\alpha_2}; \theta^{RT})$ is increasing as θ^{RT} increases.

The probability integral in (4.5.2) is given by

$$\begin{aligned}
& \int_{\chi^2_{p,\alpha_1}} \int_{\chi^2_{p,\alpha_3}} \tilde{\phi}(w_1, w_2) dw_1 dw_2 \\
= \; & \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{\kappa_1=0}^{\infty} \sum_{\kappa_2=0}^{\infty} (1-\rho^2)^p \frac{\Gamma(\frac{p}{2}+j)}{\Gamma(\frac{p}{2})j!} \frac{\Gamma(\frac{p}{2}+k)}{\Gamma(\frac{p}{2})k!} \rho^{2(j+k)} \\
& \times \left[1 - \gamma^\star\left(\frac{p}{2} + j + \kappa_1, \frac{\chi^2_{p,\alpha_1}}{2(1-\rho^2)}\right)\right] \left[1 - \gamma^\star\left(\frac{p}{2} + k + \kappa_2, \frac{\chi^2_{p,\alpha_3}}{2(1-\rho^2)}\right)\right] \\
& \times \frac{e^{-\theta^{UT}/2}(\theta^{UT}/2)^{\kappa_1}}{\kappa_1!} \frac{e^{-\theta^{PT}/2}(\theta^{PT}/2)^{\kappa_2}}{\kappa_2!}.
\end{aligned} \tag{4.5.3}$$

Here, $\gamma^\star(v, d) = \int_0^d x^{v-1} e^{-x}/\Gamma(v) dx$ is the incomplete gamma function. For details on the evaluation of the bivariate integral, see Yunus and Khan (2009). The density function of the bivariate noncentral chi-square distribution given above is a mixture of the bivariate central chi-square distribution (see Gunst and Webster, 1973, Wright and Kennedy, 2002) with the probabilities from the Poisson distribution.

Then, write the equation (4.5.3) as

$$\int_{\chi^2_{p,\alpha_1}}^{\infty}\int_{\chi^2_{p,\alpha_3}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2 = [1-H_p(\chi^2_{p,\alpha_1};\theta^{UT})][1-H_p(\chi^2_{p,\alpha_3};\theta^{PT})] \quad (4.5.4)$$

with $H_p(\chi^2_{p,\alpha_1};\theta^{UT}) = \sum_{j=0}^{\infty}\sum_{\kappa_1=0}^{\infty}\frac{(1-\rho^2)^{\frac{p}{2}}\Gamma(\frac{p}{2}+j)\rho^{2j}}{\Gamma(\frac{p}{2})j!}\gamma^\star\left(\frac{p}{2}+j+\kappa_1,\frac{\chi^2_{p,\alpha_1}}{2(1-\rho^2)}\right)e^{-\frac{\theta^{UT}}{2}}$ $\left(\frac{\theta^{UT}}{2}\right)^{\kappa_1}/\kappa_1!$ and $H_p(\chi^2_{p,\alpha_3};\theta^{PT}) = \sum_{k=0}^{\infty}\sum_{\kappa_2=0}^{\infty}\frac{(1-\rho^2)^{\frac{p}{2}}\Gamma(\frac{p}{2}+k)\rho^{2k}}{\Gamma(\frac{p}{2})k!}\gamma^\star\left(\frac{p}{2}+k+\kappa_2,\right.$ $\left.\frac{\chi^2_{p,\alpha_3}}{2(1-\rho^2)}\right)e^{-\frac{\theta^{PT}}{2}}\left(\frac{\theta^{PT}}{2}\right)^{\kappa_2}/\kappa_2!$. Note, $H_p(\chi^2_{p,\alpha_1};\theta^{UT}) \geq G_p(\chi^2_{p,\alpha_1};\theta^{UT})$ and $H_p(\chi^2_{p,\alpha_3};\theta^{PT}) \geq G_p(\chi^2_{p,\alpha_3};\theta^{PT})$. Equality is achieved when $\rho = 0$, or when $\boldsymbol{\varrho}_1 = \mathbf{0}$ and $\boldsymbol{\varrho}_2 = \mathbf{0}$.

Consider all cases when $\bar{c} > 0$ and $\rho \neq 0$. Using equation (4.5.4), we rewrite equation (4.5.2) as

$$\Pi^{PTT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2) \leq \Pi^{RT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2)[1-\Pi^{PT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2)] + \Pi^{UT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2)\Pi^{PT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2). \quad (4.5.5)$$

Equality in equation (4.5.5) is achieved when both $\boldsymbol{\varrho}_1$ and $\boldsymbol{\varrho}_2$ are $\mathbf{0}$.

Obviously $\Pi^{PTT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2) \leq \Pi^{RT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2) - \Pi^{PT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2)v_2$ for $0 \leq v_2 < 1$, and it follows that $\Pi^{PTT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2) \leq \Pi^{RT}(\boldsymbol{\varrho}_1,\boldsymbol{\varrho}_2)$ for any $\boldsymbol{\varrho}_1$ and $\boldsymbol{\varrho}_2$. Equality holds when both $\boldsymbol{\varrho}_1$ and $\boldsymbol{\varrho}_2$ are $\mathbf{0}$.

Wright and Kennedy (2002) computed the cumulative distribution function (cdf) for the bivariate central chi-square distribution. In their paper, as the correlation coefficient ρ gets larger the cdf grows larger too. In the same manner, the cdf for the bivariate noncentral chi-square distribution increases as the correlation coefficient increases (c.f. Yunus and Khan, 2009).

Rewrite equation (4.5.3) as

$$\begin{aligned}&\int_{\chi^2_{p,\alpha_1}}^{\infty}\int_{\chi^2_{p,\alpha_3}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2\\ = \;& 1-H_p(\chi^2_{p,\alpha_1};\theta^{UT})-H_p(\chi^2_{p,\alpha_3};\theta^{PT})+\int_0^{\chi^2_{p,\alpha_1}}\int_0^{\chi^2_{p,\alpha_3}}\tilde{\phi}(w_1,w_2)dw_1dw_2.\end{aligned} \quad (4.5.6)$$

When $\boldsymbol{\varrho}_2$ is not large but not $\mathbf{0}$ and $\boldsymbol{\varrho}_1$ is sufficiently large, the first term on the right hand side of the equation (4.5.2) becomes $G(\chi^2_{p,\alpha_3};\theta^{PT})$ because

θ^{RT} is sufficiently large. The second and fourth terms on the right hand side of the equation (4.5.6) becomes 0 because θ^{UT} is large. Also, note that $H_p(\chi^2_{p,\alpha_3}; \theta^{PT}) > G_p(\chi^2_{p,\alpha_3}; \theta^{PT})$. So, $\Pi^{PTT} < \Pi^{UT} = 1$ for sufficiently large $\boldsymbol{\varrho}_1$ and not so large $\boldsymbol{\varrho}_2$ $(\neq \mathbf{0})$.

When $\boldsymbol{\varrho}_2 = \mathbf{0}$ and $\boldsymbol{\varrho}_1$ is sufficiently large, the first term on the right hand side of the equation (4.5.2) becomes $1 - \alpha_3$ because θ^{RT} is large. Both $H_p(\chi^2_{p,\alpha_1}; \theta^{UT})$ and $\int_0^{\chi^2_{p,\alpha_1}} \int_0^{\chi^2_{p,\alpha_3}} \tilde{\phi}(w_1, w_2) dw_1 dw_2$ of the equation (4.5.6) become 0 while $H_p(\chi^2_{p,\alpha_3}; \theta^{PT})$ becomes $1 - \alpha_3$. So, $\Pi^{PTT} = \Pi^{UT} = 1$ when $\boldsymbol{\varrho}_1$ is sufficiently large and $\boldsymbol{\varrho}_2 = \mathbf{0}$.

In the same manner, we observe results given in (c)-(e) below

(a) When $\boldsymbol{\varrho}_2$ $(\neq \mathbf{0})$ is not large and $\boldsymbol{\varrho}_1$ is sufficiently large, then, $\Pi^{PTT} < \Pi^{UT} = 1$.

(b) When $\boldsymbol{\varrho}_2 = \mathbf{0}$ and $\boldsymbol{\varrho}_1$ is sufficiently large, then, $\Pi^{PTT} = \Pi^{UT} = 1$.

(c) When $\boldsymbol{\varrho}_1$ $(\neq \mathbf{0})$ is not large but $\boldsymbol{\varrho}_2$ is sufficiently large, then, $\Pi^{PTT} < \Pi^{UT} = 1$.

(d) Let $\alpha_1 = \alpha$. When $\boldsymbol{\varrho}_1 = \mathbf{0}$ and $\boldsymbol{\varrho}_2$ is sufficiently large, then, $\Pi^{PTT} = \Pi^{UT} = \alpha$.

(e) Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. When both $\boldsymbol{\varrho}_1$ and $\boldsymbol{\varrho}_2$ are $\mathbf{0}$, then $\Pi^{PTT} = \Pi^{UT} = \alpha$.

In this Section, the tests are analytically compared using the size and power of the tests. The relative efficiency of the power functions could also be used to compare the relative performances of the tests. Because the ultimate conclusions using the relative efficiency would be the same as those using the power function, we do not pursue this any further in this Chapter.

4.6 Simulation Study

In this Section, the size and power of the UT, RT and PTT are computed using equations (4.5.1) and (4.5.2). Obviously the size of the test is obtained when $\boldsymbol{\varrho}_1 = \mathbf{0}$ (or $\boldsymbol{\theta} = \boldsymbol{\theta}_0$) and $\boldsymbol{\varrho}_2$ any values in equations (4.5.1) and (4.5.2). The nominal sizes of the UT, RT and PT are respectively α_1, α_2 and α_3. Here, we let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$. Obviously $\Pi^{UT}(\mathbf{0}, \mathbf{0}) = \alpha_1$, $\Pi^{RT}(\mathbf{0}, \mathbf{0}) = \alpha_2$ and $\Pi^{PT}(\mathbf{0}, \mathbf{0}) = \alpha_3$ when both $\boldsymbol{\varrho}_1$ and $\boldsymbol{\varrho}_2$ are $\mathbf{0}$ in equation (4.5.1). The power of the test is obtained using equations (4.5.1) and (4.5.2) by letting $\boldsymbol{\varrho}_1 > \mathbf{0}$ (or $\boldsymbol{\theta} > \boldsymbol{\theta}_0$) and $\boldsymbol{\varrho}_2$ any values. The noncentrality parameters, θ^{UT}, θ^{RT} and θ^{PT} (see equations (4.4.4), (4.4.5) and (4.4.6)) are computed to get the size and power of the tests. The noncentrality parameters are a function of $\boldsymbol{T}$, $\boldsymbol{\varrho}_1$, $\boldsymbol{\varrho}_2$ and c_i. The regressor values, c_i, $i = 1, 2, \ldots, n$ are 0 and 1 with 50% for each. In the simulation, let $p = 2$ in the multivariate simple regression model be given by equation (4.1.1). The error term, $\boldsymbol{e}_i$, $i = 1, 2, \ldots, n$ of size n is generated randomly from distributions: (i) normal with mean 0 and variance 1, $N(0, 1)$, (ii) 10% wild: First, $\boldsymbol{e}_i$ is generated from the normal distribution with mean 0 and variance 1, then choose randomly 10% of the generated $\boldsymbol{e}_i$ and multiply them by a scalar 10, and (iii) Cauchy distribution with location 0 and scale 1. Here, take $n =$ 10, 15, 50 and 60.

Since $\boldsymbol{T}$ is in an integral form that depends on the distribution of the observations, for simulation purposes, it is estimated using

$$\hat{\boldsymbol{T}} = \begin{pmatrix} \frac{\hat{\lambda}_{11}}{\hat{\gamma}_1 \hat{\gamma}_1} & \frac{\hat{\lambda}_{12}}{\hat{\gamma}_1 \hat{\gamma}_2} \\ \frac{\hat{\lambda}_{21}}{\hat{\gamma}_2 \hat{\gamma}_1} & \frac{\hat{\lambda}_{22}}{\hat{\gamma}_2 \hat{\gamma}_2} \end{pmatrix},$$

where $\hat{\gamma}_j = \frac{1}{n} \sum_{i=1}^{n} \psi'(e_{ij})$ and $\hat{\lambda}_{jk} = \frac{1}{n} \sum_{i=1}^{n} \psi(e_{ij})\psi(e_{ik})$, $j, k = 1, 2$.

Consider three types of ψ-functions: (a) Least-square (LS), $\psi_{LS}(u) = u$ for any $u \in \Re$, (b) Huber, $\psi_H(u) = u$ if $|u| \leq k_1$, k_1 sgn(u) if $|u| > k_1$, (c) Tukey bi-square, $\psi_{TB}(u) = u(k_2^2 - u^2)^2$ if $|u| \leq k_2$, otherwise 0. Let $k_1 = 1.345$ and $k_2 = 4.685$. For the computation of the size and power of the PTT, a program

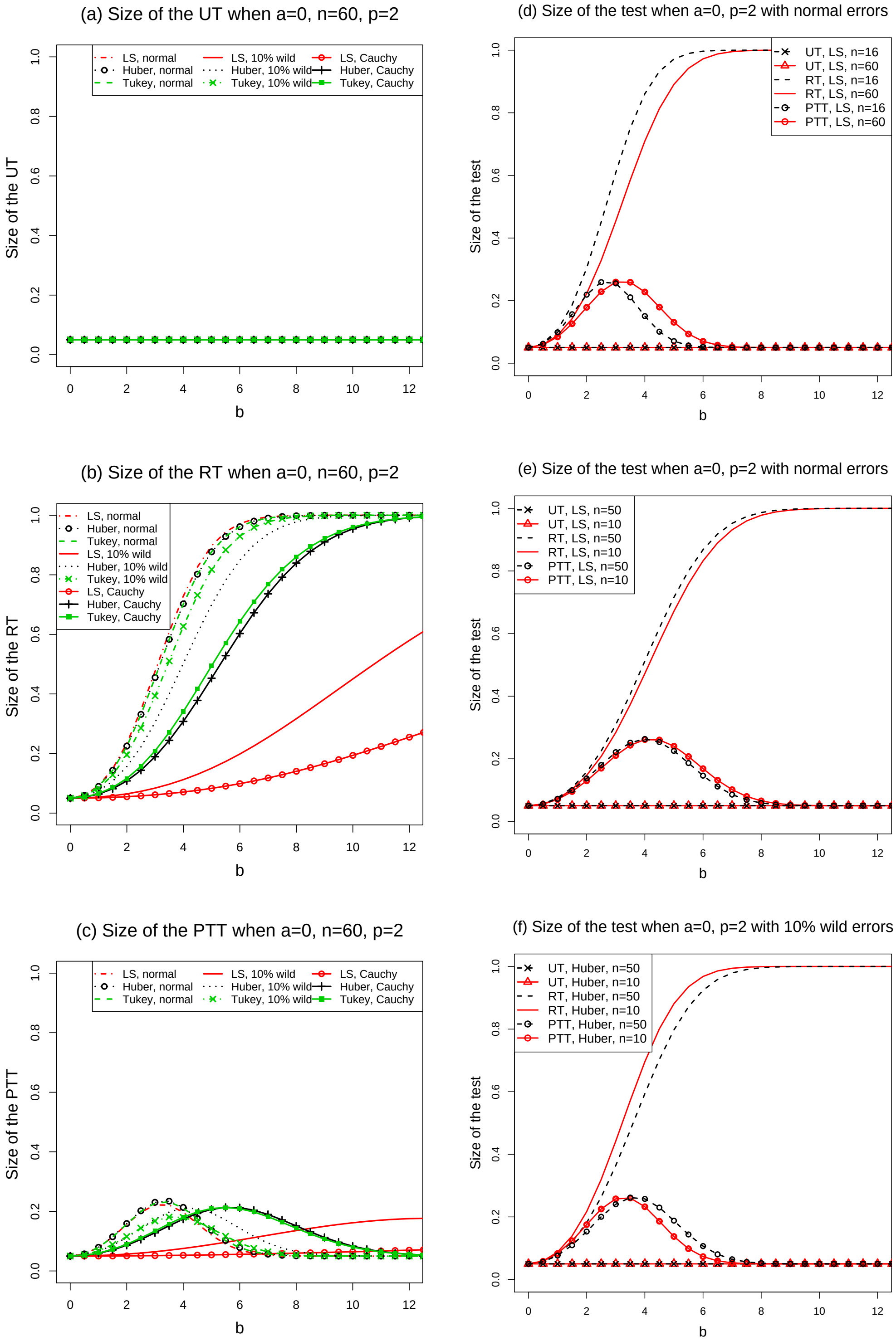


Figure 4.1: Graphs of size of the tests as a function of b for selected values of a.

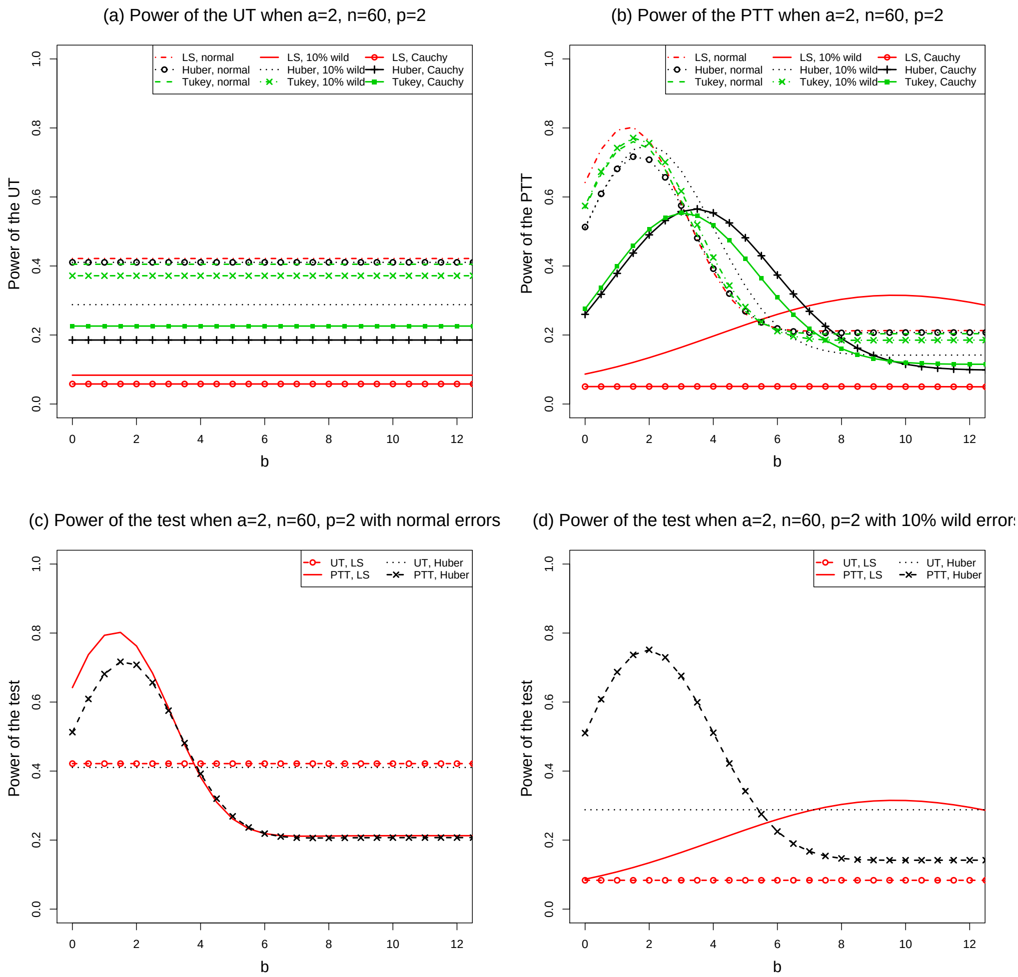


Figure 4.2: Graphs of power of the tests as a function of b for selected values of a.

is written in R to compute the bivariate integral of the noncentral chi-square distribution. The simulation is run 100 times to get 100 simulated sets of values of error terms. The average of the power function of the 100 simulated data sets is computed.

Since the UT, RT and PTT are defined based on knowledge of the slope vector, the size and power of the UT, RT and PTT are plotted with respect to b for selected a. Here, $(b, b) = (\varrho_{2_1}, \varrho_{2_2})$, where $\frac{1}{\sqrt{n}}\varrho_{2_1} = \beta_1 - \beta_{0_1}$ and $\frac{1}{\sqrt{n}}\varrho_{2_2} = \beta_2 - \beta_{0_2}$, while $(a, a) = (\varrho_{1_1}, \varrho_{1_2})$, where $\frac{1}{\sqrt{n}}\varrho_{1_1} = \theta_1 - \theta_{0_1}$ and $\frac{1}{\sqrt{n}}\varrho_{1_2} = \theta_2 - \theta_{0_2}$. So, b is actually a function of the difference between the true slope and its suspected value while a is a function of the difference between the true intercept and its suspected value. As b increases, the size of the RT increases and reaches 1 (see Figure 4.1(b)) while the size and power of the UT are constant (see Figures 4.1(a) and 4.2(a)) regardless of the values of b. It is depicted from Figures 4.1(c) and 4.2(b) that the size and power of the PTT increases to a peak (less than 1) and then decreases as b increases for selected small a.

The robustness properties of the M-test are investigated computationally through simulation in this Section. The performance of a test depends on the ψ-function and the distribution of the simulated errors. The size of the UT remains constant at nominal size, $\alpha = 0.05$ as b grows larger for all considered ψ-functions and distributions of the simulated errors (see Figure 4.1(a)). The (actual) size of the RT is significantly different from the nominal size as b grows larger. It reaches 1 as b grows larger (see Figure 4.1(b)). Therefore, the RT is not a valid test because it does not meet the asymptotic level constraint. Although the size of the PTT is different from the nominal size, it does not reach 1 for any b (see Figure 4.1(c)). Figures 4.1(d), 4.1(e) and 4.1(f) show that the size of the RT (or PTT) with a larger sample size n is not always larger or always smaller than that of with a smaller sample size n.

Since the RT is not a valid test, it is not compared for the power of the test. Only the power of the test for the UT and PTT are compared and are plotted in Figure 4.2. Figures 4.2(c) and 4.2(d) show that the UT has constant smallest power when the values of $b < q$, where q is some positive value. Also, the PTT has larger power than that of the UT when $b < q$. However, the PTT has lower power than that of the UT when $b > q$. For example, we find that the PTT has at least as much power as the UT when $b < 4 = q$ (or, equivalently, when $\beta_0 - \beta_{0_j} < 4/\sqrt{60}$) and $a = 2$ (or, equivalently, $\theta - \theta_{0_j} = 2/\sqrt{60},\ j = 1, 2$) (see Figure 4.2(c)). Although the prior information on the slope vector may be uncertain, there is a high possibility that the true values are not too far from the suspected values. Therefore, the study on the behaviour of the three tests when $b < q$ is more realistic.

Figure 4.2(a) depicts that the power of the UT for the normal simulated errors is higher than that of the non-normal simulated errors. For normal simulated errors, the PTT based on the LS ψ-function has larger power than that of the PTT using the Huber and Tukey ψ-functions when $b < q$ (see Figure 4.2(b)). By contrast, the PTT using the Huber and Tukey ψ-functions have larger power than that of using the LS ψ-function when the simulated errors are nonnormal. Also, the UT using the redecending Tukey ψ-function performs better in term of the power than that of the Huber (Figure 4.2(a)). The power of the tests using the Huber and Tukey ψ-functions are not much affected by the assumed distribution on the simulated errors compared to that of the LS (see Figures 4.2(c) and 4.2(d)). The power of the tests using the LS ψ-function is easily affected by the assumed distribution on the data and these tests perform best under normal model assumptions.

Note, a test is said to be robust if the power of the test is not significantly affected by any departure from the model assumption (see Burt and Barber, 1996, p.332) and when the nominal and actual sizes are not significantly different under slight model failure (c.f. Carolan and Rayner, 2000). Based on the results from the simulation studies, this definition allows us to conclude that the UT using the Huber or Tukey ψ-functions is more robust compared to that of the LS.

4.7 Discussion and Conclusion

The sampling distributions of the UT, RT and PT follow a univariate noncentral chi-square distribution under the alternative hypothesis when the sample size is large. However, that of the PTT is a bivariate noncentral chi-square distribution as there is a correlation between the UT and PT. Note that there is no such correlation between the RT and PT. To evaluate the power function of the PTT, new R codes are used for the computation of the distribution function of the bivariate noncentral chi-square distribution.

The power of the M-tests using the Huber and Tukey ψ-functions (score functions) are not significantly affected by slight departures from model assumptions. The test based on LS depends heavily on the model assumption and it is not robust if the normality assumption is not satisfied.

The size of the RT reaches 1 as b (a function of the difference between the true and suspected values of the slope vector) increases, so it is not a valid test because it does not satisfy the level constraint. Although the UT has the smallest constant size, it has the smallest power as well, except for very large values of b, that is when $b > q$, where q is some positive number. So, the UT fails to achieve the highest power and lowest size simultaneously. The PTT has a smaller size than the RT and its size does not reach 1 as b increases. It also has higher power than the UT, except for $b > q$. Therefore if the prior information is not far away from the true value, that is, b is near 0 (small or moderate) the PTT has a smaller size than the RT and higher power than the UT. Hence it is a better compromise between the two extremes. It is reasonable to expect b should not be too far away from 0 since the prior information is coming from previous experience or expert knowledge, and thus the PTT demonstrates a reasonable domination over the other two tests in a more realistic situation.

Chapter 5

Parallelism Model

5.1 Introduction

A researcher may model independent data sets from two random samples for two separate groups of respondents. Often, the researcher may wish to know whether the regression lines for the two groups are parallel (i.e. the slopes of the two regression lines are equal) or whether the lines have the same intercept on vertical-axis. An interesting situation would be if the researcher decides to test the equality of the intercepts when the equality of slopes is suspected, but he/she is not sure. Data for this problem can be represented by two simple linear regression equations.

A set of p (> 1) simple regression models is known as the parallelism model. Consider a set of p simple regression models

$$\boldsymbol{X}_{j_{n_j}} = \theta_j \mathbf{1}_{n_j} + \beta_j \boldsymbol{c}_j + \boldsymbol{\varepsilon}_j, \;\; j = 1, \ldots, p, \tag{5.1.1}$$

where $\boldsymbol{X}_{j_{n_j}} = (X_{j_1}, \ldots, X_{j_{n_j}})'$ is a vector of n_j observable response random variables, $\mathbf{1}_{n_j} = (1, 1, \ldots, 1)'$ is an n_j-tuple of 1's, $\boldsymbol{c}_j = (c_{j_1}, \ldots, c_{j_{n_j}})'$ is a vector of n_j independent variables, θ_j and β_j are unknown intercept and slope parameters respectively and $\boldsymbol{\varepsilon}_j = (\varepsilon_{j_1}, \ldots, \varepsilon_{j_{n_j}})'$ is a vector of errors, $\varepsilon_{j_i}, \;\; j = 1, \ldots, p, \;\; i = 1, \ldots, n_j$. Assume that $\{\varepsilon_{j_i}\} = \{X_{j_i} - \theta_j - \beta_j c_{j_i}\}$ are mutually independent and identically distributed with cumulative distribution function (cdf) F_{j_i} such that

$$F_{j_i} = F(X_{j_i} - \theta_j - \beta_j c_{j_i}) \tag{5.1.2}$$

and F is an unknown continuous distribution function.

The researcher may wish to test the intercept vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ while it is not sure if the p-slope parameters are equal. In this situation, three different scenarios associated with the value of the slopes are considered: the value of the slopes would either be (i) completely unspecified, (ii) equal at an arbitrary constant, β_0, or (iii) suspected to be equal at an arbitrary constant, β_0. The unrestricted test (UT), the restricted test (RT) and the pre-test test (PTT) are defined respectively for the three scenarios of the slope parameters. Thus, the UT is for testing $H_0^{(1)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)}: \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is unspecified, the RT is for testing $H_0^{(1)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)}: \boldsymbol{\theta} > \boldsymbol{\theta}_0$ when $\boldsymbol{\beta} = \beta_0 \mathbf{1}_p$ (fixed vector) and the PTT is for testing $H_0^{(1)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)}: \boldsymbol{\theta} > \boldsymbol{\theta}_0$ after pre-testing $H_0^\star: \boldsymbol{\beta} = \beta_0 \mathbf{1}_p$ against $H_A^\star: \boldsymbol{\beta} > \beta_0 \mathbf{1}_p$ (to remove the uncertainty). The PTT is a choice between the UT and the RT. If the null hypothesis $H_0^\star$ is rejected in the pre-test (PT), then the UT is used, otherwise the RT is used.

5.2 The M-estimation

Given an absolutely continuous function $\rho : \Re \to \Re$, the M-estimator of θ_j and β_j is defined as the values of θ_j and β_j that minimize the objective function

$$\sum_{i=1}^{n_j} \rho\left(\frac{X_{j_i} - \theta_j - \beta_j c_{j_i}}{S_n}\right). \tag{5.2.1}$$

Here S_n is an appropriate scale statistic for some functional $S = S(F) > 0$. If $\psi = \rho'$, then the M-estimator of θ_j and β_j are the solutions for the system of equations,

$$\sum_{i=1}^{n_j} \psi\left(\frac{X_{j_i} - \theta_j - \beta_j c_{j_i}}{S_n}\right) = 0, \quad \sum_{i=1}^{n_j} c_{j_i} \psi\left(\frac{X_{j_i} - \theta_j - \beta_j c_{j_i}}{S_n}\right) = 0. \tag{5.2.2}$$

Let $n = n_1 + \ldots + n_p$, $\boldsymbol{\Lambda}_n = \mathrm{Diag}\left(\frac{n_1}{n}, \ldots, \frac{n_p}{n}\right)$ and

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n_j} c_{j_i} = \lambda_j \bar{c}_j \; (|\bar{c}_j| < \infty), \;\; \lim_{n\to\infty} n^{-1} {C_{n_j}^\star}^2 = \lambda_j C_j^{\star 2},$$

where

$${C_{n_j}^\star}^2 = \sum_{i=1}^{n_j} c_{j_i}^2 - n_j \bar{c}_{n_j}^2 \text{ and } \bar{c}_{n_j} = n_j^{-1} \sum_{i=1}^{n_j} c_{j_i}.$$

Also,

$$\lim_{n\to\infty} \frac{n_j}{n} = \lambda_j \;\; (0 < \lambda_j < 1),$$

meaning,

$$\lim_{n\to\infty} \Lambda_n = \Lambda_0 = \mathrm{Diag}(\lambda_1, \ldots, \lambda_p).$$

Assume F is symmetrically distributed about 0, so

$$\int_{-\infty}^{\infty} \psi(z/S)dF(z) = 0 \text{ and } \sigma_0^2 = \int_{-\infty}^{\infty} \psi^2(z/S)dF(z). \tag{5.2.3}$$

Let

$$\gamma = \int_{-\infty}^{\infty} \frac{1}{S}\psi'(z/S)dF(z). \tag{5.2.4}$$

Further assume that σ_0 and γ are both positive and finite quantities. Let $\psi : \Re \to \Re$ be the nondecreasing and skew symmetric score function. For any real numbers a_j and b_j, consider statistics

$$\begin{aligned} \boldsymbol{M}_{n_1}(\boldsymbol{a}, \boldsymbol{b}) &= (M_{n_1}^{(1)}(a_1, b_1), \ldots, M_{n_1}^{(p)}(a_p, b_p))' \text{ and} \\ \boldsymbol{M}_{n_2}(\boldsymbol{a}, \boldsymbol{b}) &= (M_{n_2}^{(1)}(a_1, b_1), \ldots, M_{n_2}^{(p)}(a_p, b_p))', \end{aligned}$$

where $M_{n_1}^{(j)}(a_j, b_j) = \sum_{i=1}^{n_j} \psi\left(\frac{X_{j_i} - a_j - b_j c_{j_i}}{S_n}\right)$ and $M_{n_2}^{(j)}(a_j, b_j) = \sum_{i=1}^{n_j} c_{j_i}\psi\left(\frac{X_{j_i} - a_j - b_j c_{j_i}}{S_n}\right)$, with $\boldsymbol{a} = (a_1, \ldots, a_p)'$ and $\boldsymbol{b} = (b_1, \ldots, b_p)'$ being vectors of real numbers.

Let $\tilde{\boldsymbol{\beta}}$ be the constrained M-estimator of $\boldsymbol{\beta}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0 = (\theta_{0_1}, \ldots, \theta_{0_p})'$ is a vector of fixed real numbers, that is, $\tilde{\boldsymbol{\beta}}$ is the solution of $\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{b}) = \mathbf{0}$ and it may conveniently be expressed as

$$\tilde{\boldsymbol{\beta}} = [\sup\{\boldsymbol{b} : \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{b}) > 0\} \ + \ \inf\{\boldsymbol{b} : \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{b}) < 0\}]/2. \tag{5.2.5}$$

Similarly, let $\tilde{\boldsymbol{\theta}}$ be the constrained M-estimator of $\boldsymbol{\theta}$ when $\boldsymbol{\beta} = \beta_0 \mathbf{1}_p$ (β_0 is a fixed number), that is, $\tilde{\boldsymbol{\theta}}$ is the solution of $\boldsymbol{M}_{n_1}(\boldsymbol{a}, \beta_0\mathbf{1}_p) = \mathbf{0}$ and can conveniently be expressed as

$$\tilde{\boldsymbol{\theta}} = [\sup\{\boldsymbol{a} : \boldsymbol{M}_{n_1}(\boldsymbol{a}, \beta_0\mathbf{1}_p) > 0\} \ + \ \inf\{\boldsymbol{a} : \boldsymbol{M}_{n_1}(\boldsymbol{a}, \beta_0\mathbf{1}_p) < 0\}]/2. \tag{5.2.6}$$

The below theorem is used to derive the statistical tests proposed in the next Section.

Theorem 5.2.1 *Given the asymptotic properties results (i) and (ii) in Appendix B.3, as $n \to \infty$,*

$$\textit{(i) under } H_0^{(1)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \;\; n^{-\frac{1}{2}} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \xrightarrow{d} N_p(\mathbf{0}, \sigma_0^2 \boldsymbol{\Lambda}_0^\star), \tag{5.2.7}$$

$$\textit{(ii) under } H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\beta} = \beta_0 \mathbf{1}_p,$$
$$n^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p) \\ \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p) \end{pmatrix} \xrightarrow{d} N_{2p} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_2 \end{pmatrix} \right], \tag{5.2.8}$$

$$\textit{(iii) under } H_0^\star : \boldsymbol{\beta} = \beta_0 \mathbf{1}_p, \;\; n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0 \mathbf{1}_p) \xrightarrow{d} N_p(\mathbf{0}, \sigma_0^2 \boldsymbol{\Lambda}_2^\star), \tag{5.2.9}$$

where $N_p(\cdot\,, \cdot\,)$ represents a p-variate normal distribution with appropriate parameters. Here, $\boldsymbol{\Lambda}_0 = Diag(\lambda_1, \ldots, \lambda_p)$, $\boldsymbol{\Lambda}_{12} = Diag(\lambda_1 \bar{c}_1, \ldots, \lambda_p \bar{c}_p)$, $\boldsymbol{\Lambda}_{21} = \boldsymbol{\Lambda}_{12}$, $\boldsymbol{\Lambda}_0^\star = \boldsymbol{\Lambda}_0 - \boldsymbol{\Lambda}_{12} \boldsymbol{\Lambda}_2^{-1} \boldsymbol{\Lambda}_{21} = Diag(\lambda_1 C_1^{\star 2}/(C_1^{\star 2} + \bar{c}_1^2), \ldots, \lambda_p C_p^{\star 2}/(C_p^{\star 2} + \bar{c}_p^2))$ and $\boldsymbol{\Lambda}_2 = Diag(\lambda_1 (C_1^{\star 2} + \bar{c}_1^2), \ldots, \lambda_p (C_p^{\star 2} + \bar{c}_p^2))$, $\boldsymbol{\Lambda}_2^\star = \boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_{12} \boldsymbol{\Lambda}_0^{-1} \boldsymbol{\Lambda}_{21} = Diag$ $(\lambda_1 C_1^{\star 2}, \ldots, \lambda_p C_p^{\star 2})$.

The proof of Theorem 5.2.1 is given in Appendix B.3.

5.3 The UT, RT and PTT

5.3.1 The Unrestricted Test (UT)

If $\boldsymbol{\beta}$ is unspecified, ϕ_n^{UT} is the test function of $H_0^{(1)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)}: \boldsymbol{\theta} > \boldsymbol{\theta}_0$. We consider the test statistic

$$T_n^{UT} = n^{-1} \frac{\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}})' \boldsymbol{\Lambda}_0^{\star -1} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}})}{{S_n^{(1)}}^2},$$

where $\tilde{\boldsymbol{\beta}}$ (given in equation (5.2.5)) is a constrained M-estimator of $\boldsymbol{\beta}$ under $H_0^{(1)}$ and ${S_n^{(1)}}^2 = \frac{1}{n} \sum_{j=1}^{p} \sum_{i=1}^{n_j} \psi^2 \left(\frac{X_{j_i} - \theta_{0j} - \tilde{\beta}_j c_{j_i}}{S_n} \right)$. It follows from (5.2.7) that T_n^{UT} is χ_p^2 (chi-square distribution with p degrees of freedom).

Let ℓ_{n,α_1}^{UT} be the critical value of T_n^{UT} at the α_1 level of significance. So, for the test function $\phi_n^{UT} = I(T_n^{UT} > \ell_{n,\alpha_1}^{UT})$, the power function of the UT becomes $\Pi_n^{UT}(\boldsymbol{\theta}) = E(\phi_n^{UT} | \boldsymbol{\theta}) = P(T_n^{UT} > \ell_{n,\alpha_1}^{UT} | \boldsymbol{\theta})$, where $I(A)$ is an indicator function of the set A. It takes value 1 if A occurs, otherwise it is 0.

5.3.2 The Restricted Test (RT)

If $\boldsymbol{\beta} = \beta_0 \mathbf{1}_p$, the test function for testing $H_0^{(1)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_A^{(1)}: \boldsymbol{\theta} > \boldsymbol{\theta}_0$ is ϕ_n^{RT}. The proposed test statistic is

$$T_n^{RT} = n^{-1} \frac{\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p)' {\boldsymbol{\Lambda}_0}^{-1} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p)}{{S_n^{(2)}}^2},$$

where ${S_n^{(2)}}^2 = \frac{1}{n} \sum_{j=1}^{p} \sum_{i=1}^{n_j} \psi^2 \left(\frac{X_{j_i} - \theta_{0j} - \beta_0 c_{j_i}}{S_n} \right)$. It follows from equation (5.2.8) that for large n, T_n^{RT} is χ_p^2 under $H_0^{(2)}: \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\beta} = \beta_0 \mathbf{1}_p$. Again, let ℓ_{n,α_2}^{RT} be the critical value of T_n^{RT} at the α_2 level of significance. Thus, for the test function $\phi_n^{RT} = I(T_n^{RT} > \ell_{n,\alpha_2}^{RT})$, the power function of the RT becomes $\Pi_n^{RT}(\boldsymbol{\theta}) = E(\phi_n^{RT} | \boldsymbol{\theta}) = P(T_n^{RT} > \ell_{n,\alpha_2}^{RT} | \boldsymbol{\theta})$.

5.3.3 The Pre-test (PT)

For the pre-test on the slope, the test function of $H_0^\star : \boldsymbol{\beta} = \beta_0 \mathbf{1}_p$ against $H_A^\star : \boldsymbol{\beta} > \beta_0 \mathbf{1}_p$ is ϕ_n^{PT}. The proposed test statistic is

$$T_n^{PT} = n^{-1} \frac{\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0 \mathbf{1}_p)' \boldsymbol{\Lambda}_2^{\star -1} \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0 \mathbf{1}_p)}{S_n^{(3)^2}},$$

where $\tilde{\boldsymbol{\theta}}$ (given in equation (5.2.6)) is a constrained M-estimator of $\boldsymbol{\theta}$ and $S_n^{(3)^2} = \frac{1}{n} \sum_{j=1}^{p} \sum_{i=1}^{n_j} \psi^2 \left(\frac{X_{j_i} - \tilde{\theta}_j - \beta_0 c_{j_i}}{S_n} \right)$. It follows from equation (5.2.9) that as $n \to \infty$, $T_n^{PT} \xrightarrow{d} \chi_p^2$ under $H_0^\star$.

5.3.4 The Pre-test Test (PTT)

We are now in the position to formulate ϕ_n^{PTT} for testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}$. Since the PTT is a choice between the RT and the UT, define,

$$\phi_n^{PTT} = I[(T_n^{PT} < \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{RT})], \quad (5.3.1)$$

where ℓ_{n,α_3}^{PT} is the critical value of T_n^{PT} at the α_3 level of significance. The power function of the PTT is given by

$$\Pi_n^{PTT}(\boldsymbol{\theta}) = E(\phi_n^{PTT} | \boldsymbol{\theta}) \quad (5.3.2)$$

and the size of the PTT is obtained by substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ in equation (5.3.2).

5.4 Asymptotic Distributions under Local Alternatives

In this Section, the asymptotic distributions of the UT, RT, PT and PTT are derived under a sequence of local alternative hypotheses, $\{K_n^\star\}$ (see below). These distributions are essential to obtain the power functions of the UT, RT and PTT. To derive the power function of the PTT, the joint distributions of $\left[T_n^{UT}, T_n^{PT}\right]$ and $\left[T_n^{RT}, T_n^{PT}\right]$ are required.

Theorem 5.4.1 *Let $\{K_n^\star\}$ be a sequence of alternative hypotheses, where*

$$K_n^\star : (\boldsymbol{\theta}, \boldsymbol{\beta}) = (\boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0\mathbf{1}_p + n^{-\frac{1}{2}}\boldsymbol{\delta}_2), \tag{5.4.1}$$

with $\boldsymbol{\delta}_1 = n^{\frac{1}{2}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) > \mathbf{0}$ and $\boldsymbol{\delta}_2 = n^{\frac{1}{2}}(\boldsymbol{\beta}-\beta_0\mathbf{1}_p) > \mathbf{0}$. Here, $\boldsymbol{\delta}_1 = (\delta_{1_1}, \dots, \delta_{1_p})'$, $\boldsymbol{\delta}_2 = (\delta_{2_1}, \dots, \delta_{2_p})'$ are vectors of fixed real numbers. Under $\{K_n^\star\}$, for large sample,

(i)

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p) \end{bmatrix} \xrightarrow{d} N_{2p} \left[\begin{pmatrix} \gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2) \\ \gamma\boldsymbol{\Lambda}_2^\star\boldsymbol{\delta}_2 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2^\star \end{pmatrix}\right], \tag{5.4.2}$$

(ii)

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p) \end{bmatrix} \xrightarrow{d} N_{2p} \left[\begin{pmatrix} \gamma\boldsymbol{\Lambda}_0^\star\boldsymbol{\delta}_1 \\ \gamma\boldsymbol{\Lambda}_2^\star\boldsymbol{\delta}_2 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0^\star & \boldsymbol{\Lambda}_{12}^\star \\ \boldsymbol{\Lambda}_{12}^\star & \boldsymbol{\Lambda}_2^\star \end{pmatrix}\right], \tag{5.4.3}$$

where $\boldsymbol{\Lambda}_{12}^\star = -\boldsymbol{\Lambda}_{12} + \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_2^{-1}\boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{21} =$ Diag $(-\lambda_1\bar{c}_1^2C_1^{\star 2}/(C_1^{\star 2} + \bar{c}_1^2), \dots, -\lambda_p\bar{c}_p^2C_p^{\star 2}/\ (C_p^{\star 2} + \bar{c}_p^2))$.

See Appendix B.3 for the proof of Theorem 5.4.1.

Theorem 5.4.2 *Under $\{K_n^\star\}$, asymptotically (T_n^{RT}, T_n^{PT}) are independently distributed as a bivariate non-central chi-square distribution with p degrees of freedom (d.f.) and (T_n^{UT}, T_n^{PT}) are distributed as a correlated bivariate non-central* *chi-square distribution with p d.f. and non-centrality parameters,*

$$\theta^{UT} = (\gamma\boldsymbol{\Lambda}_0^\star\boldsymbol{\delta}_1)'\,\boldsymbol{\Lambda}_0^{\star-1}\,(\gamma\boldsymbol{\Lambda}_0^\star\boldsymbol{\delta}_1)\,/\sigma_0^2, \tag{5.4.4}$$

$$\theta^{RT} = [\gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2)]'\,\boldsymbol{\Lambda}_0{}^{-1}\,[\gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2)]\,/\sigma_0^2, \tag{5.4.5}$$

$$\theta^{PT} = (\gamma\boldsymbol{\Lambda}_2^\star\boldsymbol{\delta}_2)'\,\boldsymbol{\Lambda}_2^{\star-1}\,(\gamma\boldsymbol{\Lambda}_2^\star\boldsymbol{\delta}_2)\,/\sigma_0^2. \tag{5.4.6}$$

Proof The proof of this theorem is directly obtained using Theorem 5.4.1 and Theorem 1.4.1 of Muirhead (1982).

5.5 Asymptotic Properties for UT, RT and PTT

Using the results in Section 5.4, under $\{K_n^\star\}$, the asymptotic power functions for the UT, RT and PT which are denoted by $\Pi^h(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ for h any of the UT, RT and PT, are defined as

$$\Pi^h(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \lim_{n\to\infty} \Pi_n^h(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \lim_{n\to\infty} P(T_n^h > \ell_{n,\alpha_\nu}^h | K_n^\star) = 1 - G_p(\chi^2_{p,\alpha_\nu}; \theta^h), \tag{5.5.1}$$

where $G_p(\chi^2_{p,\alpha_\nu}, \theta^h)$ is the cdf of the non-central chi-square distribution with p degrees of freedom and non-centrality parameter θ^h. The level of significance, α_ν, where $\nu = 1, 2, 3$, is chosen together with the critical values ℓ_{n,α_ν}^h for the UT, RT and PT. Here, $\chi^2_{p,\alpha}$ is the upper $100\alpha\%$ critical value of a central chi-square distribution and $\ell_{n,\alpha_1}^{UT} \to \chi^2_{p,\alpha_1}$ under $H_0^{(1)}$, $\ell_{n,\alpha_2}^{RT} \to \chi^2_{p,\alpha_2}$ under $H_0^{(2)}$ and $\ell_{n,\alpha_3}^{PT} \to \chi^2_{p,\alpha_3}$under $H_0^\star$.

For testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}$, using equation (5.3.1) and the results in Section 5.4, the asymptotic power function for the PTT under $\{K_n^\star\}$

is given as

$$
\begin{aligned}
& \Pi^{PTT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) \\
= & \lim_{n\to\infty} P(T_n^{PT} \leq \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT} | K_n^{\star}) + \\
& \lim_{n\to\infty} P(T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{UT} | K_n^{\star}) \\
= & G_p(\chi^2_{p,\alpha_3}; \theta^{PT})\{1 - G_p(\chi^2_{p,\alpha_2}; \theta^{RT})\} + \int_{\chi^2_{p,\alpha_1}} \int_{\chi^2_{p,\alpha_3}} \tilde{\phi}(w_1, w_2) dw_1 dw_2,
\end{aligned}
\tag{5.5.2}
$$

where $\tilde{\phi}(\cdot)$ is the density function of a bivariate non-central chi-square distribution with p degrees of freedom, non-centrality parameters θ^{UT} and θ^{PT}, and correlation coefficient $-1 < \rho < 1$. The probability integral in (5.5.2) is given by

$$
\begin{aligned}
& \int_{\chi^2_{p,\alpha_1}} \int_{\chi^2_{p,\alpha_3}} \tilde{\phi}(w_1, w_2) dw_1 dw_2 \\
= & \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{\kappa_1=0}^{\infty} \sum_{\kappa_2=0}^{\infty} (1-\rho^2)^p \frac{\Gamma(\frac{p}{2}+j)}{\Gamma(\frac{p}{2})j!} \frac{\Gamma(\frac{p}{2}+k)}{\Gamma(\frac{p}{2})k!} \rho^{2(j+k)} \\
& \times \left[1 - \gamma^{\star}\left(\frac{p}{2} + j + \kappa_1, \frac{\chi^2_{p,\alpha_1}}{2(1-\rho^2)}\right)\right] \left[1 - \gamma^{\star}\left(\frac{p}{2} + k + \kappa_2, \frac{\chi^2_{p,\alpha_3}}{2(1-\rho^2)}\right)\right] \\
& \times \frac{e^{-\theta^{UT}/2}(\theta^{UT}/2)^{\kappa_1}}{\kappa_1!} \frac{e^{-\theta^{PT}/2}(\theta^{PT}/2)^{\kappa_2}}{\kappa_2!}.
\end{aligned}
\tag{5.5.3}
$$

Here, $\gamma^{\star}(v, d) = \int_0^d x^{v-1} e^{-x} / \Gamma(v) dx$ is the incomplete gamma function. Take $\rho^2 = \sum_{j=1}^{p} \frac{1}{p} \rho_j^2$, the mean correlation, where $\rho_j = -c_j / \sqrt{C_j^{\star 2} + \bar{c}_j^2}$ is the correlation coefficient between $\left(M_{n_1}^{(j)}(\theta_{0_j}, \tilde{\beta}_j), M_{n_2}^{(j)}(\tilde{\theta}_j, \beta_0)\right)$. For details on the evaluation of the bivariate integral, see Yunus and Khan (2009). The density function of the bivariate noncentral chi-square distribution given above, is a mixture of the bivariate central chi-square distribution of two central chi-square random variables, (see Gunst and Webster, 1973, Wright and Kennedy, 2002) with the probabilities from the Poisson distribution.

Then, write the equation (5.5.3) as

$$\begin{aligned}
&\int_{\chi^2_{p,\alpha_1}}^{\infty}\int_{\chi^2_{p,\alpha_3}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2 \\
&= 1-H_p(\chi^2_{p,\alpha_1};\theta^{UT})-H_p(\chi^2_{p,\alpha_3};\theta^{PT})+\int_0^{\chi^2_{p,\alpha_1}}\int_0^{\chi^2_{p,\alpha_3}}\tilde{\phi}(w_1,w_2)dw_1dw_2,
\end{aligned} \tag{5.5.4}$$

with $H_p(\chi^2_{p,\alpha_1};\theta^{UT})=\sum_{j=0}^{\infty}\sum_{\kappa_1=0}^{\infty}\frac{(1-\rho^2)^{\frac{p}{2}}\Gamma(\frac{p}{2}+j)\rho^{2j}}{\Gamma(\frac{p}{2})j!}\gamma^{\star}\left(\frac{p}{2}+j+\kappa_1,\frac{\chi^2_{p,\alpha_1}}{2(1-\rho^2)}\right)e^{-\frac{\theta^{UT}}{2}}\left(\frac{\theta^{UT}}{2}\right)^{\kappa_1}/\kappa_1!$ and $H_p(\chi^2_{p,\alpha_3};\theta^{PT})=\sum_{k=0}^{\infty}\sum_{\kappa_2=0}^{\infty}\frac{(1-\rho^2)^{\frac{p}{2}}\Gamma(\frac{p}{2}+k)\rho^{2k}}{\Gamma(\frac{p}{2})k!}\gamma^{\star}\left(\frac{p}{2}+k+\kappa_2,\frac{\chi^2_{p,\alpha_3}}{2(1-\rho^2)}\right)\frac{e^{-\frac{\theta^{PT}}{2}}\left(\frac{\theta^{PT}}{2}\right)^{\kappa_2}}{\kappa_2!}$. Note, $H_p(\chi^2_{p,\alpha_1};\theta^{UT})\geq G_p(\chi^2_{p,\alpha_1};\theta^{UT})$ and $H_p(\chi^2_{p,\alpha_3};\theta^{PT})\geq G_p(\chi^2_{p,\alpha_3};\theta^{PT})$. Equality is achieved when $\rho=0$, or when $\boldsymbol{\delta}_1=\mathbf{0}$ and $\boldsymbol{\delta}_2=\mathbf{0}$.

Consider all cases when $\bar{c}_j>0$ and $\rho\neq 0$. Using equation 5.5.4, we write equation (5.5.2) as

$$\Pi^{PTT}(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)\ \leq\ \Pi^{RT}(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)[1-\Pi^{PT}(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)]+\Pi^{UT}(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)\Pi^{PT}(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2). \tag{5.5.5}$$

Equality in equation (5.5.5) is achieved when both $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are $\mathbf{0}$.

Wright and Kennedy (2002) computed the cumulative distribution function (cdf) for the bivariate central chi-square distribution. In their paper, as the correlation coefficient ρ gets larger the cdf grows larger too. In the same manner, the cdf for the bivariate noncentral chi-square distribution increases as the correlation coefficient increases (c.f. Yunus and Khan, 2009). Rewrite equation (5.5.3) as

$$\begin{aligned}
&\int_{\chi^2_{p,\alpha_1}}^{\infty}\int_{\chi^2_{p,\alpha_3}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2 \\
&= 1-H_p(\chi^2_{p,\alpha_1};\theta^{UT})-H_p(\chi^2_{p,\alpha_3};\theta^{PT})+\int_0^{\chi^2_{p,\alpha_1}}\int_0^{\chi^2_{p,\alpha_3}}\tilde{\phi}(w_1,w_2)dw_1dw_2.
\end{aligned} \tag{5.5.6}$$

From equations (5.5.1)-(5.5.6), we observe the followings:

(i) $\Pi^{RT} \geq \Pi^{UT}$, since $\theta^{RT} \geq \theta^{UT}$ when $\bar{c}_j \geq 0$ or $\boldsymbol{\delta}_2 = \mathbf{0}$ and $\alpha_1 = \alpha_2$. We conclude that the asymptotic size of the RT is larger than that of the UT but the asymptotic power of the UT is smaller than that of the RT.

(ii) Obviously from equation (5.5.5), $\Pi^{PTT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) \leq \Pi^{RT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) - \Pi^{PT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ v_2 for $0 \leq v_2 < 1$, and it follows that $\Pi^{PTT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) \leq \Pi^{RT}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ for any $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$. Equality holds when both $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are $\mathbf{0}$.

(iii) When $\boldsymbol{\delta}_2$ is not large but not $\mathbf{0}$ and $\boldsymbol{\delta}_1$ is sufficiently large, the first term on the right hand side of the equation (5.5.2) becomes $G(\chi^2_{p,\alpha_3}; \theta^{PT})$ because θ^{RT} is sufficiently large. The second and fourth terms on the right hand side of the equation (5.5.4) becomes 0 because θ^{UT} is large. Also, note that $H_p(\chi^2_{p,\alpha_3}; \theta^{PT}) > G_p(\chi^2_{p,\alpha_3}; \theta^{PT})$. So, $\Pi^{PTT} < \Pi^{UT} = 1$ for sufficiently large $\boldsymbol{\delta}_1$ and not so large $\boldsymbol{\delta}_2$ $(\neq \mathbf{0})$. Thus, when $\boldsymbol{\delta}_2$ $(\neq \mathbf{0})$ is not large and $\boldsymbol{\delta}_1$ is sufficiently large, then, $\Pi^{PTT} < \Pi^{UT} = 1$.

(iv) When $\boldsymbol{\delta}_2 = \mathbf{0}$ and $\boldsymbol{\delta}_1$ is sufficiently large, the first term on the right hand side of the equation (5.5.2) becomes $1 - \alpha_3$ because θ^{RT} is large. Both $H_p(\chi^2_{p,\alpha_1}; \theta^{UT})$ and $\int_0^{\chi^2_{p,\alpha_1}} \int_0^{\chi^2_{p,\alpha_3}} \tilde{\phi}(w_1, w_2) dw_1 dw_2$ of the equation (5.5.4) become 0 while $H_p(\chi^2_{p,\alpha_3};\ \theta^{PT})$ becomes $1 - \alpha_3$. So, $\Pi^{PTT} = \Pi^{UT} = 1$ when $\boldsymbol{\delta}_1$ is sufficiently large and $\boldsymbol{\delta}_2 = \mathbf{0}$. Thus, when $\boldsymbol{\delta}_2 = \mathbf{0}$ and $\boldsymbol{\delta}_1$ is sufficiently large, then, $\Pi^{PTT} = \Pi^{UT} = 1$.

(v) When $\boldsymbol{\delta}_1$ $(\neq \mathbf{0})$ is not large but $\boldsymbol{\delta}_2$ is sufficiently large, then, $\Pi^{PTT} < \Pi^{UT} = 1$.

(vi) Let $\alpha_1 = \alpha$. When $\boldsymbol{\delta}_1 = \mathbf{0}$ and $\boldsymbol{\delta}_2$ is sufficiently large, then, $\Pi^{PTT} = \Pi^{UT} = \alpha$.

(vii) Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. When both $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are $\mathbf{0}$, then $\Pi^{PTT} = \Pi^{UT} = \alpha$.

5.6 Illustrative Example

Consider two simple linear regression lines,

$$\begin{aligned} \boldsymbol{X}_{1_{n_1}} &= \theta_1 \mathbf{1}_{n_1} + \beta_1 \boldsymbol{c}_1 + \boldsymbol{\varepsilon}_1, \text{ and} \\ \boldsymbol{X}_{2_{n_2}} &= \theta_2 \mathbf{1}_{n_2} + \beta_2 \boldsymbol{c}_2 + \boldsymbol{\varepsilon}_2. \end{aligned} \tag{5.6.1}$$

The power functions given in equations (5.5.1) and (5.5.2) are computed for their graphical view. For $p = 2$, the non-centrality parameters for the UT, RT and PT are respectively

$$\begin{aligned}
\theta^{UT} &= \begin{bmatrix} \xi_{1_1}\lambda_1 \frac{C_1^{\star 2}}{C_1^{\star 2}+\bar{c}_1^2} \\ \xi_{1_2}\lambda_2 \frac{C_2^{\star 2}}{C_2^{\star 2}+\bar{c}_2^2} \end{bmatrix}' \begin{bmatrix} \lambda_1 \frac{C_1^{\star 2}}{C_1^{\star 2}+\bar{c}_1^2} & 0 \\ 0 & \lambda_2 \frac{C_2^{\star 2}}{C_2^{\star 2}+\bar{c}_2^2} \end{bmatrix}^{-1} \begin{bmatrix} \xi_{1_1}\lambda_1 \frac{C_1^{\star 2}}{C_1^{\star 2}+\bar{c}_1^2} \\ \xi_{1_2}\lambda_2 \frac{C_2^{\star 2}}{C_2^{\star 2}+\bar{c}_2^2} \end{bmatrix}, \\
\theta^{RT} &= \begin{bmatrix} \xi_{1_1}\lambda_1 + \xi_{2_1}\lambda_1\bar{c}_1 \\ \xi_{1_2}\lambda_2 + \xi_{2_2}\lambda_2\bar{c}_1 \end{bmatrix}' \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} \xi_{1_1}\lambda_1 + \xi_{2_1}\lambda_1\bar{c}_1 \\ \xi_{1_2}\lambda_2 + \xi_{2_2}\lambda_2\bar{c}_1 \end{bmatrix} \text{ and} \\
\theta^{PT} &= \begin{bmatrix} \xi_{2_1}\lambda_1 C_1^{\star 2} \\ \xi_{2_2}\lambda_2 C_2^{\star 2} \end{bmatrix}' \begin{bmatrix} \lambda_1 C_1^{\star 2} & 0 \\ 0 & \lambda_2 C_2^{\star 2} \end{bmatrix}^{-1} \begin{bmatrix} \xi_{2_1}\lambda_1 C_1^{\star 2} \\ \xi_{2_2}\lambda_2 C_2^{\star 2} \end{bmatrix},
\end{aligned}$$

where $\xi_{k_l} = \delta_{k_l}\gamma/\sigma_0$ for $k, l = 1, 2$ and $\delta_{1_l} = \sqrt{n}(\theta_l - \theta_{0_l})$ and $\delta_{2_l} = \sqrt{n}(\beta_l - \beta_{0_l})$.

A special case of the two sample problem (Saleh, 2006, p.67) is considered with $n_j = n_{j_1} + n_{j_2}$ for $j = 1, 2$, $n_{j_1}/n_j \to 1 - P$, $c_{j_1} = \ldots = c_{j_{n_1}} = 0$ and $c_{j_{n_1+1}} = \ldots = c_{j_n} = 1$. So $\bar{c}_j = 1 - P$ and $C_j^{\star 2} = P(1 - P)$. In this example, let $P = 0.5$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$. Also, let $n_1, n_2 = 50$ so $n = n_1 + n_2 = 100$. As a results, the correlation coefficient $\rho_j, j = 1, 2$ for both regression lines are the same since $\bar{c}_1^2 = \bar{c}_2^2 = \bar{c}^2$ for both samples, (X_{n_1}, c_1) and (X_{n_2}, c_2), of the two regression lines. Note, in plotting the power functions for the PTT, a bivariate non-central chi-square distribution is used.

Let $\xi_{1_1} = \xi_{1_2} = a$ and $\xi_{2_1} = \xi_{2_2} = b$. Figure 5.1 shows the power of the test against b at selected values of ξ_{1_1} and ξ_{1_2}. A test with a higher size and lower power is a test which makes a small probability of Type I and Type II errors. In Figure 5.1, except for small b, the UT has the smallest size and the PTT has

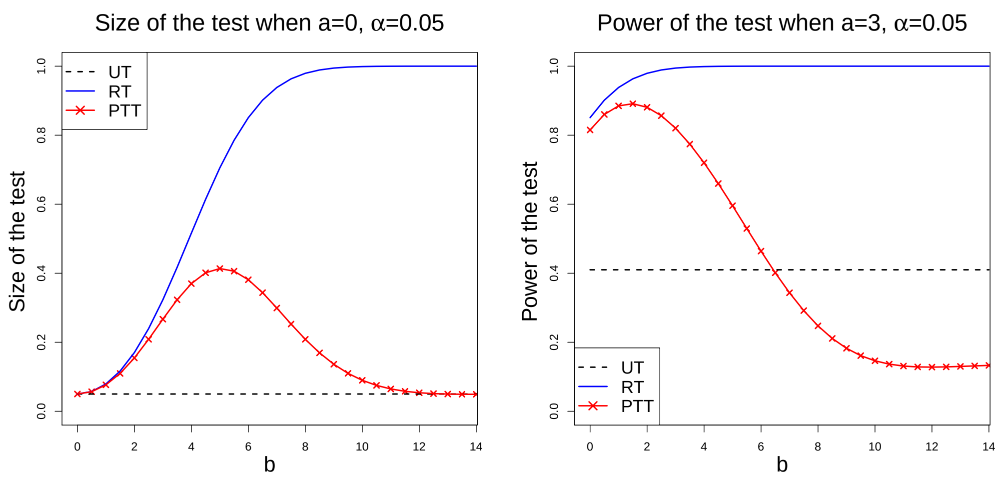


Figure 5.1: Graphs of power functions as a function of b $(=\xi_{2_1}=\xi_{2_2})$ for selected values of ξ_{1_1} and ξ_{1_2} with $\bar{c}>0$ and $\alpha_1=\alpha_2=\alpha_3=\alpha=0.05$. Here, $\xi_{k_l}=\delta_{k_l}\gamma/\sigma_0, k,l=1,2$.

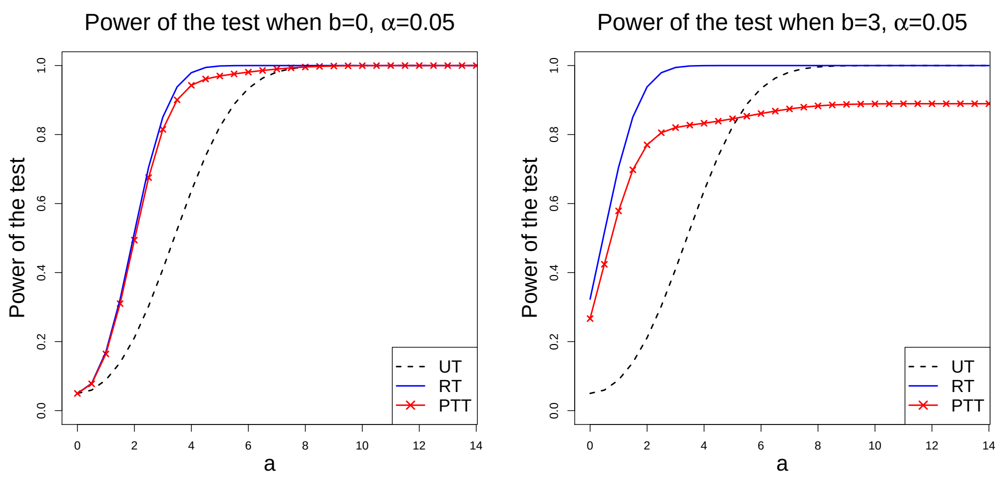


Figure 5.2: Graphs of power functions as a function of a $(=\xi_{1_2}=\xi_{1_2})$ for selected values of ξ_{2_1} and ξ_{2_2} with $\bar{c}>0$ and $\alpha_1=\alpha_2=\alpha_3=\alpha=0.05$. Here, $\xi_{k_l}=\delta_{k_l}\gamma/\sigma_0, k,l=1,2$.

a smaller size than that of the RT. The RT has the largest power as b grows. The PTT has higher power than that of the UT except for large b.

In Figure 5.2, the power of the UT, RT and PTT is plotted against a at selected values of ξ_{2_1} and ξ_{2_2}. As a grows large, the power of all tests grows large too. Although the power of the UT and RT is increasing to 1 as a is increasing, the power of the PTT is increasing to a value that is less than 1. The analytical findings in the previous Section support these graphical results.

5.7 Discussion and Conclusion

The sampling distributions of the UT, RT and PT follow a univariate noncentral chi-square distribution under the alternative hypothesis when the sample size is large. However, that of the PTT is a bivariate noncentral chi-square distribution as there is a correlation between the UT and PT. Note that there is no such correlation between the RT and PT.

The size of the RT reaches 1 as b (a function of the difference between the true and suspected values of the slopes) increases. This means the RT does not satisfy the asymptotic level constraint, so it is not a valid test. The UT has the smallest constant size; however, it has the smallest power as well, except for very large values of b, that is, when $b > q$, where q is some positive number. Thus, the UT fails to achieve the highest power and lowest size simultaneously. The PTT has a smaller size than the RT and its size does not reach 1 as b increases. It also has higher power than the UT, except for $b > q$.

Therefore, if the prior information is not far away from the true value, that is, b is near 0 (small or moderate), the PTT has a smaller size than the RT and more power than the UT. So, the PTT is a better compromise between the two extremes. Since the prior information comes from previous experience or expert assessment, it is reasonable to expect b should not be too far from 0, although it may not be 0, and hence the PTT achieves a reasonable dominance over the other two tests in a more realistic situation.

Chapter 6

Multiple Linear Regression Model

6.1 Introduction

Let $X_i,\ i = 1, \ldots, n$, be n observable response variables of a multiple regression model,

$$X_i = \boldsymbol{\beta}'\boldsymbol{c}_i + e_i, \tag{6.1.1}$$

where $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ is a p-dimensional row vector of unknown regression parameters, $\boldsymbol{c}_i' = (c_{1i}, \ldots,\ c_{pi})$ is a p-dimensional row vector of known real constants of the independent variables, e_i is the error term which is identically and independently distributed symmetric about 0 with a distribution function, $F_i,\ i = 1, \ldots, n$. The vector of p-regression parameters can be expressed as $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1',\ \boldsymbol{\beta}_2')$, where $\boldsymbol{\beta}_1'$ is a sub-vector of order r and $\boldsymbol{\beta}_2'$ is a sub-vector of dimension s such that $r + s = p$. Similarly, partition $\boldsymbol{c}_i'$ as $(\boldsymbol{c}_{i1}',\ \boldsymbol{c}_{i2}')$ with $\boldsymbol{c}_{i1}' = (c_{1i}, \ldots,\ c_{ri})$ and $\boldsymbol{c}_{i2}' = (c_{(r+1)i}, \ldots, c_{pi})$.

Consider testing the significance of the sub-vector $\boldsymbol{\beta}_1$ under three conditions on the values of the sub-vector $\boldsymbol{\beta}_2$: (i) unspecified (ii) specified and fixed (iii) uncertain. For case (i), we want to test $H_0^{(1)} : \boldsymbol{\beta}_1 = \mathbf{0}$ against $H_A^{(1)} : \boldsymbol{\beta}_1 > \mathbf{0}$ with test function, ϕ_n^{UT}. This test is called the unrestricted test (UT). For case (ii), the test for testing $H_0^{(1)} : \boldsymbol{\beta}_1 = \mathbf{0}$ against $H_A^{(1)} : \boldsymbol{\beta}_1 > \mathbf{0}$ with test function ϕ_n^{RT} is called the restricted test (RT). For case (iii), testing $H_0^{(2)} : \boldsymbol{\beta}_2 = \mathbf{0}$ is recommended to remove the uncertainty of the suspicious values of $\boldsymbol{\beta}_2 = \mathbf{0}$ before testing the significance of $\boldsymbol{\beta}_1$. The testing on $H_0^{(2)} : \boldsymbol{\beta}_2 = \mathbf{0}$ against $H_A^{(2)} : \boldsymbol{\beta}_2 > \mathbf{0}$ with test function ϕ_n^{PT} is known as a pre-test (PT). If the null hypothesis of this pre-test is rejected, the UT is used to test $H_0^{(1)}$, otherwise the RT is used. The ultimate test for testing $H_0^{(1)}$ following a pre-test on $H_0^{(2)}$ is defined as the pre-test test (PTT) and the test function is denoted by ϕ_n^{PTT}.

6.2 The M-estimation

Given an absolutely continuous function $\rho : \Re \to \Re$, M-estimator of $\boldsymbol{\beta}$ is defined as the solution of minimizing the objective function

$$\sum_{i=1}^{n} \rho\left(\frac{X_i - \boldsymbol{\beta}'\boldsymbol{c}_i}{S_n}\right) \tag{6.2.1}$$

with respect to $\boldsymbol{\beta} \in \Re_p$. Here S_n is an appropriate scale statistic for some functional $S = S(F) > 0$. If $\psi = \rho'$, then the M-estimator of $\boldsymbol{\beta}$ is the solution of the system of equations,

$$\sum_{i=1}^{n} \boldsymbol{c}_i \psi\left(\frac{X_i - \boldsymbol{\beta}'\boldsymbol{c}_i}{S_n}\right) = \mathbf{0}. \tag{6.2.2}$$

For any r and s dimensional column vectors, $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ ($r, s \in \Re$), consider the statistics below

$$\begin{aligned} \boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \boldsymbol{t}_2) &= \sum_{i=1}^{n} \boldsymbol{c}_{i1} \psi\left(\frac{X_i - \boldsymbol{t}_1'\boldsymbol{c}_{i1} - \boldsymbol{t}_2'\boldsymbol{c}_{i2}}{S_n}\right), & (6.2.3) \\ \boldsymbol{M}_{n_2}(\boldsymbol{t}_1, \boldsymbol{t}_2) &= \sum_{i=1}^{n} \boldsymbol{c}_{i2} \psi\left(\frac{X_i - \boldsymbol{t}_1'\boldsymbol{c}_{i1} - \boldsymbol{t}_2'\boldsymbol{c}_{i2}}{S_n}\right). & (6.2.4) \end{aligned}$$

For a nondecreasing $\psi : \Re \to \Re$, let $\tilde{\boldsymbol{\beta}}_2$ be the constrained M-estimator of $\boldsymbol{\beta}_2$ when $\boldsymbol{\beta}_1 = \mathbf{0}$, that is, $\tilde{\boldsymbol{\beta}}_2$ is the solution of $\boldsymbol{M}_{n_2}(\mathbf{0}, \boldsymbol{t}_2) = \mathbf{0}$ and it may be conveniently expressed as

$$\tilde{\boldsymbol{\beta}}_2 = [\sup\{\boldsymbol{t}_2 : \boldsymbol{M}_{n_2}(\mathbf{0}, \boldsymbol{t}_2) > \mathbf{0}\} + \inf\{\boldsymbol{t}_2 : \boldsymbol{M}_{n_2}(\mathbf{0}, \boldsymbol{t}_2) < \mathbf{0}\}]/2 \tag{6.2.5}$$

(c.f. Sen, 1982). Note that for the nondecreasing ψ-function, $\boldsymbol{M}_{n_2}(\mathbf{0}, \boldsymbol{t}_2)$ is decreasing as $\boldsymbol{t}_2$ is increasing (Jurečková and Sen, 1996, p.85). Similarly, let $\tilde{\boldsymbol{\beta}}_1$ be the constrained M-estimator of $\boldsymbol{\beta}_1$ when $\boldsymbol{\beta}_2 = \mathbf{0}$, that is, $\tilde{\boldsymbol{\beta}}_1$ is the solution of $\boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \mathbf{0}) = \mathbf{0}$ and may conveniently be expressed as

$$\tilde{\boldsymbol{\beta}}_1 = [\sup\{\boldsymbol{t}_1 : \boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \mathbf{0}) > \mathbf{0}\} + \inf\{\boldsymbol{t}_1 : \boldsymbol{M}_{n_1}(\boldsymbol{t}_1, \mathbf{0}) < \mathbf{0}\}]/2. \tag{6.2.6}$$

Let

$$\sigma_0^2 = \int_{-\infty}^{\infty} \psi^2\left(\frac{X - \boldsymbol{\beta}'\boldsymbol{c}}{S}\right) dF(X - \boldsymbol{\beta}'\boldsymbol{c}). \tag{6.2.7}$$

Here σ_0^2 is the second moment of $\psi(\cdot)$ while the first moment is zero by assuming F is symmetrically distributed at 0 and ψ is a skew symmetric function.

Theorem 6.2.1 *Given the asymptotic properties of $\boldsymbol{M}_{n_1}(\cdot,\cdot)$ and $\boldsymbol{M}_{n_2}(\cdot,\cdot)$ in equations (B.4.1), (B.4.2) and (B.4.3) in the Appendix B.4, asymptotically,*

$$\text{(i)} \quad n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) \xrightarrow{d} N_r(\boldsymbol{0},\sigma_0^2\boldsymbol{Q}_1^\star) \text{ under } H_0^{(1)}:\boldsymbol{\beta}_1=\boldsymbol{0}, \tag{6.2.8}$$

$$\text{(ii)} \quad n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0}) \xrightarrow{d} N_s(\boldsymbol{0},\sigma_0^2\boldsymbol{Q}_2^\star) \text{ under } H_0^{(2)}:\boldsymbol{\beta}_2=\boldsymbol{0}, \tag{6.2.9}$$

where $\boldsymbol{Q}_1^\star=\boldsymbol{Q}_{11}-\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$ and $\boldsymbol{Q}_2^\star=\boldsymbol{Q}_{22}-\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$. Here, $N_r(\cdot\ ,\cdot\ )$ represents an r-variate normal distribution with appropriate parameters. Take $\boldsymbol{Q}_{11}=\lim_{n\to\infty}\frac{1}{n}\boldsymbol{Q}_{n_{11}}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\boldsymbol{c}_{i1}\boldsymbol{c}_{i1}'$, $\boldsymbol{Q}_{12}=\lim_{n\to\infty}\frac{1}{n}\boldsymbol{Q}_{n_{12}}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\boldsymbol{c}_{i1}\boldsymbol{c}_{i2}'$, $\boldsymbol{Q}_{21}=\lim_{n\to\infty}\frac{1}{n}\boldsymbol{Q}_{n_{21}}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\boldsymbol{c}_{i2}\boldsymbol{c}_{i1}'$ and $\boldsymbol{Q}_{22}=\lim_{n\to\infty}\frac{1}{n}\boldsymbol{Q}_{n_{22}}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\boldsymbol{c}_{i2}\boldsymbol{c}_{i2}'$. Assume that $|\boldsymbol{Q}_{11}|\neq 0$, $|\boldsymbol{Q}_{22}|\neq 0$, $|\boldsymbol{Q}_1^\star|\neq 0$ and $|\boldsymbol{Q}_2^\star|\neq 0$.

See Appendix B.4 for the proof of Theorem 6.2.1.

6.3 The UT, RT and PTT

6.3.1 The Unrestricted Test (UT)

If $\boldsymbol{\beta}_2$ is unspecified, ϕ_n^{UT} is the test function of $H_0^{(1)}:\boldsymbol{\beta}_1=\boldsymbol{0}$ against $H_A^{(1)}:\boldsymbol{\beta}_1>\boldsymbol{0}$. Under $H_0^{(1)}$, $X_i=\boldsymbol{\beta}_2'\boldsymbol{c}_{i2}+e_i$. We consider test statistic

$$L_n^{UT}=\frac{\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2)'{\boldsymbol{Q}_{n_1}^\star}^{-1}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2)}{{S_n^{(1)}}^2},$$

where $\tilde{\boldsymbol{\beta}}_2$ (given in equation (6.2.5)) is a constrained M-estimator of $\boldsymbol{\beta}_2$ under $H_0^{(1)}$. It follows from equation (6.2.8) that L_n^{UT} is χ_r^2 (chi-square distribution with r degrees of freedom) under $H_0^{(1)}$ as $n \to \infty$, with $\boldsymbol{Q}_{n_1}^\star = \boldsymbol{Q}_{n_{11}} - \boldsymbol{Q}_{n_{12}}\boldsymbol{Q}_{n_{22}}^{-1}\boldsymbol{Q}_{n_{21}}$ and ${S_n^{(1)}}^2 = n^{-1}\sum \psi^2\left(\frac{X_i - \tilde{\boldsymbol{\beta}}_2'\boldsymbol{c}_{i2}}{S_n}\right)$.

Let ℓ_{n,α_1}^{UT} be the critical value of L_n^{UT} at the α_1 level of significance. So, for the test function $\phi_n^{UT} = I(L_n^{UT} > \ell_{n,\alpha_1}^{UT})$, the power function of the UT becomes $\tilde{\Pi}_n^{UT}(\boldsymbol{\beta}_1) = E(\phi_n^{UT}|\boldsymbol{\beta}_1) = P(L_n^{UT} > \ell_{n,\alpha_1}^{UT}|\boldsymbol{\beta}_1)$, where $I(A)$ is an indicator function of the set A. It takes value 1 if A occurs, otherwise it is 0.

6.3.2 The Restricted Test (RT)

If $\boldsymbol{\beta}_2 = \boldsymbol{0}$, ϕ_n^{RT} is the test function for testing $H_0^{(1)} : \boldsymbol{\beta}_1 = \boldsymbol{0}$ against $H_A^{(1)} : \boldsymbol{\beta}_1 > \boldsymbol{0}$. The proposed test statistic is

$$L_n^{RT} = \frac{\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0})'{\boldsymbol{Q}_{n_{11}}}^{-1}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0})}{{S_n^{(2)}}^2},$$

where ${S_n^{(2)}}^2 = n^{-1}\sum \psi^2\left(\frac{X_i}{S_n}\right)$. It follows from equation (B.4.3) that for large n, $L_n^{RT} \xrightarrow{d} \chi_r^2$ under $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}, \boldsymbol{\beta}_2 = \boldsymbol{0}$, Again, let ℓ_{n,α_2}^{RT} be the critical value of L_n^{RT} at the α_2 level of significance. So, for the test function $\phi_n^{RT} = I(L_n^{RT} > \ell_{n,\alpha_2}^{RT})$, the power function of the RT becomes $\tilde{\Pi}_n^{RT}(\boldsymbol{\beta}_1) = E(\phi_n^{RT}|\boldsymbol{\beta}_1) = P(L_n^{RT} > \ell_{n,\alpha_2}^{RT}|\boldsymbol{\beta}_1)$.

6.3.3 The Pre-test (PT)

For the pre-test on the slope, ϕ_n^{PT} is the test function for testing $H_0^{(2)} : \boldsymbol{\beta}_2 = \boldsymbol{0}$ against $H_A^{(2)} : \boldsymbol{\beta}_2 > \boldsymbol{0}$. Under $H_0^{(2)}$, $X_i = \boldsymbol{\beta}_1'\boldsymbol{c}_{i1} + e_i$. The proposed test statistic is

$$L_n^{PT} = \frac{\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0})'{\boldsymbol{Q}_{n_2}^\star}^{-1}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0})}{{S_n^{(3)}}^2},$$

where $\tilde{\boldsymbol{\beta}}_1$ (given in equation (6.2.6)) is a constrained M-estimator of $\boldsymbol{\beta}_1$. It follows from equation (6.2.9) that $L_n^{PT} \xrightarrow{d} \chi_s^2$ under $H_0^{(2)}$, where $\boldsymbol{Q}_{n_2}^\star = \boldsymbol{Q}_{n_{22}} -$

$\boldsymbol{Q}_{n_{21}}\boldsymbol{Q}_{n_{11}}^{-1}\boldsymbol{Q}_{n_{12}}$ and ${S_n^{(3)}}^2 = n^{-1}\sum \psi^2\left(\frac{X_i - \tilde{\boldsymbol{\beta}}_1'\boldsymbol{c}_{i1}}{S_n}\right)$.

6.3.4 The Pre-test Test (PTT)

Let ϕ_n^{PTT} be the test function for testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}$. Since the PTT is a choice between the RT and UT, define,

$$\phi_n^{PTT} = I[(L_n^{PT} < \ell_{n,\alpha_3}^{PT}, L_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (L_n^{PT} > \ell_{n,\alpha_3}^{PT}, L_n^{UT} > \ell_{n,\alpha_1}^{UT})], \quad (6.3.1)$$

where ℓ_{n,α_3}^{PT} is the critical value of L_n^{PT} at the α_3 level of significance. The power function of the PTT is given by

$$\tilde{\Pi}_n^{PTT}(\boldsymbol{\beta}_1) = E(\phi_n^{PTT}|\boldsymbol{\beta}_1) \quad (6.3.2)$$

and the size of the PTT is obtained by substituting $\boldsymbol{\beta}_1 = \mathbf{0}$ in equation (6.3.2).

6.4 Asymptotic Distributions of UT, RT, PT and PTT

In this Section, the asymptotic distributions of the UT, RT, PT and PTT are derived under local alternative hypotheses, $\{K_n\}$ (see below). These distributions are essential to obtain the power functions of the UT, RT and PTT. To derive the power function of the PTT, we require to find the joint distributions of $\left[L_n^{UT}, L_n^{PT}\right]$ and $\left[L_n^{RT}, L_n^{PT}\right]$.

Theorem 6.4.1 *Let $\{K_n\}$ be a sequence of local alternative hypotheses, where*

$$K_n : (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, n^{-\frac{1}{2}}\boldsymbol{\lambda}_2), \quad (6.4.1)$$

with $\boldsymbol{\lambda}_1 = n^{\frac{1}{2}}\boldsymbol{\beta}_1 > \mathbf{0}$ and $\boldsymbol{\lambda}_2 = n^{\frac{1}{2}}\boldsymbol{\beta}_2 > \mathbf{0}$ are (fixed) real numbers. Under

$\{K_n\}$, *asymptotically,*

$$\text{(i)} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0}) \end{bmatrix} \xrightarrow{d} N_p \left[\begin{pmatrix} \gamma(\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{12}\boldsymbol{\lambda}_2) \\ \gamma\boldsymbol{Q}_2^\star\boldsymbol{\lambda}_2 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2^\star \end{pmatrix} \right], \quad (6.4.2)$$

$$\text{(ii)} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0}) \end{bmatrix} \xrightarrow{d} N_p \left[\begin{pmatrix} \gamma\boldsymbol{Q}_1^\star\boldsymbol{\lambda}_1 \\ \gamma\boldsymbol{Q}_2^\star\boldsymbol{\lambda}_2 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{Q}_1^\star & \boldsymbol{Q}_{12}^\star \\ \boldsymbol{Q}_{21}^\star & \boldsymbol{Q}_2^\star \end{pmatrix} \right], \quad (6.4.3)$$

where $\boldsymbol{Q}_{12}^\star = \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12} - \boldsymbol{Q}_{12}$, $\boldsymbol{Q}_{21}^\star = \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21} - \boldsymbol{Q}_{21}$ *and* $\gamma = \frac{1}{S}\int_{-\infty}^{\infty}\psi'\left(\frac{X-\boldsymbol{\beta}'\boldsymbol{c}}{S}\right)dF(X-\boldsymbol{\beta}'\boldsymbol{c})$.

The proof of Theorem 6.4.1 is given in Appendix B.4.

Theorem 6.4.2 *Under* $\{K_n\}$, *asymptotically* (L_n^{RT}, L_n^{PT}) *are independently distributed as bivariate noncentral chi-square distribution with* (r,s) *degrees of freedom (d.f.) and* (L_n^{UT}, L_n^{PT}) *are distributed as correlated bivariate noncentral chi-square distribution with* (r,s) *d.f. and noncentrality parameters,*

$$\theta^{UT} = \frac{\gamma^2}{\sigma_0^2}(\boldsymbol{\lambda}_1'\boldsymbol{Q}_1^\star\boldsymbol{\lambda}_1), \quad (6.4.4)$$

$$\theta^{RT} = \frac{\gamma^2}{\sigma_0^2}(\boldsymbol{\lambda}_1'\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_1'\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_2'\boldsymbol{Q}_{21}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2), \quad (6.4.5)$$

$$\theta^{PT} = \frac{\gamma^2}{\sigma_0^2}(\boldsymbol{\lambda}_2'\boldsymbol{Q}_2^\star\boldsymbol{\lambda}_2). \quad (6.4.6)$$

Proof The proof of this theorem is directly obtained using Theorem 6.4.1 and Theorem 1.4.1 of Muirhead (1982).

6.5 Asymptotic Properties for UT, RT and PTT

Using results of Section 6.4, under $\{K_n\}$, the asymptotic power function for the UT is

$$\begin{aligned}\tilde{\Pi}^{UT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)=\lim_{n\to\infty}\tilde{\Pi}_n^{UT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) &= \lim_{n\to\infty}P(L_n^{UT}>\ell_{n,\alpha_1}^{UT}|K_n)\\ &= 1-G_r(\chi^2_{r,\alpha_1};\theta^{UT}),\end{aligned} \tag{6.5.1}$$

the asymptotic power function for the RT is

$$\begin{aligned}\tilde{\Pi}^{RT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)=\lim_{n\to\infty}\tilde{\Pi}_n^{RT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) &= \lim_{n\to\infty}P(L_n^{RT}>\ell_{n,\alpha_2}^{RT}|K_n)\\ &= 1-G_r(\chi^2_{r,\alpha_2};\theta^{RT}),\end{aligned} \tag{6.5.2}$$

and the asymptotic power function for the PT is

$$\begin{aligned}\tilde{\Pi}^{PT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)=\lim_{n\to\infty}\tilde{\Pi}_n^{PT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) &= \lim_{n\to\infty}P(L_n^{PT}>\ell_{n,\alpha_3}^{PT}|K_n)\\ &= 1-G_s(\chi^2_{s,\alpha_3};\theta^{PT}),\end{aligned} \tag{6.5.3}$$

where $G_k(\chi^2_{k,\alpha_\nu};\theta^h)$ is the cumulative density function of the noncentral chi-square distribution with k degrees of freedom (d.f.) and the noncentrality parameter θ^h in which h is any of the UT, RT and PTT. The level of significance, $\alpha_\nu, \nu = 1,2,3$ is chosen together with the critical values ℓ^h_{n,α_ν} for the UT, RT and PT. Here, $\chi^2_{k,\alpha}$ is the upper $100\alpha\%$ critical value of a central chi-square distribution and $\ell^{UT}_{n,\alpha_1}\to\chi^2_{r,\alpha_1}$ under $H_0^{(1)}$, $\ell^{RT}_{n,\alpha_2}\to\chi^2_{r,\alpha_2}$ under H_0 and $\ell^{PT}_{n,\alpha_3}\to\chi^2_{s,\alpha_3}$under $H_0^{(2)}$.

For a large n (fixed), when $\theta^{RT}\geq\theta^{UT}$, we observe from equations (6.5.1) and (6.5.2) that the asymptotic size of the RT is larger than that of the UT but the asymptotic power of the UT is smaller than that of the RT.

For testing $H_0^{(1)}$ following a pre-test on $\boldsymbol{\beta}_2$, using equation (6.3.1) and the results of Section 6.4, the asymptotic power function for the PTT under $\{K_n\}$

is given by

$$
\begin{aligned}
&\tilde{\Pi}^{PTT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \\
= \; & \lim_{n\to\infty} P(L_n^{PT} \le \ell_{n,\alpha_3}^{PT}, L_n^{RT} > \ell_{n,\alpha_2}^{RT} | K_n) \\
& + \lim_{n\to\infty} P(L_n^{PT} > \ell_{n,\alpha_3}^{PT}, L_n^{UT} > \ell_{n,\alpha_1}^{UT} | K_n) \\
= \; & G_s(\chi^2_{s,\alpha_3}; \theta^{PT})\{1 - G_r(\chi^2_{r,\alpha_2}; \theta^{RT})\} + \int_{\chi^2_{r,\alpha_1}}^{\infty} \int_{\chi^2_{s,\alpha_3}}^{\infty} \tilde{\phi}(w_1, w_2) dw_1 dw_2,
\end{aligned}
\tag{6.5.4}
$$

where $\tilde{\phi}(\cdot)$ is the density function of a bivariate noncentral chi-square distribution. It is observed that $G_s(\chi^2_{s,\alpha_3};\ \theta^{PT})$ is decreasing as the value of θ^{PT} is increasing and $1 - G_r(\chi^2_{r,\alpha_2}; \theta^{RT})$ is increasing as the value of θ^{RT} is increasing.

Let

$$
\gamma^{\star}(v, d) = \int_0^d x^{v-1} e^{-x} / \Gamma(v) dx
$$

be the incomplete gamma function.

The probability integral in (6.5.4) is given by

$$
\begin{aligned}
& \int_{\chi^2_{r,\alpha_1}}^{\infty} \int_{\chi^2_{s,\alpha_3}}^{\infty} \tilde{\phi}(w_1, w_2) dw_1 dw_2 \\
= \; & \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{\delta_1=0}^{\infty} \sum_{\delta_2=0}^{\infty} (1-\rho^2)^{(r+s)/2} \frac{\Gamma(\frac{r}{2}+j)}{\Gamma(\frac{r}{2}) j!} \frac{\Gamma(\frac{s}{2}+k)}{\Gamma(\frac{s}{2}) k!} \rho^{2(j+k)} \\
& \times \left[1 - \gamma^{\star}\left(\frac{r}{2} + j + \delta_1, \frac{\chi^2_{r,\alpha_1}}{2(1-\rho^2)}\right)\right] \left[1 - \gamma^{\star}\left(\frac{s}{2} + k + \delta_2, \frac{\chi^2_{s,\alpha_3}}{2(1-\rho^2)}\right)\right] \\
& \times \frac{e^{-\theta^{UT}/2} (\theta^{UT}/2)^{\delta_1}}{\delta_1 !} \frac{e^{-\theta^{PT}/2} (\theta^{PT}/2)^{\delta_2}}{\delta_2 !},
\end{aligned}
\tag{6.5.5}
$$

with (r, s) degrees of freedom, noncentrality parameters, θ^{UT} and θ^{PT} and correlation coefficient, $-1 < \rho < 1$. For details on the evaluation of the bivariate integral, see Yunus and Khan (2009). The density function of the bivariate noncentral chi-square distribution given above is a mixture of the bivariate central chi-square distribution of two central chi-square random variables with different degrees of freedom (see Gunst and Webster, 1973, Wright and Kennedy, 2002), with the probabilities from the Poisson distribution. Let $\rho^2 = \sum_{j=1}^{p} \frac{1}{p} \rho_j^2$ be the mean correlation, where ρ_j is the correlation coefficient for any two different elements of the augmented vector $\left[\, n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\tilde{\boldsymbol{\beta}}_2), n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\mathbf{0})\,\right]$ in equation (6.4.3).

Write the second term on the right hand side of the equation (6.5.4) as

$$\int_{\chi^2_{r,\alpha_1}}^{\infty}\int_{\chi^2_{s,\alpha_3}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2 = [1-H_r(\chi^2_{r,\alpha_1};\theta^{UT})][1-H_s(\chi^2_{s,\alpha_3};\theta^{PT})], \tag{6.5.6}$$

with $H_r(\chi^2_{r,\alpha_1};\theta^{UT}) = \sum_{j=0}^{\infty}\sum_{\delta_1=0}^{\infty}\frac{(1-\rho^2)^{\frac{r}{2}}\Gamma(\frac{r}{2}+j)\rho^{2j}}{\Gamma(\frac{r}{2})j!}\gamma^{\star}\left(\frac{r}{2}+j+\delta_1, \frac{\chi^2_{r,\alpha_1}}{2(1-\rho^2)}\right)e^{-\frac{\theta^{UT}}{2}}\left(\frac{\theta^{UT}}{2}\right)^{\delta_1}/\delta_1!$ and $H_s(\chi^2_{s,\alpha_3};\theta^{PT}) = \sum_{k=0}^{\infty}\sum_{\delta_2=0}^{\infty}\frac{(1-\rho^2)^{\frac{s}{2}}\Gamma(\frac{s}{2}+k)\rho^{2k}}{\Gamma(\frac{s}{2})k!}\gamma^{\star}\left(\frac{s}{2}+k+\delta_2, \frac{\chi^2_{s,\alpha_3}}{2(1-\rho^2)}\right)\frac{e^{-\frac{\theta^{PT}}{2}}\left(\frac{\theta^{PT}}{2}\right)^{\delta_2}}{\delta_2!}$. Note, $H_r(\chi^2_{r,\alpha_1};\theta^{UT}) \geq G_r(\chi^2_{r,\alpha_1};\theta^{UT})$ and $H_s(\chi^2_{s,\alpha_3};\theta^{PT}) \geq G_s(\chi^2_{s,\alpha_3};\theta^{PT})$. Equality is archived when $\rho = 0$, or when $\boldsymbol{\lambda}_1 = \mathbf{0}$ and $\boldsymbol{\lambda}_2 = \mathbf{0}$.

Consider all the cases when $\theta^{RT} \geq \theta^{UT}$ and $\rho \neq 0$. So, using equation (6.5.6), we write equation (6.5.4) as

$$\tilde{\Pi}^{PTT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) \;\leq\; \tilde{\Pi}^{RT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)[1-\tilde{\Pi}^{PT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)] + \tilde{\Pi}^{UT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)\tilde{\Pi}^{PT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2). \tag{6.5.7}$$

Equality in equation (6.5.7) is achieved when both $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are $\mathbf{0}$. It is obvious that $\tilde{\Pi}^{PTT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) \leq \tilde{\Pi}^{RT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) - \tilde{\Pi}^{PT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)v_2$ for $0 \leq v_2 < 1$, and it follows that $\tilde{\Pi}^{PTT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) \leq \tilde{\Pi}^{RT}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)$ for any $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. Equality holds when both $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are $\mathbf{0}$.

Rewrite equation (6.5.5) as

$$\begin{aligned}&\int_{\chi^2_{r,\alpha_1}}^{\infty}\int_{\chi^2_{s,\alpha_3}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2\\ =\;& 1-H_r(\chi^2_{r,\alpha_1};\theta^{UT})-H_s(\chi^2_{s,\alpha_3};\theta^{PT})+\int_0^{\chi^2_{r,\alpha_1}}\int_0^{\chi^2_{s,\alpha_3}}\tilde{\phi}(w_1,w_2)dw_1dw_2.\end{aligned} \tag{6.5.8}$$

When $\boldsymbol{\lambda}_2$ is not large but not $\mathbf{0}$ and $\boldsymbol{\lambda}_1$ is sufficiently large, the first term on the right hand side of the equation (6.5.4) becomes $G(\chi^2_{s,\alpha_3};\theta^{PT})$ because θ^{RT} is sufficiently large. The second and fourth terms on the right hand side of the equation (6.5.8) becomes 0 because θ^{UT} is large. Also, note that $H_s(\chi^2_{s,\alpha_3}; \theta^{PT}) > G_s(\chi^2_{s,\alpha_3}; \theta^{PT})$. So, $\tilde{\Pi}^{PTT} < \tilde{\Pi}^{UT} = 1$ for sufficiently large $\boldsymbol{\lambda}_1$ and not so large $\boldsymbol{\lambda}_2$ $(\neq \mathbf{0})$.

Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. When $\boldsymbol{\lambda}_2 = \mathbf{0}$ and $\boldsymbol{\lambda}_1$ is sufficiently large, the first term on the right hand side of the equation (6.5.4) becomes $1-\alpha$ because θ^{RT} is large. Both $H_r(\chi^2_{r,\alpha}; \theta^{UT})$ and $\int_0^{\chi^2_{r,\alpha}} \int_0^{\chi^2_{s,\alpha}} \tilde{\phi}(w_1, w_2) dw_1 dw_2$ of the equation (6.5.8) become 0 while $H_s(\chi^2_{r,\alpha}; \theta^{PT})$ becomes $1-\alpha$. So, $\tilde{\Pi}^{PTT} = \tilde{\Pi}^{UT} = 1$ when $\boldsymbol{\lambda}_1$ is sufficiently large and $\boldsymbol{\lambda}_2 = \mathbf{0}$.

In the same manner, we observe results given in (c)-(e) below

(a) When $\boldsymbol{\lambda}_2$ $(\neq \mathbf{0})$ is not large and $\boldsymbol{\lambda}_1$ is sufficiently large, then, $\tilde{\Pi}^{PTT} < \tilde{\Pi}^{UT} = 1$.

(b) When $\boldsymbol{\lambda}_2 = \mathbf{0}$ and $\boldsymbol{\lambda}_1$ is sufficiently large, then, $\tilde{\Pi}^{PTT} = \tilde{\Pi}^{UT} = 1$.

(c) When $\boldsymbol{\lambda}_1$ $(\neq \mathbf{0})$ is not large but $\boldsymbol{\lambda}_2$ is sufficiently large, then, $\tilde{\Pi}^{PTT} < \tilde{\Pi}^{UT} = 1$.

(d) When $\boldsymbol{\lambda}_1 = \mathbf{0}$ and $\boldsymbol{\lambda}_2$ is sufficiently large, then, $\tilde{\Pi}^{PTT} = \tilde{\Pi}^{UT} = \alpha$.

(e) When both $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are $\mathbf{0}$, then $\tilde{\Pi}^{PTT} = \tilde{\Pi}^{UT} = \alpha$.

This confirms that the asymptotic size of the PTT is larger than that of the UT but less than that of the RT. For small and moderate values of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, the asymptotic power of the PTT is larger than that of the UT but less than that of the RT. But for large $\boldsymbol{\lambda}_1$ or $\boldsymbol{\lambda}_2$, the asymptotic power of the PTT may be smaller than that of the UT as well as the RT.

6.6 Illustrative Example

For this illustrative example, we consider samples of size 100 from the multiple linear regression model in equation (6.1.1) with $p = 3$, $r = 1$ and $s = 2$. The random errors, e_i's ($i = 1, 2, \ldots, 100$) are generated from the standard normal distribution using a code in R. Then, set $\beta_\nu = 1$ for $\nu = 1, 2, 3$. Let $c_{1i} = 1$ while c_{2i} and c_{3i} are 0 or 1 with 50% for each. In practice, often the normality assumption is not met due to the presence of contaminants in the collected data. In this example, to create contaminated observations, we randomly choose to replace $m (< n)$ of the n responses with some additive contamination, such that the contaminated responses X_i' is $X_i' = \beta_1 + \beta_2 c_{2i} + \beta_3 c_{3i} + d_i$ with d_i are generated from uniform distribution, $U[-5, -3.5]$ and $U[3.5, 5]$ with 50% for each. Only 10% contamination in the data is considered for simulation. For the contaminated data, the power functions of the UT, RT and PTT are calculated by equations (6.5.1), (6.5.2) and (6.5.4) using the Huber ψ-function, $\psi_H(U_i) = -k$ if $U_i < -k$, U_i if $|U_i| \leq k$, k if $U_i > k$, where $U_i = (X_i - \beta_1 - \beta_2 c_{2i} - \beta_3 c_{3i})/S_n$ with $S_n = MAD/0.6745$ and MAD is known as the mean absolute deviation. As suggested in many reference books (e.g. Wilcox, 2005, p.76), the value of $k = 1.28$ is chosen because $k = 1.28$ is the 90th quantile of a standard normal distribution, so there is a 0.8 probability that a randomly sampled observation will have a value between $-k$ and k. The estimate for σ_0^2 is taken to be $\sum \psi_H^2(U_i)/n$. For the estimation of γ, an R-estimate from the Wilcoxon sign rank statistic is used. The estimate of γ is the value of t such that $S(V_1, \ldots, V_n, t) = \sum_{i=1}^n \operatorname{sign}(V_i - t) a_n(R_{n_i}^+(t)) = 0$, where $R_{n_i}^+(t)$ is the rank of $V_i - t$ and $a_n(k) = k/(n+1), k = 1, \ldots, n$. Here, $V_i = \psi_H'(U_i)/S_n$ where $\psi_H'(U_i)$ is just the derivative of the Huber ψ-function.

Let $\boldsymbol{\lambda}_1 = [\lambda_1]$ and $\boldsymbol{\lambda}_2 = [\lambda_2,\ \lambda_3]'$. Here, we set $\alpha_\nu = 0.05$ for $\nu = 1, 2, 3$ and consider all the cases when $\theta^{RT} \geq \theta^{UT}$. In Figure 6.1, the power of the UT, RT and PTT are plotted against λ_1 for the selected values of $[\lambda_2,\ \lambda_3]$. As λ_1 grows large, power of all tests grow large too. Although the power of the UT and RT are increasing to 1 as λ_1 is increasing, the power of the PTT is increasing to a value that is less than 1. The analytical findings in the previous Section supports these graphical results.

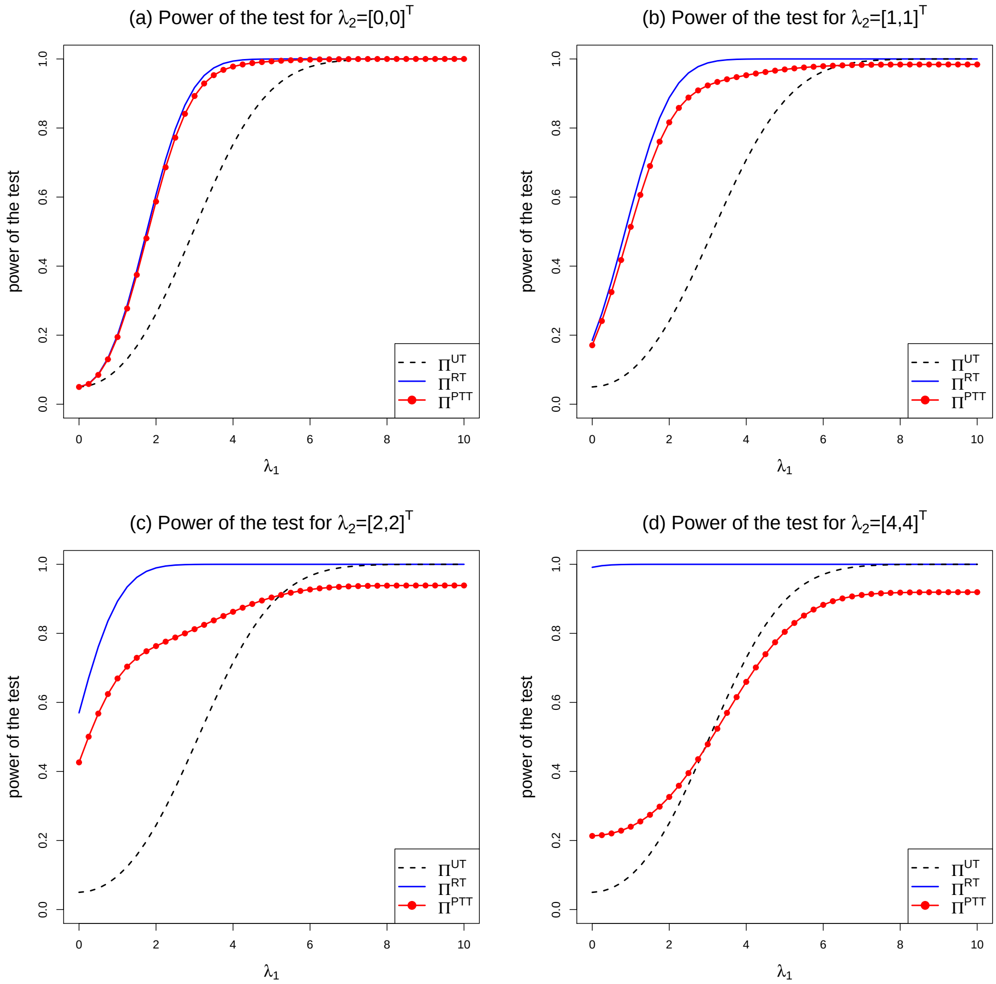


Figure 6.1: Graphs of power of the tests as a function of $\boldsymbol{\lambda}_1$ for selected values of $\boldsymbol{\lambda}_2$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$.

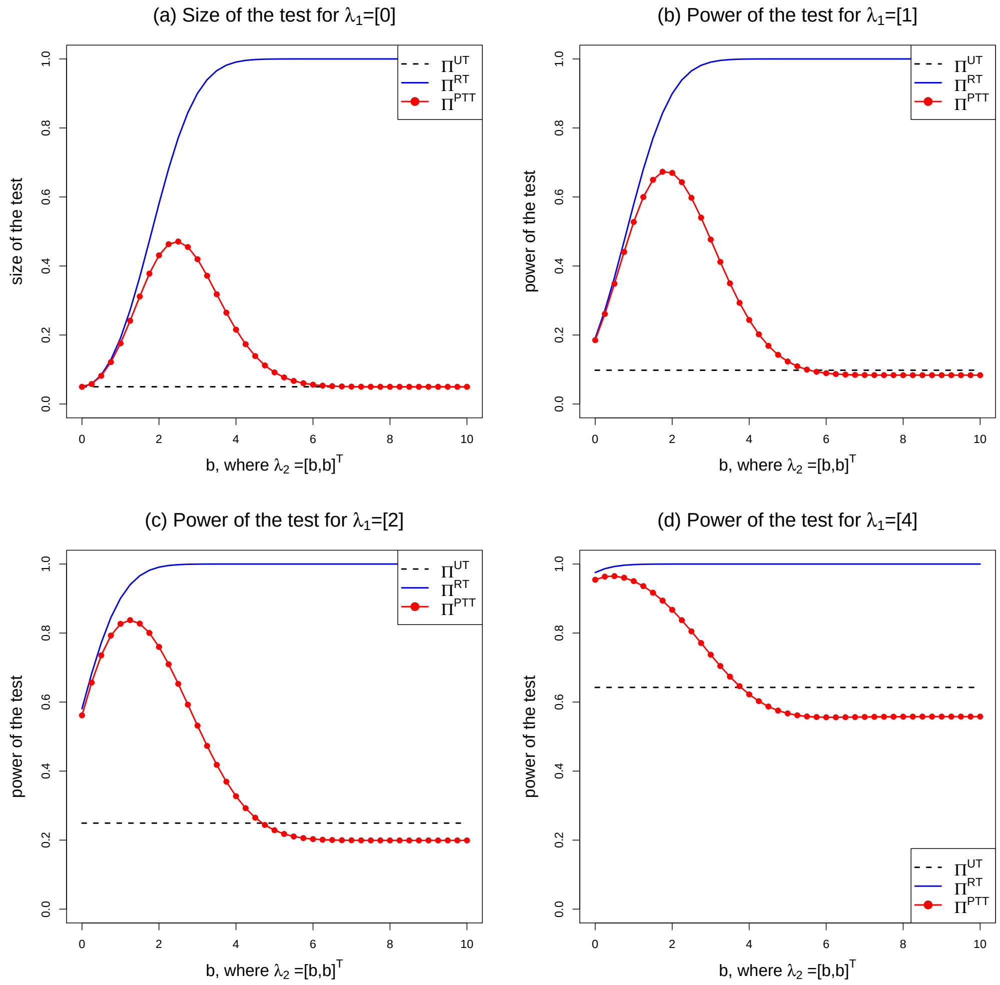


Figure 6.2: Graphs of power of the tests as a function of $\boldsymbol{\lambda}_2$ for selected values of $\boldsymbol{\lambda}_1$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$.

Since the UT, RT and PTT are defined based on the knowledge of $\boldsymbol{\beta}_2 = [\beta_2,\ \beta_3]'$, the size and power of each test are plotted against b such that $\boldsymbol{\lambda}_2 = [b,\ b]'$ in Figure 6.2. Figure 6.2 depicts that the RT has the largest power but also the largest size as b grows larger. By contrast, the UT has the constant smallest size regardless of the value of b but the constant smallest power when $b < q$, where q is some positive value. From the observations, the PTT is a compromise in minimizing the size and maximizing the power when $b < q$. This is because it has a smaller size than the RT but larger power than the UT. However, the PTT has the lowest power compared to the other tests when $b > q$. Although the prior information on the $\boldsymbol{\beta}_2$ vector may be uncertain, there is a high possibility that the true values are not too far from the suspected values. Therefore, the study on the behaviour of the three tests when $b < q$ is more realistic.

6.7 The Comparison of the Gaussian and Chi-square Tests

Consider the simple regression model given in equation (3.1.1) of Chapter 3. Recall the test statistics T_n^{UT}, T_n^{RT} and T_n^{PTT} in Section 3.3 of Chapter 3, that are asymptotically normal with a given mean and variance. In this Section, define all of the tests derived in Chapter 3 as the Gaussian tests. The power functions for these tests are plotted in Figures 3.4 and 3.5 of Chapter 3.

We know the simple regression model of equation (3.1.1) is a special case of the multiple linear regression model of equation (6.1.1). Now, consider the test statistics, L_n^{UT}, L_n^{RT} and L_n^{PTT} in Section 6.3 of this Chapter after letting $p = 2$, $s = 1$, $r = 1$, $\beta_1 = \theta$, $\beta_2 = \beta$, $c_{1i} = 1$ and $c_{2i} = c_i$ in equation (6.1.1). We find $L_n^{UT} = n^{-1}M_{n_1}(0, \tilde{\beta})^2 / \left(\frac{\sigma_0^2 C^{\star 2}}{C^{\star 2} + \bar{c}^2}\right) \xrightarrow{d} \chi_1^2$ under $H_0^{(1)} : \theta = 0$, $L_n^{RT} = n^{-1}M_{n_1}(0,0)^2/\sigma_0^2 \xrightarrow{d} \chi_1^2$ under $H_0 : \theta = 0, \beta = 0$ and

$L_n^{PT} = n^{-1}M_{n_2}(\tilde{\theta}, 0)^2/\left(\sigma_0^2 C^{\star 2}\right) \xrightarrow{d} \chi_1^2$ under $H_0^{(2)}: \beta = 0$. We then call these tests the chi-square tests.

Considering the simple regression model, the aim is to compare the performance of the Gaussian test and the chi-square test for the PTT. The power functions for the UT, RT and PTT from the Gaussian test are given in equations (3.5.5), (3.5.7) and (3.5.9) of Chapter 3. The power functions for the chi-square tests are given in equations (6.5.1), (6.5.2) and (6.5.4) with $s,\ r = 1$. The noncentrality parameters $\theta^{UT} = (\gamma/\sigma_0)^2\lambda_1^2\left[C^{\star 2}/(C^{\star 2} + \bar{c}^2)\right]$, $\theta^{RT} = (\gamma/\sigma_0)^2(\lambda_1 + \lambda_2\bar{c})^2$ and $\theta^{PT} = (\gamma/\sigma_0)^2(\lambda_2 C^\star)^2$ are obtained using equations (6.4.4), (6.4.5) and (6.4.6) of this Chapter and equation (3.2.4) of Chapter 3. Again, the density function of the bivariate noncentral chi-square distribution $\tilde{\phi}(\cdot)$, as a mixture of the bivariate central chi-square distribution (see Gunst and Webster, 1973, Wright and Kennedy, 2002), with the probabilities from the Poisson distribution, proposed by Yunus and Khan (2009) is used.

We use the same simulated data as in Section 3.7.2 of Chapter 3 to plot the power of the UT, RT and PTT for the chi-square test in Figures 6.3 and 6.4. From Figures 3.4(d) and 6.4(c), it is observed that the power functions of the UT, RT and PTT for both Gaussian and chi-square tests are not behaving in the same manner though both tests are derived from the same statistics. Figure 6.4(c) depicts that the power of the PTT from the chi-square test is lower than that of the UT when $\lambda_2 > 4$. However, the power of the PTT using the Gaussian test is at least as much as that of the UT when λ_2 is large (see Figure 3.4(d)). We also find that the power of the PTT using the chi-square test does not reach 1 as λ_1 increases, and it is lower than those of the UT and RT (see Figure 6.3(c)) when $\lambda_2 = 2$ and λ_1 is large. However, in Figure 3.5(d), the power of the UT, RT and PTT for the Gaussian test approaches 1 as λ_1 increases.

Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, we rewrite the bivariate integral given by equation

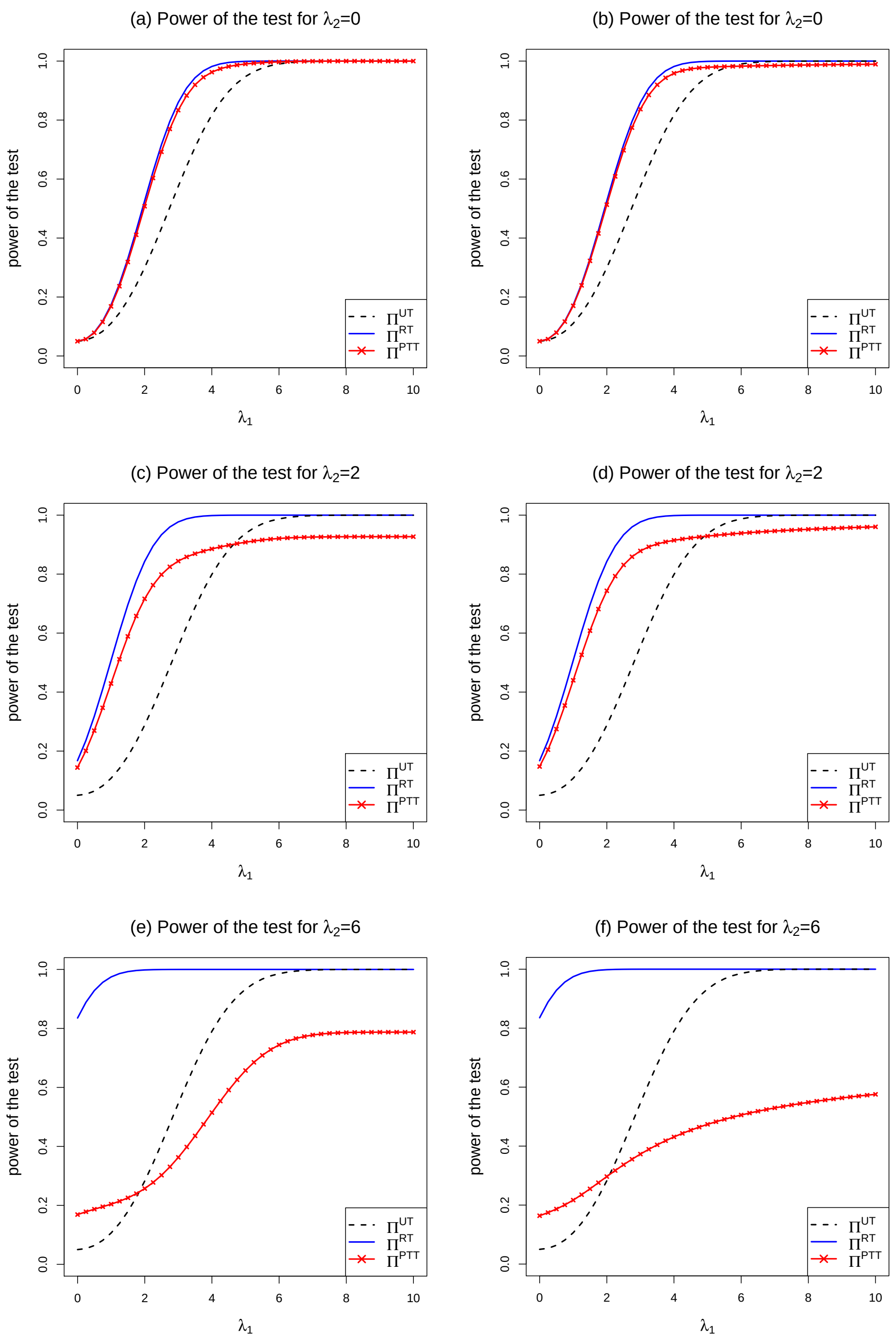


Figure 6.3: Graphs of power functions as a function of λ_1 for selected values of λ_2, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$ and $\bar{c} > 0$. The power of the PTT in (a), (c) and (e) are obtained using the bivariate noncentral chi-square while (b), (d) and (f) are their approximation using the Steyn-Roux method.

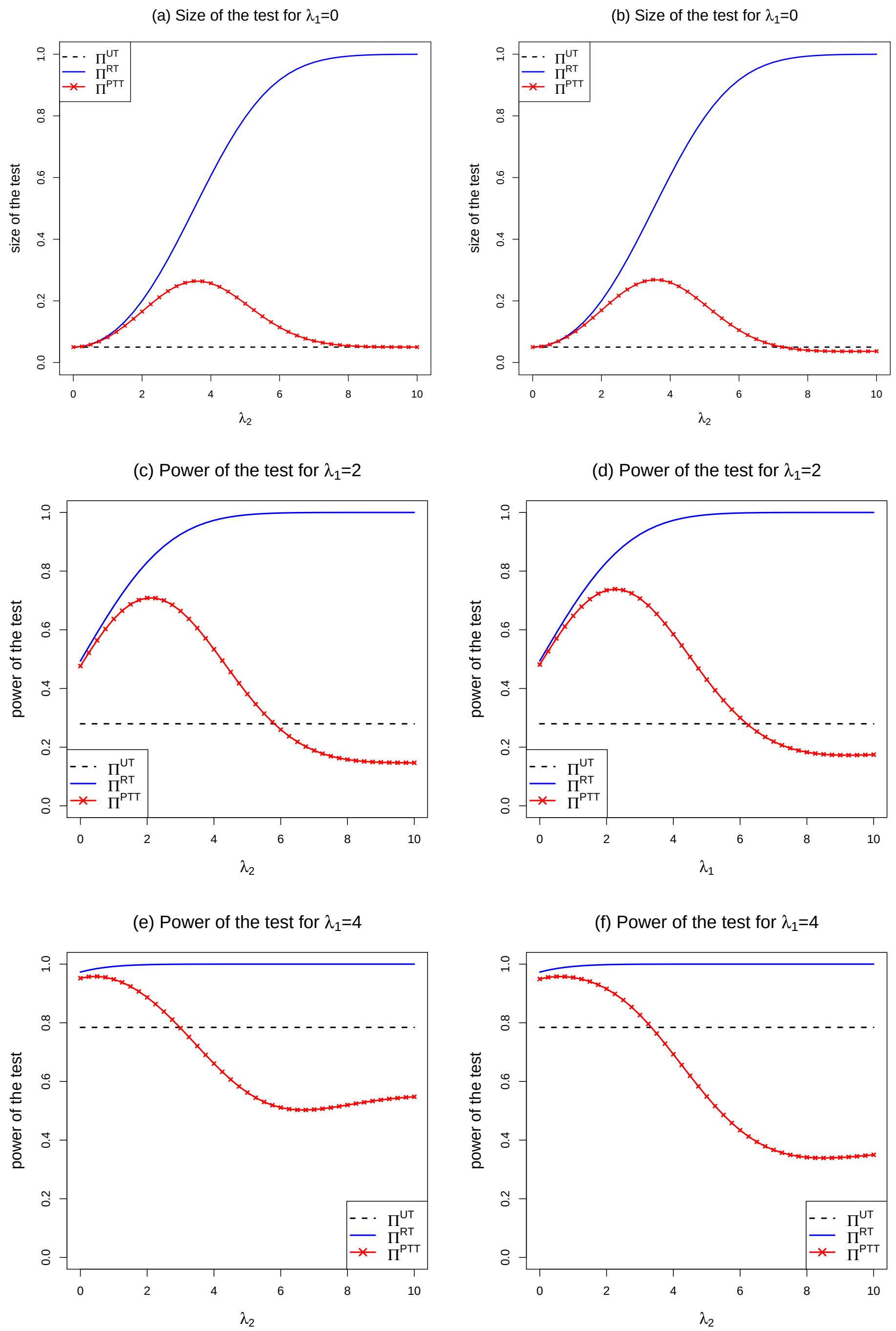


Figure 6.4: Graphs of power functions as a function of λ_2 for selected values of λ_1, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$ and $\bar{c} > 0$. The power of the PTT in (a), (c) and (e) are obtained using the bivariate noncentral chi-square while (b), (d) and (f) are their approximation using the Steyn-Roux method.

(6.5.6) as

$$\begin{aligned}\int_{\chi^2_{1,\alpha}}^{\infty}\int_{\chi^2_{1,\alpha}}^{\infty}\tilde{\phi}(w_1,w_2)dw_1dw_2 &= 1-H_1(\chi^2_{1,\alpha};\theta^{UT})-H_1(\chi^2_{1,\alpha};\theta^{PT}) \\ &\quad +H_1(\chi^2_{1,\alpha};\theta^{UT})H_1(\chi^2_{1,\alpha};\theta^{PT}), \qquad (6.7.1)\end{aligned}$$

where $H_1(\cdot,\cdot)$ is as defined in Section 6.5. Consider a sufficiently small λ_2 $(\neq 0)$ and a sufficiently large λ_1, such that $\tilde{\Pi}^{UT}=\tilde{\Pi}^{RT}=1$. The 2nd and 4th terms on the right hand side of the equation (6.7.1) become 0 because θ^{UT} and θ^{RT} are large. The remaining terms of the equation (6.7.1) are smaller than $1-G(\chi^2_{1,\alpha},\theta^{PT})=\tilde{\Pi}^{PT}$, where $G(\chi^2_{1,\alpha},\theta^{PT})$ is the cdf of the noncentral bivariate chi-square with 1 d.f. and noncentrality parameter θ^{PT}. The first term on the right of the equation (6.5.4) becomes $1-\tilde{\Pi}^{PT}$ because θ^{RT} is sufficiently large. Thus, we find that $\tilde{\Pi}^{PTT}<1$ when λ_2 $(\neq 0)$ is small and λ_1 is large. This analytical result means it is not impossible that the power of the PTT is less than 1, that is, it is less than the power of the UT and the RT, when λ_1 is sufficiently large and λ_2 is small but not 0. This analytical result supports the graphical view of the power function for the PTT from the chi-square test as shown in Figures 6.3(c) and 6.4(c).

Kocherlakota and Kocherlakota (1999) suggested an approximation for the bivariate noncentral chi-square distribution using the bivariate central chi-square distribution. An approximation of the bivariate noncentral chi-square distribution by the central chi-square distribution using some transformations on the random variables, correlation coefficient and degree of freedom, is also considered in this Section. Note, we can write part (ii) of Theorem 6.4.1 as

$$\boldsymbol{Z}=n^{-\frac{1}{2}}\begin{bmatrix}\frac{M_{n_1}(0,\tilde{\beta})}{\sigma_0\sqrt{\frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2}}}\\ \frac{M_{n_2}(\tilde{\theta},0)}{\sigma_0\sqrt{C^{\star 2}}}\end{bmatrix}\xrightarrow{d}N_2\left[\begin{pmatrix}\frac{\gamma\lambda_1}{\sigma_0}\sqrt{\frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2}}\\ \frac{\gamma\lambda_2}{\sigma_0}\sqrt{C^{\star 2}}\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right], \qquad (6.7.2)$$

where $\rho=-\frac{\bar{c}}{\sqrt{C^{\star 2}+\bar{c}^2}}$. Then, $\boldsymbol{ZZ'}\sim W_2(\text{d.f.}=1,\Sigma,\Omega)$ (Wishart distribution) with parameters $\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$, and $\Omega=\Sigma^{-1}M'M$, where $M=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=$

$$\begin{pmatrix} \frac{\gamma\lambda_1}{\sigma_0}\sqrt{\frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2}} \\ \frac{\gamma\lambda_2}{\sigma_0}\sqrt{C^{\star 2}} \end{pmatrix}.$$

Note, the diagonal of $\boldsymbol{ZZ}'$ are L_n^{UT} and L_n^{PT}. We find that L_n^{UT} and L_n^{PT} have the noncentral bivariate chi-square distribution with correlation coefficient ρ, degree of freedom 1 and noncentrality parameters $\theta^{UT} = \mu_1^2$ and $\theta^{PT} = \mu_2^2$.

Steyn and Roux (1972) proposed an approximation for the noncentral Wishart distribution by a central Wishart distribution. Kocherlakota and Kocherlakota (1999) used the idea by Steyn and Roux (1972) for the approximation of the noncentral bivariate chi-square distribution. Following Kocherlakota and Kocherlakota (1999), the random variables for the noncentral bivariate chi-square, (L_n^{UT}, L_n^{PT}) are approximately distributed as the bivariate central chi-square distribution with transformed variables,

$$\frac{L_n^{UT}}{1+\mu_1^2} \text{ and } \frac{L_n^{PT}}{1+\mu_2^2},$$

degree of freedom 1 and correlation coefficient,

$$\rho^{\star 2} = \frac{(\rho+\mu_1\mu_2)^2}{(1+\mu_1^2)(1+\mu_2^2)}.$$

We find that the power of the PTT, computed directly using the bivariate noncentral chi-square distribution and using the approximation by the Steyn-Roux method, is behaving quite similar (see Figures 6.3 and 6.4). It is observed that the power of the PTT is also less than that of the UT and RT for sufficiently large λ_1 and small λ_2 (or small λ_1 and large λ_2) using the approximation by the Steyn-Roux method.

6.8 Discussion and Conclusion

In this Chapter, the asymptotic sampling distributions of the UT, RT and PT follow a univariate noncentral chi-square distribution under the local alternative hypothesis when the sample size is large. However, the sampling distribution of the PTT is a bivariate noncentral chi-square distribution as there is a correlation between the UT and PT. Note that there is no such correlation between the RT and PT. The new R code is used for the computation of the distribution function of the bivariate noncentral chi-square distribution (Yunus and Khan, 2009) to evaluate the power function of the PTT .

The RT has the largest power among the three tests, but it also has the largest size. So, it is not a valid test because it does not satisfy the asymptotic level constraint. On the other hand the UT has the smallest size, but it has the smallest power as well except when $\boldsymbol{\lambda}_1 = n^{\frac{1}{2}}\boldsymbol{\beta}_1$ or $\boldsymbol{\lambda}_2 = n^{\frac{1}{2}}\boldsymbol{\beta}_2$ is large. So, both UT and RT fail to achieve the highest power and lowest size simultaneously. The PTT has a smaller size than the RT. It also has higher power than the UT, except for the very large values of $\boldsymbol{\lambda}_1$ or $\boldsymbol{\lambda}_2$. Therefore if the prior information is not far away from the true value, that is, $\boldsymbol{\lambda}_2$ is near $\mathbf{0}$ (small or moderate) the PTT has a smaller size than the RT and higher power than the UT. Hence it is a better compromise between the two extremes. Since the prior information is coming from previous experience or expert knowledge, it is reasonable to expect $\boldsymbol{\lambda}_2$ should not be too far away from $\mathbf{0}$, although it may not be $\mathbf{0}$, and hence the PTT demonstrates a reasonable domination over the other two tests in more a realistic situation.

Chapter 7

Discussions, Conclusions and Future Research

7.1 Discussion and Conclusion

Under a sequence of local alternative hypotheses, the sampling distributions for the UT, RT and PT of a simple regression model when the sample size is large, follow a normal distribution with appropriate mean and variance as discussed in Chapter 3. However, that of the PTT is a bivariate normal distribution. There is a correlation between the UT and PT, but there is no such correlation between the RT and PT.

The size of the test is the probability of rejecting the null hypothesis when it is true. For all cases where a mean of regressor of the simple regression model is larger than 0, the probability of rejecting the null hypothesis $H_0 : \theta = \theta_0$, when it is true for all UT, RT and PTT, increases and tends to unity as the suspected intercept θ_0 moves away from the true intercept θ. The size of the RT increases and tends to unity as the suspected value of the slope β_0 moves away from the true slope β. Hence, the RT is not a valid test because it does not satisfy the asymptotic size constraint especially when $\lambda_2 = \sqrt{n}|\beta - \beta_0|$ is large. For the UT, the probability of rejecting H_0 when it is true is constant regardless of the distance between the true slope and its suspected value. The PTT has a significantly smaller size of the test than the RT for moderate and large λ_2. Thus, the PTT is better in terms of size than the RT, though the UT remains as the most preferable.

The power of the test is the probability of rejecting the null hypothesis when it is false. The PTT has a larger probability of rejecting $H_0 : \theta = \theta_0$ when it is false than that of the UT for smaller and moderate λ_2. Although the RT has the largest power of the other two tests, it does not satisfy the asymptotic level constraint.

Therefore, the power function of the PTT is found to behave similar to the MSE of the PTE, in the sense that although it is not uniformly the best statistical test with the smallest size and the largest power, it does protect from the risk of size being too large and power being too small. Thus, the power function of the PTT is a compromise between that of the UT and RT. In the face of uncertainty on the value of the slope, if the objective of a researcher is to minimize the size and maximize the power of the test, the PTT is therefore the best choice.

The performance (size and power) of the PTT depends on its arguments, namely the slope (via λ_2) and the nominal sizes (preassigned significance level) of the UT, RT and PT. The values of the nominal sizes for the UT, RT and PT are set before testing is carried out, and they affect the actual size of the PTT. In order to get a small probability of a Type I error for the PTT, the investigations concentrate on small nominal sizes of the UT, RT and PT with a view to achieving small (actual) size of the PTT.

This study revealed that for small and moderate values of λ_2, the smaller the nominal size of the RT, the smaller the (actual) size of the PTT, when other nominal sizes are kept fixed and small. For moderate and large values of λ_2, a large size of the PTT is observed when the nominal size of the PT is set close to 0. The performance of the PTT improves when a larger value of nominal size of the PT is selected. However, setting the nominal size of the PT as a large value means the probability of a Type I error is large for the PT (testing on the slope). The size of the PTT behaves much like that of the RT when the nominal size of the PT is small, but it behaves more like that of the UT when the nominal size of the PT is large.

The power of the PTT is larger for moderate values of λ_2 than for smaller and larger values of λ_2. It is shown analytically that the power of the PTT approaches the power of the RT when the nominal size of the PT is closer to 0, but that it approaches the power of the UT when the nominal size of the PT is closer to 1. In practical applications, the size of the PT should be small (ideally close to 0), and in such cases the power of the PTT is close to that of the RT (which is much higher than that of the UT).

To avoid the larger size of the RT, practitioners are recommended to use the PTT as it achieves a smaller size (than the RT) and higher power (than the UT) when the distance between the slope and its suspected value is small or moderate. Even for large values of the distance between the slope and its suspected value, the PTT has at least as much power as the UT and has at least as small a size as the RT.

For the multivariate simple regression model, multiple linear regression model and parallelism model, the sampling distributions of the UT, RT and PT follow a univariate noncentral chi-square distribution under the alternative hypothesis when the sample size is large. However, that of the PTT is a bivariate noncentral chi-square distribution as there is a correlation between the UT and PT. Note that there is no such correlation between the RT and PT. New R codes are written for the computation of the distribution function of the bivariate noncentral chi-square distribution proposed, as by Yunus and Khan (2009), to evaluate the power function of the PTT.

For the multivariate simple regression model, the size of the RT reaches 1 as b (a function of the difference between the true and suspected values of the slope vector) increases, so the RT is not a valid test as it does not satisfy the asymptotic level constraint. Although the UT has the smallest constant size, it has the smallest power as well, except for very large values of b, that is when $b > q$, where q is some positive number. So, the UT fails to achieve the highest power and lowest size simultaneously. The PTT has a smaller size than the RT and its size does not reach 1 for any b. It also has higher power than the UT, except for $b > q$. Therefore if the prior information is not far away from the true value, that is, b is near 0 (small or moderate), the PTT has a smaller size than the RT and higher power than the UT. Hence it is a better compromise between the two extremes. Since the NSPI comes from previous experience or expert knowledge, it is reasonable to expect b should not be too far away from 0, although it may not be 0, and hence the PTT demonstrates a reasonable domination over the other two tests in a more realistic situation.

The power of the UT, RT and PTT for the multiple linear regression and parallelism models demonstrates similar behavior as those of the multivariate simple regression model. The PTT shows a reasonable domination over the other two tests asymptotically when the suspected NSPI value of the parameter interest is not too far away from its true value (that is under the null hypothesis). Similar to that of the multivariate simple regression model, the PTTs for the multiple linear regression and parallelism models also have lower power than those of the UTs when the suspected NSPI value is far away from that under the null hypothesis. Since the NSPI comes from previous experience or expert knowledge, it is reasonable to expect the suspected NSPI value is not too far away from that under the null hypothesis.

The bivariate noncentral chi-square distribution is involved in the formula of the power function of the PTT for the multivariate simple regression model, multiple linear regression model and parallelism model. The PTTs for these regression models do not show the same behavior when it comes to the power of the test as that of the simple regression model. The power of the PTT for the simple regression model tends to unity, whereas those of the other regression models do not reach 1 as the intercept (or intercept vector) moves away from its suspected value. Also, as the slope (or slope vector) moves away from its suspected value, the PTT has at least as much power as the UT for the simple regression model, whereas it is lower than that of the UT for the other regression models. In Section 6.7, we find that the power of the PTT computed, using an approximation to the noncentral chi-square (Kocherlakota and Kocherlakota, 1999) approach, also behaves in the same manner as if we were to use the proposed noncentral bivariate chi-square distribution by Yunus and Khan (2009).

The UT, RT and PTT are proposed using the score function in the robust M-estimation methodology. The robustness of the UT, RT and PTT based M-tests is investigated on data simulated using the Monte Carlo method. The simulated data is generated for both normal and contaminated normal responses. The power of the UT, RT and PTT using the Huber and Tukey ψ functions (score functions) is not significantly affected by slight departures from the assumed normal model. The UT, RT and PTT based on the LS method depend heavily on the model assumptions and they are not robust if the normality assumption is not satisfied. The Huber M-test for the UT and RT enjoys the maximin power (or minimax property) of a test which is formerly suggested by Huber (1965) for any values of $\lambda_1 = \sqrt{n}|\theta - \theta_0|$ and λ_2. However, the Huber M-test for the PTT does not enjoy the minimax property for some λ_1 and λ_2 (see result (vii) of Section 3.6).

7.2 Limitations and Future Directions

This thesis considers a one sided hypothesis, $H_0 : \theta = \theta_0$ against $H_0 : \theta > \theta_0$. This means, if the null hypothesis is rejected, then the true intercept is larger than its suspected value. For this one-sided alternative, it would not be the case that the true intercept is less than its suspected value if the null hypothesis is rejected. Nonetheless, it is recommended to propose a PTT for a two sided test. However, it is suspected that a more complicated form of power function for the PTT may be derived if a two-sided alternative is considered.

The complicated formula of the power function of the PTT limits the studies on the minimax property of a test. In this thesis, the robustness property of the PTT is investigated through computational analysis. The Monte-Carlo method is used and the simulated data is generated for both contaminated and uncontaminated cases. In spite of the graphical view, the theoretical analysis for the robustness property of the PTT is not thoroughly discussed due to the complexity form of the power function of the PTT, especially the PTT that involves the bivariate noncentral chi-square distribution in the formula of its power function.

The M-estimators based on the componentwise estimating equations for the multivariate simple regression model are considered in this thesis. However, the assumption of a strong correlation between elements of error vector $\boldsymbol{e}_i$ is questionable because the M-estimates, obtained using the method of componentwise equations, are more appropriate when there is small dependence between the elements of $\boldsymbol{e}_i$. The PTT for a general multivariate model which is not defined in a componentwise way is recommended for future work.

In this thesis, the study of the behavior of the power functions of the UT, RT and PTT only requires the stated regularity conditions to guarantee the consistency of the M-estimators. The entire asymptotic theory is directly adapted from Jurečková and Sen (1996) without any updating of the regularity conditions for the existence of a consistent M-estimator during the past 14 years. Since the idea was to use the existing asymptotic theory of the M-estimation method in the pre-testing framework, the updated regularity conditions were not essential for this study. However, the PTT may be proposed under weaker and updated regularity conditions in the future.

The PTT should be proposed for the other robust methodologies such as the GM-estimator, LM-estimator and S-estimator methodologies. Most probably however, this suggestion is difficult to attain due to the complicated form of the sampling distribution of the tests of these robust methodologies.

References

Abraham, B. and Ledolter, J. (2006), *Introduction to Regression Modeling*, Thomson Brooks/Cole, US.

Adichie, J. N. (1967), 'Estimates of regression parameters based on rank tests', *Ann. Math. Stat.* **38**, 894–904.

Ahmed, S. E., Hussein, A. A. and Sen, P. K. (2006), 'Risk comparison of some shrinkage m-estimators in linear models', *Jour. of nonparametric statist.* **18**(4-6), 401–415.

Ahmed, S. E. and Saleh, A. K. M. E. (1989), 'Estimation strategy using a preliminary test in some normal models', *Soochow. J. of Math.* **14**, 135–165.

Ahsanullah, M. and Saleh, A. K. M. E. (1972), 'Estimation of intercept in a linear regression model with one dependent variable after a preliminary test on the regression coefficient', *Int. Statist. Rev.* **40**(2), 139–145.

Akritas, M. G., Saleh, A. K. M. E. and Sen, P. K. (1984), Nonparametric estimation of intercepts after a preliminary test on parallelism of several regression lines., In: P. K. Sen (Ed). *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences* The Bernard G. Greenberg Vol., North-Holland, New York.

Ali, A. M. and Saleh, A. K. M. E. (1990), 'Estimation of the mean vector of a multivariate normal distribution under symmetry', *Jou. Statist. Comp. Simul.* **35**(3), 209–226.

Amado, C. and Pires, A. M. (2004), 'Robust boostrap with non random weights based on the influence function', *Comm. Stat.- Simul & Comp.* **33**, 377–396.

Bachmaier, M. (2007), 'Consistency of completely outlier-adjusted simultaneous redescending m-estimators of location and scale', *AStA* **91**, 197–219.

Bancroft, T. A. (1944), 'On biases in estimation due to the use of preliminary tests of significance', *An. Math. Statist.* **15**, 190–204.

Bancroft, T. A. (1964), 'Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance', *Biometrics* **20**(3), 427–442.

Bancroft, T. A. (1965), Inference for incompletely specified models in the physical sciences (with discussion), Bull. ISI, Proc. 35th Section, Beograde, Vol. 41, pp. 497–515.

Bantli, F. E. (2004), 'M-estimation in linear models under nonstandard conditions', *Journal of Statistical Planning and Inference* **121**, 231–248.

Barnett, V. and Lewis, T. (1995), *Outliers in statistical data*, Wiley, England.

Bechhofer, R. E. (1951), The effect of preliminary test of significance on the size and power of certain tests of univariate linear hypotheses, PhD thesis, Columbia Univ.

Becker, C. and Gather, U. (1999), 'The masking breakdown point of multivariate outlier identification rules', *J. Amer. Statist. Assoc.* **94**, 947–955.

Bozivich, H., Bancroft, T. A. and Hartley, H. O. (1956), 'Power of analysis of variance test procedures for certain incompletely specified models', *Ann. Math. Statist.* **27**, 1017–1043.

Büning, H. (2000), 'Robustness and power of parametric, nonparametric, robustified and adaptive tests - the multi-sample location problem', *Statist. Papers* **41**, 381–407.

Burt, J. E. and Barber, G. M. (1996), *Elementary statistics for geographers, 2nd edition*, Guilford, New York.

Carolan, A. and Rayner, J. C. W. (2000), 'One sample tests of location for non-normal symmetric data', *Communications in Statistics - Theory and Methods* **29**, 1569–1581.

Carroll, R. J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman & Hall, US.

Chakraborty, B. (1999), 'On multivariate median regression', *Bernoulli* **5**, 683–703.

Coakley, C. W. and Hettmansperger, T. P. (1993), 'A bounded influence, high breakdown, efficient regression estimator', *J. Amer. Statist. Assoc.* **88**, 872–880.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap methods and their application*, Cambridge University Press, London.

De Veaux, R. D., Velleman, P. F. and Bock, D. E. (2009), *Intro Stats, 3rd Ed.*, Pearson, USA.

Dharmawansa, P. and McKay, M. R. (2009), 'Diagonal distribution of a complex non-central wishart matrix: A new trivariate non-central chi-squared density', *Jour. Multivariate Analysis* **100**, 561–580.

Donoho, D. L. and Huber, P. J. (1983), The notion of breakdown point, In: P.J Bickel, K. Doksum & J.L. Hodges (Eds). *A Festschrift for E.L.Lehmann*, Wadsworth, Belmont, Calif., pp. 157–184.

Dubey, S. D. (1970), 'Compound gamma, beta and F distributions.', *Metrika* **16**, 27–31.

Efron, B. and Tibshirani, R. (1993), *An introduction to the bootstrap*, Chapman & Hall, London.

Flaherty, J. (1999), Nonparametric statistics, Ch.8, Lecture notes PR220 - property data analysis 2, RMIT Univ.
URL: *http://www.bf.rmit.edu.au/Property/subjects/pr220/guide.htm*

Gagliardini, P., Trojani, F. and Urga, G. (2005), 'Robust GMM tests for structural breaks', *Journal of Econometrics* **129**, 139 – 182.

Garthwaite, P. H., Jolliffe, I. T. and Jones, B. (2002), *Statistical inference, 2nd ed.*, Oxford Science, US.

Gerstenkorn, T. (2004), 'A compound of the generalized negative binomial distribution with the generalized beta distribution.', *CEJM* **2**, 527–537.

Gilchrist, W. (1984), *Statistical Modelling*, Wiley, Chichester, UK.

Goodall, C. (1983), M-estimator of location: An outline of the theory, In: D.C. Hoaglin, F. Mosteller & J.W. Tukey (Eds). *Understanding robust and explanatory data analysis*, Wiley, New York, pp. 339–403.

Gunst, R. F. and Webster, J. T. (1973), ‘Density functions of the bivariate chi-square distribution’, *Journal of Statistical Computation and Simulation A* **2**(3), 275–288.

Hájek, J., Šidák, Z. and Sen, P. K. (1999), *Theory of Rank Tests, 2nd Ed*, Academia Press, US.

Hampel, F. R. (1968), Contributions to the theory of robust estimation, PhD thesis, Univ. California, Berkeley.

Hampel, F. R. (1971), ‘A general qualitative definition of robustness’, *Ann. Math. Statist.* **42**, 1887–1896.

Hampel, F. R. (1974), ‘The influence curve and its role in robust estimation’, *J. Amer. Statist. Assoc.* **69**, 383–393.

Hampel, F. R. (2001), Robust statistics: A brief introduction and overview, Vol. No.94, Eidgenössische Technische Hochschule (ETH), Zurich, Switzerland.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986), *Robust Statistics - The Approach based on influence functions*, Wiley, New York.

He, X., Simpson, D. G. and Portnoy, S. L. (1990), ‘Breakdown robustness of tests’, *J. Amer. Statist. Assoc.* **85**, 446–452.

Heritier, S. and Ronchetti, E. (1994), ‘Robust bounded-influence tests in general parametric models’, *J. Amer. Statist. Assoc.* **89**, 897–904.

Hettmansperger, T. P. and McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, Wiley, New York.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1983), *Understanding robust and explanatory data analysis*, Wiley, New York.

Huber, P. J. (1964), ‘Robust estimator of a location parameter’, *Ann. Math. Statist* **35**(1), 73–101.

Huber, P. J. (1965), 'A robust version of the probability ratio test', *Ann. Math. Statist* **36**(6), 1753–1758.

Huber, P. J. (1973), 'Robust regression: Asymptotics, conjectures and Monte Carlo', *Ann. Stat.* **1**, 799–821.

Huber, P. J. (1981), *Robust Statistics*, Wiley, New York.

Huber, P. J. (1983), 'Minimax aspects of bounded-influence regression', *J. Amer. Statist. Assoc.* **78**, 66–73.

Huber, P. J. (1984), 'Finite sample breakdown of M- and P- estimators.', *Ann. Statist.* **12**, 119–126.

Hutchinson, T. P. (1981), 'Compound gamma bivariate distributions.', *Metrika* **28**, 263–271.

Imon, A. H. M. R. (2003), 'Residuals from deletion in added variable plots', *Journal of Applied Statistics* **30**, 827–841.

James, W. and Stein, C. (1961), Estimation with quadratic loss, Proc. Fourth Berkeley Symp. Math. Statist. Prob., Vol. 1, University of California Press, Berkeley, pp. 361–379.

Judge, G. G. and Bock, M. E. (1978), *The statistical implications of pre-test and Stein-rule estimators in econometrics*, North-Holland, Amsterdam.

Jung, K. M. (2005), 'Multivariate least-trimmed squares regression estimator', *Computational Statistics and Data Analysis* **48**, 307–316.

Jurečková, J. (1977), 'Asymptotic relations of M-estimates and R-estimates in linear regression model', *Ann. Statist.* **5**, 464–472.

Jurečková, J. and Sen, P. K. (1981), 'Sequential procedures based on m-estimators with discontinuous score functions', *Journal of Statistical Planning and Inference* **5**, 253–266.

Jurečková, J. and Sen, P. K. (1996), *Robust Statistical Procedures: Asymptotics and Interrelations*, John Wiley & Sons, US.

Jurečková, J. and Picek, J. (2006), *Robust Statistical Methods with R*, Chapman, US.

Kabir, M. E. and Khan, S. (2009), 'Estimation of univariate normal mean using p-value', *Advances and Applications in Statistics* **13**, 241–256.

Kelly, G. E. (1992), 'Robust regression estimators - the choice of tuning constants', *JRSS (The Statistician)* **41**(3), 303–314.

Kent, R. (2008), Energy miser - know your plant's 'energy fingerprint', (accessed 6 August 2009).
URL: *http://www.ptonline.com/articles/200902em1.html*

Khan, S. (1998), 'On the estimation of the mean vector of Student-t population with uncertain prior information', *Pak Jou. Statist.* **14**, 161–175.

Khan, S. (2000), 'Improved estimation of the mean vector for Student-t model.', *Communications in Statistics: Theory & Methods* **29**, 507–527.

Khan, S. (2003), 'Estimation of the parameters of two parallel regression lines under uncertain prior information', *Biometrical Journal* **45**(1), 73–90.

Khan, S. (2005), 'Estimation of parameters of the multivariate regression model with uncertain prior information and Student-t errors', *Jour. of Statist. Research* **39**(2), 79–94.

Khan, S. (2006*a*), 'Prediction distribution of future regression and residual sum of squares matrices for multivariate simple regression model with correlated normal responses', *Jour. of Appl. Prob. and Statist.* **1**, 15–30.

Khan, S. (2006*b*), 'Shrinkage estimation of the slope parameters of two parallel regression lines under uncertain prior information', *Model Assisted Statist. Analysis* **1**(3), 193–205.

Khan, S. (2008), 'Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes', *Comm. in Statist. - Theory and Methods* **37**, 247–260.

Khan, S., Hoque, Z. and Saleh, A. K. M. E. (2002), 'Estimation of the slope parameter for linear regression model with uncertain prior information', *Jour. of Statist. Research* **36**, 55–73.

Khan, S. and Saleh, A. K. M. E. (1995), 'Preliminary test estimators of the mean for the sampling from some Student-t populations', *Jour. of Statist. Research* **29**, 67–88.

Khan, S. and Saleh, A. K. M. E. (1997), 'Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors', *Biometrical Jour.* **39**(2), 131–147.

Khan, S. and Saleh, A. K. M. E. (2001), 'On the comparison of pre-test and shrinkage estimators for the univariate normal mean', *Statistical papers* **42**(4), 451–473.

Khan, Z. and Dellaert, F. (2004), Robust generative subspace modeling: The subspace t distribution, Technical Report GIT-GVU-04-11, Georgia Institute of Technology.

Kim, H. M. (2003), On some contributions to measurement error models, PhD thesis, Carleton Univ., Ottawa, Canada.

Kitagawa, T. (1963), 'Estimation after preliminary tests of significance', *Univer. of California Publ. in Statist.* **3**, 147–186.

Kocherlakota, S. and Kocherlakota, K. (1999), 'Aproximations for central and noncentral bivariate chi-square distributions.', *Comm. in Statist. - Simul & Compt.* **28**, 909–930.

Koenker, R. and Portnoy, S. (1990), 'M estimation of multivariate regressions', *JASA* **85**, 1060 – 1068.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000), *Continuous multivariate distributions, vol. 1: models and applications, 2nd ed.*, Wiley, New York.

Krasker, W. S. and Welsch, R. E. (1982), 'Efficient bounded-influence regression estimation', *J. Amer. Statist. Assoc.* **77**, 595–604.

Krishnaiah, P. R., Hagis, P. and Steinberg, L. (1963), 'A note on the bivariate chi distribution', *SIAM Review* **5**(2), 140–144.

Krishnan, M. (1967), 'The noncentral bivariate chi distribution.', *SIAM Review* **9**, 708–714.

Lambert, A., Saleh, A. K. M. E. and Sen, P. K. (1985a), 'On least squares estimation of intercepts after a preliminary test on parallelism of regression lines', *Comm. Statist.- theor. meth.* **14**, 793–807.

Lambert, A., Saleh, A. K. M. E. and Sen, P. K. (1985b), 'Test of homogeneity of intercepts after a preliminary test on parallelism of several regression lines: from parametric to asymptotic', *Comm. Statist.- theo. meth.* **14**(9), 2243–2259.

Lindgren, B. W. (1993), *Statistical theory, 4th Ed*, Chapman & Hall, US.

Maatta, J. M. and Casella, G. (1990), 'Developments in decision-theoretic variance estimation', *Statistical Sciences* **5**, 90–101.

Mahdi, T. N., Ahmad, S. E. and Ahsanullah, M. (1998), 'Improves prediction: Pooling two identical regression lines', *Jour. Aplied. Statist. Sc.* **7**, 63–86.

Markatou, M. and He, X. (1994), 'Bounded-influence and high-breakdown-point testing procedures in linear models', *JASA* **89**, 543–549.

Markatou, M. and Hettmansperger, T. P. (1990), 'Robust bounded-influence tests in linear models', *J. Amer. Statist. Assoc.* **85**, 187–190.

Marshall, A. W. and Olkin, I. (1990), Multivariate distributions generated from mixtures of convolution and product families, In: H.W. Block, A.R. Sampson & T.H. Savits (Eds). *Topics in Statistical Dependence*, IMS Lecture Notes-Monograph Series, Hayward, CA, pp. 371–393.

Mead, R., Bancroft, T. A. and Han, C. P. (1975), 'Power of analysis of variance test procedures for incompletely specified fixed models', *Ann. Statist.* **3**, 797–808.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2001), *Introduction to linear regression analysis*, Wiley, New York.

Mosteller, F. (1948), 'On pooling data', *Jour American Stat. Soc.* **43**(242), 231–242.

Muirhead, R. J. (1982), *Aspects of multivariate statistical theory*, John Wiley & Sons, New York.

Muller, C. (1998), 'Optimum robust testing in linear models', *Ann. Statist.* **26**, 1126–1146.

Ohtani, K. (1988), 'Optimal levels of significance of a pre-test in estimating the disturbance variance after the pre-test for a linear hypothesis on coefficients in a linear regression', *Economics Letters* **28**(2), 151–156.

Ohtani, K. and Giles, J. (1993), 'Testing linear restrictions on coefficients in a linear regression model with proxy variables and spherically symmetric disturbances', *Journal of Econometrics* **57**(1-3), 393–406.

Ohtani, K. and Toyoda, T. (1985), 'Testing linear hypothesis on regression coefficients after a pre-test for disturbance variance', *Economics Letters* **17**(1-2), 111–114.

Paull, A. E. (1950), 'On a preliminary test for pooling mean squares in the analysis of variance', *Ann. Math. Statist.* **21**, 539–556.

Pinto, J. V., Ng, P. and Allen, D. S. (2003), 'Logical extremes, beta, and the power of the test', *Journal of Statistics Education [Online]* **11**.
URL: *www.amstat.org/publications/jse/v11n1/pinto.html*

Puri, M. L. and Sen, P. K. (1985), *Nonparametric methods in general linear models*, Wiley, New York.

Rieder, H. (1982), 'Qualitative robustness of rank tests', *Ann. Statist.* **10**, 205–211.

Rieder, H. (1994), *Robust Asymptotic Statistics*, Springer-Verlag, New York.

Rosenberger, J. L. and Gasko, M. (1983), Comparing location estimators: timmed means, medians and trimean, In: D.C. Hoaglin, F. Mosteller & J.W. Tukey (Eds). *Understanding robust and explanatory data analysis.* Wiley. New York, pp. 297–338.

Rousseeuw, P. J. (1984), 'Least median of squares regression', *J. Amer. Statist. Assoc.* **79**, 871–880.

Rousseeuw, P. J. and Leroy, A. M. (1987), *Robust regression and outlier detection*, Wiley, New York.

Rousseeuw, P. J. and Van Driessen, K. V. (2006), 'Computing lts regression for large data sets', *Data Mining and Knowledge Discovery* **12**, 29–45.

Royen, T. (1995), 'On some central and non-central multivariate chi-square distributions.', *Statistica Sinica* **5**, 373–397.

Ryan, T. P. (1997), *Modern regression methods*, Wiley, New York.

Saleh, A. K. M. E. (2006), *Theory of Preliminary test and Stein-type estimation with applications*, John Wiley & Sons, New Jersey.

Saleh, A. K. M. E. and Han, C. P. (1990), 'Shrinkage estimation in regression analysis', *Estadistica* **42**, 40–63.

Saleh, A. K. M. E. and Sen, P. K. (1978), 'Nonparametric estimation of location parameter after a preliminary test on regression', *Annals of Statistics* **6**, 154–168.

Saleh, A. K. M. E. and Sen, P. K. (1982), 'Non-parametric tests for location after a preliminary test on regression', *Comm. in Statist.: Theory and Methods* **11**, 639–651.

Saleh, A. K. M. E. and Sen, P. K. (1983), 'Nonparametric tests of location after a preliminary test on regression in the multivariate case', *Comm. in Statist.: Theory and Methods* **12**(16), 1855–1872.

Saleh, A. K. M. E. and Sen, P. K. (1985), Shrinkage least squares estimation in a general multivariate linear model, Proceedings of the Fifth Pannonian Symposium on Mathematical Statist., pp. 307–325.

Saleh, A. K. M. E. and Sen, P. K. (1987), 'Relative performance of Stein-rule and preliminary test estimators in linear models least squares theory', *Comm. in Statist.: Theory and Methods* **16**(2), 461–476.

Saleh, A. K. M. E. and Shiraishi, T. (1989), 'On some R- and M-estimators of regression parameters under uncertain restriction', *J.Japan Stat. Soc.* **19**, 129–137.

Schrader, R. M. and Hettmansperger, T. P. (1980), 'Robust analysis of variance based upon a likelihood ratio criterion', *Biometrika* **67**, 93–101.

Schrader, R. M. and McKean, J. W. (1977), 'Robust analysis of variance', *Comm. Statist.- theor. Meth.* **6**, 879–894.

Sclove, S. L., Morris, C. and Rao, C. R. (1972), 'Non-optimality of preliminary test estimators for the mean of a multivariate normal distribution', *Ann. Math. Statist.* **43**, 1481–1490.

Sen, P. K. (1982), 'On M-tests in linear models', *Biometrika* **69**(1), 245–248.

Sen, P. K. and Saleh, A. K. M. E. (1979), 'Nonparametric estimation of location parameter after a preliminary test on regression in the multivariate case', *Jour. Multivariate Analysis* **9**(2), 322–331.

Sen, P. K. and Saleh, A. K. M. E. (1987), 'On preliminary test and shrinkage M-estimation in linear models', *Ann. Statist.* **15**, 1580–1592.

Shiraishi, T. (1990), 'M-tests in multivariate models', *Metrika* **37**, 189–197.

Silvapulle, M. J. (1992), 'Robust Wald type tests of one sided hypotheses in linear model', *J. Amer. Statist. Assoc.* **87**, 156–161.

Silvey, S. D. (1975), *Statistical Inference (3rd ed.)*, Chapman and Hall, London.

Snedecor, G. W. (1938), *Statistical Methods*, Collegiate Press, Iowa.

Stein, C. (1956), 'Inadmissibility of the usual estimator for the mean of a multivariate normal distribution.', Proc. 3rd Berkeley Sympos. Math. Statist. Probability 1, 197-206.

Stein, C. (1981), 'Estimation of the mean of a multivariate normal distribution', *Ann. Statist.* **9**, 1135–1151.

Steyn, H. S. and Roux, J. J. (1972), 'Approximations for the noncentral Wishart distribution', *South African Statist. J.* **6**, 165–173.

Tamura, R. (1965), 'Nonparametric inferences with a preliminary test', *Bull. Math. Stat* **11**, 38–61.

Thisted, R. A. (1988), *Elements of Statistical Computing: Numerical Computation*, Chapman & Hall, New York.

Toyoda, T. and Ohtani, K. (1986), 'Testing equality between sets of coefficients after a preliminary test for equality of disturbance variances in two linear regressions', *Journal of Econometrics* **31**(1), 67–80.

van der Vaart, A. W. (1998), *Asymptotic statistics*, Cambridge University Press, UK.

Wackerly, D., Mendenhall, W. and Scheaffer, R. L. (2008), *Mathematical Statistics with Applications, 7th Ed.*, Thompson, Belmont, CA, USA.

Wald, A. (1943), 'Tests of statistical hypotheses concerning several parameters when the number of observations is large', *Trans. Am. Math.Soc.* **54**, 426–482.

Wang, J. and Wang, J. (2007), Mitigating the effect of multiple outliers on GNSS navigation with M-estimation schemes, International Global Navigation Satellite Systems Society IGNSS Symposium, The University of New South Wales, Sydney, Australia, 4-6 Dec.

Wang, L. and Qu, A. (2007), 'Robust tests in regression models with omnibus alternatives and bounded influence', *JASA* **109**, 347 – 358.

Wilcox, R. R. (2005), *Introduction to Robust Estimation and Hypothesis Testing*, Elsevier Inc, US.

Wright, K. and Kennedy, W. J. (2002), 'Self-validated computations for the probabilities of the central bivariate chi-square distribution and a bivariate F distribution.', *J. Statist. Comput. Simul.* **72**, 63–75.

Wu, X., Yang, Y. and Zhao, L. (2007), 'Approximation by random weighting method for M-test in linear models', *Science in China Series A: Mathematics* **50**, 87–99.

Yohai, V. J. (1987), 'High breakdown-point and high efficiency robust estimates for regression', *Ann. Stat.* **15**, 642–656.

Yohai, V. J. and Zamar, R. H. (1988), 'High breakdown point estimates of regression by means of the minimization of an efficient scale', *J. Amer. Statist. Assoc.* **83**, 406–413.

Yunus, R. M. and Khan, S. (2009), The bivariate noncentral chi-square distribution - a compound distribution approach, USQ Working Paper Series SC-MC-0902.
URL: *http://www.sci.usq.edu.au/research/workingpapers.php*

Zhang, J. and Li, G. (1998), 'Breakdown properties of location M-estimators', *Ann. Statist.* **26**, 1170–1189.

Appendix A

A.1 Regularity Conditions

(Jurečková and Sen, 1996, p.217) The ψ function is decomposed into the sum

$$\psi = \psi_a + \psi_c + \psi_s,$$

where ψ_a is an absolutely continuous function with an absolutely continuous derivative, ψ_c is a continuous, piecewise linear function that is constant in a neighbourhood of $\pm\infty$, and ψ_s is a nondecreasing step function. The following conditions imposed on (3.2.1):

M1. $S_n(\boldsymbol{X})$ is a regression-invariant and scale-equivariant, $S_n > 0$ a.s. and

$$n^{\frac{1}{2}}(S_n - S) = O_p(1)$$

for some functional $S = S(F) > 0$.

M2. The function $h(t) = \int \rho((z-t)/S)dF(z)$ has a unique minimum at $t = 0$.

M3. For some $\delta > 0$ and $\eta > 1$,

$$\int_{-\infty}^{\infty} \left\{ |z| \sup_{|u|\leq\delta} \sup_{|v|\leq\delta} |\psi_a''\left(\frac{e^{-v}(z+u)}{S}\right)| \right\}^{\eta} dF(z) < \infty$$

and

$$\int_{-\infty}^{\infty} \left\{ |z|^2 \sup_{|u|\leq\delta} |\psi_a''\left(\frac{e^{-v}(z+u)}{S}\right)| \right\}^{\eta} dF(z) < \infty,$$

where the derivative of $\psi_a(\cdot)$ are taken with respect to z.

M4. ψ_c is a continuous, piecewise linear function with knots at $-\infty = \mu_0 < \mu_1, \ldots, \mu_k < \mu_{k+1} = \infty$, that is constant in a neighborhood of $\pm\infty$. Further we assume that $\frac{dF(z)}{dz}$ is bounded in neighborhoods of S_{μ_i}, $i = 0, 1, \ldots, k+1$.

M5. $\psi_s(z) = \lambda_v,\ q_v < z \leq q_{v+1}, v = 1, \ldots, m$, where $-\infty = q_0 < q_1 < \ldots < q_{m+1} = \infty$ and $-\infty < \lambda_0 < \lambda_1 < \ldots < \lambda_m < \infty$. We assume that the first and second derivative of $F(z)$ are bounded in neighborhoods of $S_{q_i},\ i = 0, \ldots, m+1$.

The class of ψ_a covers the ML score function $\psi_{ML}(U_i) = U_i$, where $U_i = \frac{X_i - \theta - \beta c_i}{S_n}$ while the class of ψ_c covers the Huber score function

$$\psi_H(U_i) = \begin{cases} U_i & |U_i| \leq k, \\ k\ \mathrm{sign}(U_i) & |U_i| > k, \end{cases} \tag{A.1.1}$$

where k is known as the tuning constant and it is chosen to achieve desired efficiencies.

A.2 Le Cam's Lemma

Le Cam's first lemma (see Hájek et al., 1999, p.251)
For any likelihood ratio statistic $L_v(x_v)$,

$$L_v(x_v) = \begin{cases} q_v(x_v)/p_v(x_v), & \text{if} \quad p_v(x_v) > 0, \\ 1, & \text{if} \quad p_v(x_v) = q_v(x_v) = 0, \\ \infty, & \text{if} \quad 0 = p_v(x_v) < q_v(x_v), \end{cases}$$

where x_v denotes the typical point of the space $X_v,\ v \geq 1$, we find $\{Q_v\}$ is contiguous to $\{P_v\}$ if

$$\log L_v(X) \xrightarrow{D} N\left(-\frac{1}{2}\,\sigma^2, \sigma^2\right) \quad (\text{under } \{P_v\}),$$

with $\sigma^2 > 0$.

Le Cam's second lemma (see Hájek et al., 1999, p.253)
Assume that

$$\lim_{v\to\infty} \max_{1\leq i\leq N_v} P_v\left(\left|\frac{g_{vi}(X_i)}{f_{vi}(X_i)} - 1\right|\right) = 0 \tag{A.2.2}$$

and statistics

$$W_v = 2\sum_{i=1}^{N_v}\left\{\sqrt{g_{vi}(X_i)/f_{vi}(X_i)} - 1\right\} \tag{A.2.3}$$

are asymptotically normal $(-\frac{1}{4}\sigma^2, \sigma^2)$ under P_v, with $p_v(x_v) = \prod_{i=1}^{N_v} f_{vi}(x_i)$, $q_v(x_v) = \prod_{i=1}^{N_v} g_{vi}(x_i)$ and $x_v = (x_1, \ldots, x_{N_v})$. Then, statistics $\log L_v$ satisfy

$$\lim_{v\to\infty} P_v \left(|\log L_v - W_v + \frac{1}{4}\sigma^2| > \epsilon \right) = 0, \quad \epsilon > 0, \tag{A.2.4}$$

and are asymptotically normal $(-\frac{1}{2}\sigma^2, \sigma^2)$ under P_v.

Le Cam's third lemma (see Hájek et al., 1999, p.257)
If

$$\begin{bmatrix} T_v \\ \log L_v \end{bmatrix} \xrightarrow{D} N_2 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right) \quad \text{(under } \{P_v\}\text{)},$$

where T_v is a statistic with $\mu_2 = -\frac{1}{2}\sigma_{22}$, then

$$T_v \xrightarrow{D} N(\mu_1 + \sigma_{12},\ \sigma_{11}) \quad \text{(under } \{Q_v\}\text{)}.$$

The Le Cam's second lemma (Hájek et al., 1999, p.253) gives conditions when $\log L_v \xrightarrow{D} N(-\frac{1}{2}\,\sigma^2, \sigma^2)$.

Appendix B

Proof of Theorems

B.1 Simple Regression Model

Interested readers are referred to papers by Jurečková (1977), Sen (1982) and Jurečková and Sen (1996, p.221) for the following asymptotic properties. For simplicity, we omit condition (3.2.17) and let $S_n = S$ in equation (5.5.29) of Jurečková and Sen (1996, p.221),

(i) Under $H_0^{\star(2)} : \beta = \beta_0$, for every positive K, as $n \to \infty$, in probability

$$\begin{aligned}\sup\{n^{-\frac{1}{2}}|M_{n_1}\{(\theta, \beta_0) + (t_1, t_2)\} - M_{n_1}(\theta, \beta_0) + n\gamma(t_1 + t_2\bar{c})| :\\ |t_1|, |t_2| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} 0,\end{aligned} \tag{B.1.1}$$

$$\begin{aligned}\sup\{n^{-\frac{1}{2}}|M_{n_2}\{(\theta, \beta_0) + (t_1, t_2)\} - M_{n_2}(\theta, \beta_0) +\\ n\gamma\{t_1\bar{c} + t_2(C^{\star 2} + \bar{c}^2)\}| : |t_1|, |t_2| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} 0.\end{aligned} \tag{B.1.2}$$

(ii) Under $H_0 : \theta = 0, \beta = 0$, as n grows large,

$$n^{-\frac{1}{2}} \begin{pmatrix} M_{n_1}(0,0) \\ M_{n_2}(0,0) \end{pmatrix} \xrightarrow{d} N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \right], \tag{B.1.3}$$

where $N_2(\cdot\,,\cdot\,)$ represents a bivariate normal distribution with appropriate parameters.

(iii) Under $H_0 : \theta = 0, \beta = 0$,

$$\begin{aligned}\sup\{n^{-\frac{1}{2}}|M_{n_1}(a, b) - M_{n_1}(0, 0) + n\gamma(a + b\bar{c})| :\\ |a|, |b| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} 0,\end{aligned} \tag{B.1.4}$$

$$\sup\{n^{-\frac{1}{2}}|M_{n_2}(a,b) - M_{n_2}(0,0) + n\gamma\{a\bar{c} + b(C^{\star 2} + \bar{c}^2)\}| : \\ |a| \leq n^{-\frac{1}{2}}K, |b| \leq n^{-\frac{1}{2}}K\} \xrightarrow{p} 0 \quad \text{(B.1.5)}$$

as $n \to \infty$, and K is a positive constant. The above convergence is in probability, meaning the sequences of random variables converge in probability to a fix value (0).

Proof of part (a) of Theorem 3.8.1: Under $H_0^{\star(2)} : \beta = \beta_0$, we obtain

$$n^{-\frac{1}{2}}M_{n_2}(\check{\theta}, \beta_0) = n^{-\frac{1}{2}}M_{n_2}(\theta, \beta_0) - \bar{c}n^{-\frac{1}{2}}M_{n_1}(\theta, \beta_0) + o_p(1). \quad \text{(B.1.6)}$$

by equations (3.8.2), (B.1.1) and (B.1.2).

Further, the distribution of $n^{-\frac{1}{2}}M_{n_2}(\check{\theta}, \beta_0)$ under $H_0^{\star(2)} : \beta = \beta_0$ is the same as the distribution of $n^{-\frac{1}{2}}M_{n_2}(0,0) - n^{-\frac{1}{2}}\bar{c}M_{n_1}(0,0)$ under $H_0 : \theta = 0, \beta = 0$ using equation (B.1.6) and due to the fact that the distribution of $M_{n_1}(a,b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ when $\theta = 0, \beta = 0$, and similar to $M_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332).

Therefore, by utilizing equation (B.1.3), we find

$$n^{-\frac{1}{2}}M_{n_2}(\check{\theta}, \beta_0) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2}).$$

under $H_0^{\star(2)} : \beta = \beta_0$ as $n \to \infty$.

The proof of parts (b) and (c) is similarly obtained.

Proof of part (ii) of Theorem 3.8.2: Note that under (3.1.1), (3.2.4), (3.2.5) and (3.8.10), the contiguity of the sequence of probability measures under $\{K_n^\star\}$ to those under $H_0^\star$ follows from Le Cam's first and second lemmas (Hájek et al., 1999, Ch.7). We are interested in the asymptotic joint distribution of the joint statistics $\left[n^{-\frac{1}{2}}T_n^{\star RT}, \ n^{-\frac{1}{2}}T_n^{\star PT}\right]$. Here, convergence of $\left[n^{-\frac{1}{2}}T_n^{\star RT}, n^{-\frac{1}{2}}T_n^{\star PT}\right]$ $+\Upsilon \to [0,0]$ under $H_0^\star$ implies $\left[n^{-\frac{1}{2}}T_n^{\star RT}, \ n^{-\frac{1}{2}}T_n^{\star PT}\right] + \Upsilon \to [0,0]$ under $\{K_n^\star\}$ since the probability measures under $\{K_n^\star\}$ are contiguous to those under $H_0^\star$ (c.f. Saleh, 2006, p.44). Here, Υ is a known vector.

From Jurečková (1977), under $H_0^\star : \theta = \theta_0, \beta = \beta_0$ for every positive K, as $n \to \infty$,

$$\sup\{n^{-\frac{1}{2}}|M_{n_1}\{(\theta_0, \beta_0) + (t_1, t_2)\} - M_{n_1}(\theta_0, \beta_0) + n\gamma(t_1 + t_2\bar{c})| : \\ |t_1|, |t_2| \leq n^{-\frac{1}{2}}K\} \xrightarrow{p} 0, \quad \text{(B.1.7)}$$

$$\sup\{n^{-\frac{1}{2}}|M_{n_2}\{(\theta_0,\beta_0)+(t_1,t_2)\}-M_{n_2}(\theta_0,\beta_0)+ \\ n\gamma\{t_1\bar{c}+t_2(C^{\star 2}+\bar{c}^2)\}|:|t_1|,|t_2|\le n^{-\frac{1}{2}}K\} \xrightarrow{p} 0. \tag{B.1.8}$$

Under $H_0^\star:\theta=\theta_0,\ \beta=\beta_0$, with relation to (B.1.7) and (B.1.8)

$$n^{-\frac{1}{2}}M_{n_1}(\check{\theta},\beta_0)=n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0)-n^{1/2}\gamma(\check{\theta}-\theta_0)+o_p(1) \text{ and} \tag{B.1.9}$$

$$n^{-\frac{1}{2}}M_{n_2}(\check{\theta},\beta_0)=n^{-\frac{1}{2}}M_{n_2}(\theta_0,\beta_0)-n^{1/2}\gamma(\check{\theta}-\theta_0)\bar{c}+o_p(1). \tag{B.1.10}$$

Recalling definition (3.8.2), the equation (B.1.9) reduces to

$$n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0)=n^{1/2}\gamma(\check{\theta}-\theta_0)+o_p(1), \tag{B.1.11}$$

and hence the equation (B.1.10) becomes

$$n^{-\frac{1}{2}}M_{n_2}(\check{\theta},\beta_0)=n^{-\frac{1}{2}}M_{n_2}(\theta_0,\beta_0)-n^{-\frac{1}{2}}\bar{c}M_{n_1}(\theta_0,\beta_0)+o_p(1). \tag{B.1.12}$$

Therefore, under $H_0^\star$, we find

$$\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0) \\ n^{-\frac{1}{2}}M_{n_2}(\check{\theta},\beta_0)\end{bmatrix}-\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0) \\ n^{-\frac{1}{2}}M_{n_2}(\theta_0,\beta_0)-n^{-\frac{1}{2}}\bar{c}M_{n_1}(\theta_0,\beta_0)\end{bmatrix} \tag{B.1.13}$$

$$=\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0) \\ n^{-\frac{1}{2}}M_{n_2}(\check{\theta},\beta_0)\end{bmatrix}-\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1\end{bmatrix}\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0) \\ n^{-\frac{1}{2}}M_{n_2}(\theta_0,\beta_0)\end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ 0\end{bmatrix} \tag{B.1.14}$$

from equations (B.1.7) and (B.1.12). Now utilizing the contiguity of probability measures under $\{K_n^\star\}$ to those under $H_0^\star$, the equation (B.1.14) implies that

$$\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0) & n^{-\frac{1}{2}}M_{n_2}(\check{\theta},\beta_0)\end{bmatrix}',$$

which under $\{K_n^\star\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1\end{bmatrix}\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0,\beta_0) \\ n^{-\frac{1}{2}}M_{n_2}(\theta_0,\beta_0)\end{bmatrix}$$

under $H_0^\star$. But the asymptotic distribution of the above random vector under $\{K_n^\star\}$ is the same as

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1\end{bmatrix}\begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\theta_0-n^{-\frac{1}{2}}\delta_1,\beta_0-n^{-\frac{1}{2}}\delta_2) \\ n^{-\frac{1}{2}}M_{n_2}(\theta_0-n^{-\frac{1}{2}}\delta_1,\beta_0-n^{-\frac{1}{2}}\delta_2)\end{bmatrix}$$

under $H_0^\star$ due to the fact that the distribution of $M_{n_1}(a,b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ under $\theta = 0, \beta = 0$, and similar to $M_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332).

Note that under $H_0^\star : \theta = \theta_0, \beta = \beta_0$, from (B.1.7) and (B.1.8),

$$\begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0 - n^{-\frac{1}{2}}\delta_1, \beta_0 - n^{-\frac{1}{2}}\delta_2) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0 - n^{-\frac{1}{2}}\delta_1, \beta_0 - n^{-\frac{1}{2}}\delta_2) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \beta_0) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0, \beta_0) \end{bmatrix}$$
$$- \begin{bmatrix} \gamma(\delta_1 + \delta_2 \bar{c}) \\ \gamma\{\delta_1 \bar{c} + \delta_2(C^{\star 2} + \bar{c}^2)\} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{B.1.15}$$

Hence, by equation (3.8.5), under $H_0^\star$,

$$\begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0 - n^{-\frac{1}{2}}\delta_1, \beta_0 - n^{-\frac{1}{2}}\delta_2) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0 - n^{-\frac{1}{2}}\delta_1, \beta_0 - n^{-\frac{1}{2}}\delta_2) \end{bmatrix} \xrightarrow{d}$$
$$N_2 \left[\begin{pmatrix} \gamma(\delta_1 + \delta_2 \bar{c}) \\ \gamma\{\delta_1 \bar{c} + \delta_2(C^{\star 2} + \bar{c}^2)\} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \right]. \tag{B.1.16}$$

Thus, as $n \to \infty$, the distribution of

$$\begin{bmatrix} n^{-\frac{1}{2}} T_n^{\star RT} & n^{-\frac{1}{2}} T_n^{\star PT} \end{bmatrix}' = \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \beta_0) & n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) \end{bmatrix}'$$

under $\{K_n^\star\}$ is bivariate normal with mean vector

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} \gamma(\delta_1 + \delta_2 \bar{c}) \\ \gamma\{\delta_1 \bar{c} + \delta_2(C^{\star 2} + \bar{c}^2)\} \end{bmatrix} = \begin{bmatrix} \gamma(\delta_1 + \delta_2 \bar{c}) \\ \gamma \delta_2 C^{\star 2} \end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix}' = \sigma_0^2 \begin{bmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{bmatrix}.$$

Since the two statistics $n^{-\frac{1}{2}} T_n^{\star RT}$ and $n^{-\frac{1}{2}} T_n^{\star PT}$ are uncorrelated, asymptotically, they are independently distributed normal variables.

Proof of part (i) of Theorem 3.8.2: Under $H_0^\star : \theta = \theta_0, \beta = \beta_0$, using equations (3.8.1), (B.1.7), (B.1.8) and (B.1.12), we find

$$\begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \check{\beta}) \\ n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \beta_0) - \frac{n^{-\frac{1}{2}} \bar{c}}{(C^{\star 2}+\bar{c}^2)} M_{n_2}(\theta_0, \beta_0) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0, \beta_0) - n^{-\frac{1}{2}} \bar{c} M_{n_1}(\theta_0, \beta_0) \end{bmatrix}$$

$$= \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \check{\beta}) \\ n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) \end{bmatrix} - \begin{bmatrix} 1 & -\frac{\bar{c}}{(C^{\star 2}+\bar{c}^2)} \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \beta_0) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0, \beta_0) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{B.1.17}$$

Now by using the contiguity of probability measures under $\{K_n^\star\}$ to those under $H_0^\star$, the equation (B.1.17) implies that

$$\begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \check{\beta}) & n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) \end{bmatrix}'$$

under $\{K_n^\star\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} 1 & -\bar{c}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \beta_0) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0, \beta_0) \end{bmatrix}.$$

But the asymptotic distribution of the above random vector under $\{K_n^\star\}$ is the same as

$$\begin{bmatrix} 1 & -\bar{c}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0 - n^{-\frac{1}{2}}\delta_1, \beta_0 - n^{-\frac{1}{2}}\delta_2) \\ n^{-\frac{1}{2}} M_{n_2}(\theta_0 - n^{-\frac{1}{2}}\delta_1, \beta_0 - n^{-\frac{1}{2}}\delta_2) \end{bmatrix}$$

under $H_0^\star$ due to the fact that the distribution of $M_{n_1}(a, b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ under $\theta = 0, \beta = 0$ and similar to $M_{n_2}(0, 0)$ (c.f. Saleh, 2006, p.332).

Then it follows that by equation (B.1.16), as n grows large, the distribution of

$$\begin{bmatrix} n^{-\frac{1}{2}} T_n^{\star UT} & n^{-\frac{1}{2}} T_n^{\star PT} \end{bmatrix}' = \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\theta_0, \check{\beta}) & n^{-\frac{1}{2}} M_{n_2}(\check{\theta}, \beta_0) \end{bmatrix}'$$

is bivariate normal with mean vector

$$\begin{bmatrix} 1 & -\bar{c}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} \gamma(\delta_1 + \delta_2 \bar{c}) \\ \gamma\{\delta_1 \bar{c} + \delta_2 (C^{\star 2} + \bar{c}^2)\} \end{bmatrix} = \begin{bmatrix} \gamma \delta_1 C^{\star 2}/(C^{\star 2} + \bar{c}^2) \\ \gamma \delta_2 C^{\star 2} \end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix} 1 & -\bar{c}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c} & 1 \end{bmatrix} \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \begin{bmatrix} 1 & -\bar{c}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c} & 1 \end{bmatrix}'$$

$$= \sigma_0^2 \begin{bmatrix} C^{\star 2}/(C^{\star 2} + \bar{c}^2) & -\bar{c}\, C^{\star 2}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c}\, C^{\star 2}/(C^{\star 2} + \bar{c}^2) & C^{\star 2} \end{bmatrix}.$$

Clearly, the two test statistics $n^{-\frac{1}{2}}T_n^{\star UT}$ and $n^{-\frac{1}{2}}T_n^{\star PT}$ are not independent, but rather correlated.

B.2 Multivariate Simple Regression Model

The following asymptotic results of Jurečková (1977), Sen (1982) and Jurečková and Sen (1996, p.221) are used in deriving the distribution of the proposed tests. For simplicity, we assume S is known or consider the nonstudentized M-estimator, so we omit condition M1 of Jurečková and Sen (1996, p.217) and let $S_n = S$ in equation (5.5.29) of Jurečková and Sen (1996, p.221). Thus,

(i) Under $\theta_j = a_j, \beta_j = b_j$ and as n grows large,

$$\sup \Big\{ n^{-\frac{1}{2}} |M_{n_1 j}\{(a_j, b_j) + (t_{1_j}, t_{2_j})\} - M_{n_1 j}(a_j, b_j) + n\gamma_j (t_{1_j} + t_{2_j}\bar{c})| : |t_{1_j}| \le n^{-\frac{1}{2}} K_1, |t_{2_j}| \le n^{-\frac{1}{2}} K_2 \Big\} \xrightarrow{p} 0, \tag{B.2.1}$$

$$\sup \Big\{ n^{-\frac{1}{2}} |M_{n_2 j}\{(a_j, b_j) + (t_{1_j}, t_{2_j})\} - M_{n_2 j}(a_j, b_j) + n\gamma_j (t_{1_j}\bar{c} + t_{2_j}(C^{\star 2} + \bar{c}^2))| : |t_{1_j}| \le n^{-\frac{1}{2}} K_1, |t_{2_j}| \le n^{-\frac{1}{2}} K_2 \Big\} \xrightarrow{p} 0, \tag{B.2.2}$$

where K_1, K_2 are positive constants.

(ii) Under $\boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\beta} = \mathbf{0}$, as n grows large,

$$n^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ \boldsymbol{M}_{n_2}(\mathbf{0},\mathbf{0}) \end{pmatrix} \xrightarrow{d} N_{2p} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \otimes \boldsymbol{\Lambda} \right], \tag{B.2.3}$$

where $N_{2p}(\cdot\ ,\cdot\ )$ represents a $2p$-variate normal distribution with appropriate parameters.

Proof of part (iii) of Theorem 4.2.1: By equations (4.2.6), (B.2.1) and (B.2.2), under $H_0^{(3)}$ we obtain

$$n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) = n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\boldsymbol{\theta}, \boldsymbol{\beta}_0) - n^{-\frac{1}{2}} \bar{c} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}, \boldsymbol{\beta}_0) + \boldsymbol{o}_p(1). \tag{B.2.4}$$

Further, the distribution of $n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)$ under $H_0^{(3)}$ is the same as the distribution of $n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\mathbf{0},\mathbf{0}) - n^{-\frac{1}{2}} \bar{c} \boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0})$ under $H_0 : \boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\beta} = \mathbf{0}$ using equation (B.2.4) and the fact that the distribution of $M_{n_1}(a, b)$ under $\theta = a, \beta = b$

is the same as that of $M_{n_1}(\theta - a, \beta - b)$ when $\theta = 0, \beta = 0$, and similar to $M_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332).

Therefore, utilizing equation (B.2.3), under $H_0^{(3)} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ as $n \to \infty$, the proof of part (iii) of Theorem 4.2.1 is completed.

Proof of part(i) of Theorem 4.4.1: Under $H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0,\ \boldsymbol{\beta} = \boldsymbol{\beta}_0$, with relation to (B.2.1) and (B.2.2),

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) - n^{\frac{1}{2}}\boldsymbol{\gamma}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(1), \tag{B.2.5}$$

and

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) - n^{\frac{1}{2}}\bar{c}\boldsymbol{\gamma}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(1). \tag{B.2.6}$$

The equation (B.2.5) is then reduced to

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) = n^{\frac{1}{2}}\boldsymbol{\gamma}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(1) \tag{B.2.7}$$

by definition (4.2.6). Substituting equation (B.2.7) in equation (B.2.6) yields

$$\begin{aligned}
& \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) - n^{-\frac{1}{2}}\bar{c}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{bmatrix} \\
= & \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) \end{bmatrix} - \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{bmatrix} \\
& \xrightarrow{p} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}.
\end{aligned} \tag{B.2.8}$$

Now utilizing the contiguity of probability measures under $\{K_n\}$ to those under $H_0^{(2)}$, the equation (B.2.8) implies that under $\{K_n\}$ $[n^{-\frac{1}{2}}, \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0), n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)]'$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{bmatrix}$$

under $H_0^{(2)}$. But the asymptotic distribution of the above random vector under $\{K_n\}$ is the same as

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_2) \end{bmatrix}$$

under $H_0^{(2)}$ due to the fact that the distribution of $M_{n_1}(a,b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ under $\theta = 0, \beta = 0$, and similar to $M_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332).

Note that under $H_0^{(2)}$, by equations (B.2.1), (B.2.2) and part (ii) of Theorem 4.2.1,

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_2) \end{bmatrix} \xrightarrow{d}$$
$$N_{2p}\left(\begin{pmatrix} \boldsymbol{\gamma}(\boldsymbol{\varrho}_1 + \boldsymbol{\varrho}_2\bar{c}) \\ \boldsymbol{\gamma}\{\boldsymbol{\varrho}_1\bar{c} + \boldsymbol{\varrho}_2(C^{\star 2} + \bar{c}^2)\} \end{pmatrix}, \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \otimes \boldsymbol{\Lambda}\right). \quad \text{(B.2.9)}$$

Thus, the distribution of $[n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0), n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)]'$ under $\{K_n\}$ is a $2p$-variate normal with mean vector

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}(\boldsymbol{\varrho}_1 + \boldsymbol{\varrho}_2\bar{c}) \\ \boldsymbol{\gamma}\{\boldsymbol{\varrho}_1\bar{c} + \boldsymbol{\varrho}_2(C^{\star 2} + \bar{c}^2)\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}(\boldsymbol{\varrho}_1 + \boldsymbol{\varrho}_2\bar{c}) \\ \boldsymbol{\gamma}\boldsymbol{\varrho}_2 C^{\star 2} \end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \otimes \boldsymbol{\Lambda} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix}' = \begin{bmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{bmatrix} \otimes \boldsymbol{\Lambda}. \quad \text{(B.2.10)}$$

Since the two statistics $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$ and $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)$ are uncorrelated, asymptotically, they are independently distributed normal variables.

Proof of part(ii) of Theorem 4.4.1: Under $H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\beta} = \boldsymbol{\beta}_0$, by equations (4.2.5), (4.2.6), (B.2.1) and (B.2.2), as $n \to \infty$,

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0) \end{bmatrix} - \begin{bmatrix} \boldsymbol{I}_p & \frac{-\bar{c}}{C^{\star 2}+\bar{c}^2}\boldsymbol{I}_p \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{bmatrix}$$
$$\xrightarrow{p} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}. \quad \text{(B.2.11)}$$

Now using the contiguity of probability measures under $\{K_n\}$ to those under $H_0^{(2)}$, the equation (B.2.11) implies that $[n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}), n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)]'$ under $\{K_n\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} \boldsymbol{I}_p & -\frac{\bar{c}}{C^{\star 2}+\bar{c}^2}\boldsymbol{I}_p \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{bmatrix}$$

under $H_0^{(2)}$. But the asymptotic distribution of the above random vector under $\{K_n\}$ is the same as

$$\begin{bmatrix} \boldsymbol{I}_p & \frac{-\bar{c}}{C^{\star 2}+\bar{c}^2}\boldsymbol{I}_p \\ -\bar{c}\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_1, \boldsymbol{\beta}_0 - n^{-\frac{1}{2}}\boldsymbol{\varrho}_2) \end{bmatrix}$$

under $H_0^{(2)}$.

The proof then follows from equations (B.2.3) and (B.2.9). Clearly, the two test statistics $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}})$ and $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\beta}_0)$ are not independent, but rather correlated.

B.3 Parallelism Model

The following asymptotic results of Jurečková (1977), Sen (1982)and Jurečková and Sen (1996, p.221) are used in deriving the distribution of the proposed tests. For simplicity, we assume S is known or consider the nonstudentized M-estimator, so we omit condition M1 of Jurečková and Sen (1996, p.217) and let $S_n = S$ in equation (5.5.29) of Jurečková and Sen (1996, p.221). Thus,

(i) Under $\theta_j = a_j, \beta_j = b_j$ and as n grows large,

$$\begin{aligned} &\sup\{n^{-\frac{1}{2}}|M_{n_1}^{(j)}\{(a_j, b_j) + (t_{1_j}, t_{2_j})\} - M_{n_1}^{(j)}(a_j, b_j) + \\ &\quad \lambda_j n\gamma(t_{1_j} + t_{2_j}\bar{c}_j)| : |t_{1_j}| \le n^{-\frac{1}{2}}K, |t_{2_j}| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} 0, \qquad \text{(B.3.1)} \\ &\sup\{n^{-\frac{1}{2}}|M_{n_2}^{(j)}\{(a_j, b_j) + (t_{1_j}, t_{2_j})\} - M_{n_2}^{(j)}(a_j, b_j) + \lambda_j n\gamma \\ &\quad (t_{1_j}\bar{c}_j + t_{2_j}(C_j^{\star 2} + \bar{c}_j^2))| : |t_{1_j}| \le n^{-\frac{1}{2}}K, |t_{2_j}| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} 0. \qquad \text{(B.3.2)} \end{aligned}$$

(ii) Under $\boldsymbol{\theta} = \boldsymbol{0}, \boldsymbol{\beta} = \boldsymbol{0}$, as n grows large,

$$n^{-\frac{1}{2}}\begin{pmatrix} \boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ \boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0}) \end{pmatrix} \xrightarrow{d} N_{2p}\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix}\right], \qquad \text{(B.3.3)}$$

where $N_{2p}(\cdot\,,\cdot\,)$ represents a $2p$-variate normal distribution with appropriate parameters and $K \in \Re$.

Proof of part (iii) of Thereom 5.2.1: By equations (5.2.6), (B.3.1) and (B.3.2), under $H_0^\star : \boldsymbol{\beta} = \beta_0\boldsymbol{1}_p$, we obtain

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\boldsymbol{1}_p) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}, \beta_0\boldsymbol{1}_p) - n^{-\frac{1}{2}}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}, \beta_0\boldsymbol{1}_p) + \boldsymbol{o}_p(1). \qquad \text{(B.3.4)}$$

Further, the distribution of $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p)$ under $H_0^\star : \boldsymbol{\beta} = \beta_0\mathbf{1}_p$ is the same as the distribution of $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\mathbf{0},\mathbf{0}) - n^{-\frac{1}{2}}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0})$ under $H_0 : \boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\beta} = \mathbf{0}$ using equation (B.3.4) and the fact that the distribution of $M_{n_1}(a,b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ when $\theta = 0, \beta = 0$, and similar to $M_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332).

Therefore, utilizing (B.3.3), under $H_0^\star : \boldsymbol{\beta} = \beta_0\mathbf{1}_p$, as $n \to \infty$, we obtain the result in part (iii). The proof of parts (i) and (ii) of Theorem (5.2.1) is similarly obtained.

Proof of part (i) of Theorem 5.4.1: Under $H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0,\ \boldsymbol{\beta} = \beta_0\mathbf{1}_p$, with relation to (B.3.1) and (B.3.2),

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) - n^{\frac{1}{2}}\gamma\boldsymbol{\Lambda}_0(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(1) \quad \text{(B.3.5)}$$

and

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) - n^{\frac{1}{2}}\gamma\boldsymbol{\Lambda}_{12}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \boldsymbol{o}_p(1). \quad \text{(B.3.6)}$$

But, the equation (B.3.5) reduces to

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) = n^{\frac{1}{2}}\gamma\boldsymbol{\Lambda}_0(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(1) \quad \text{(B.3.7)}$$

by equation (5.2.6). Therefore, under $H_0^{(2)}$, we obtain

$$\begin{aligned}
&\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \\ n^{-\frac{1}{2}}M_{n_2}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) - n^{\frac{1}{2}}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12}M_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \end{bmatrix} \\
&= \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p) \end{bmatrix} - \begin{bmatrix} \boldsymbol{I}_p & \mathbf{0} \\ -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12} & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \\ n^{-\frac{1}{2}}M_{n_2}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p) \end{bmatrix} \\
&\xrightarrow{p} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \qquad \text{(B.3.8)}
\end{aligned}$$

Now utilizing the contiguity of probability measures (see Hájek et al., 1999, Ch.7) under $\{K_n^\star\}$ to those under $H_0^{(2)}$, the equation (B.3.8) implies that

$$[n^{-\frac{1}{2}}\boldsymbol{M}'_{n_1}(\boldsymbol{\theta}_0, \beta_0\mathbf{1}_p), n^{-\frac{1}{2}}\boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\mathbf{1}_p)]'$$

under $\{K_n^\star\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12} & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p) \\ n^{-\frac{1}{2}} M_{n_2}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p) \end{bmatrix}$$

under $H_0^{(2)}$. But the asymptotic distribution of the above random vector under $\{K_n^\star\}$ is the same as

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12} & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0 \mathbf{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \\ n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0 \mathbf{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \end{bmatrix}$$

under $H_0^{(2)}$ due to the fact that the distribution of $M_{n_1}(a, b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ under $\theta = 0, \beta = 0$, and similar to $M_{n_2}(0, 0)$ (c.f. Saleh, 2006, p.332).

Note that under $H_0^{(2)}$, with relation to (B.3.1) and (B.3.2),

$$\begin{bmatrix} n^{-\frac{1}{2}} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0 \mathbf{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \\ n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0 \mathbf{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \end{bmatrix} = \begin{bmatrix} n^{-\frac{1}{2}} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p) \\ n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p) \end{bmatrix} + \begin{bmatrix} \gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2) \\ \gamma(\boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{22}\boldsymbol{\delta}_2) \end{bmatrix} + \begin{bmatrix} \boldsymbol{o}_p(1) \\ \boldsymbol{o}_p(1) \end{bmatrix}.$$

Hence, by equation (5.2.8), under $H_0^{(2)}$,

$$\begin{bmatrix} n^{-\frac{1}{2}} \boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0 \mathbf{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \\ n^{-\frac{1}{2}} \boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0 \mathbf{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \end{bmatrix} \xrightarrow{d} N_{2p} \left[\begin{pmatrix} \gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2) \\ \gamma(\boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{22}\boldsymbol{\delta}_2) \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \right]. \quad \text{(B.3.9)}$$

Thus, as $n \to \infty$, the distribution of

$$[n^{-\frac{1}{2}} \boldsymbol{M}'_{n_1}(\boldsymbol{\theta}_0, \beta_0 \mathbf{1}_p), n^{-\frac{1}{2}} \boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0 \mathbf{1}_p)]'$$

under $\{K_n^\star\}$ is multivariate normal with mean vector

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12} & \boldsymbol{I}_p \end{bmatrix} \begin{pmatrix} \gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2) \\ \gamma(\boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{22}\boldsymbol{\delta}_2) \end{pmatrix} = \begin{bmatrix} \gamma(\boldsymbol{\Lambda}_0\boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\delta}_2) \\ \gamma(\boldsymbol{\Lambda}_{22} - \boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{21})\boldsymbol{\delta}_2 \end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12} & \boldsymbol{I}_p \end{bmatrix} \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda}_{12} & \boldsymbol{I}_p \end{bmatrix}' = \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_2^\star \end{pmatrix}.$$

Since the two statistics $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p)$ and $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\boldsymbol{1}_p)$ are uncorrelated, asymptotically, they are independently distributed normal vectors.

Proof of part (ii) of Theorem 5.4.1: Under $H_0^{(2)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0\ \boldsymbol{\beta} = \beta_0\boldsymbol{1}_p$, using equations (5.2.5), (B.3.1), (B.3.2) and (B.3.8), we find

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\boldsymbol{1}_p) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) - n^{\frac{1}{2}}\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) - n^{\frac{1}{2}}\boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) \end{bmatrix}$$

$$= \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\boldsymbol{1}_p) \end{bmatrix} + \begin{bmatrix} -\boldsymbol{I}_p & \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1} \\ \boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{11}^{-1} & -\boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) \\ n^{-\frac{1}{2}}M_{n_2}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) \end{bmatrix}$$

$$\xrightarrow{p} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}.$$

Now utilizing the contiguity of probability measures under $\{K_n^\star\}$ to those under $H_0^{(2)}$, the above equation implies that

$$[n^{-\frac{1}{2}}\boldsymbol{M}'_{n_1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\beta}}), n^{-\frac{1}{2}}\boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\theta}}, \beta_0\boldsymbol{1}_p)]'$$

under $\{K_n^\star\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} \boldsymbol{I}_p & -\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1} \\ -\boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{11}^{-1} & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) \\ n^{-\frac{1}{2}}M_{n_2}(\boldsymbol{\theta}_0, \beta_0\boldsymbol{1}_p) \end{bmatrix}$$

under $H_0^{(2)}$. But the asymptotic distribution of the above random vector under $\{K_n^\star\}$ is the same as

$$\begin{bmatrix} \boldsymbol{I}_p & -\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1} \\ -\boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{11}^{-1} & \boldsymbol{I}_p \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}M_{n_1}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0\boldsymbol{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \\ n^{-\frac{1}{2}}M_{n_2}(\boldsymbol{\theta}_0 - n^{-\frac{1}{2}}\boldsymbol{\delta}_1, \beta_0\boldsymbol{1}_p - n^{-\frac{1}{2}}\boldsymbol{\delta}_2) \end{bmatrix}$$

under $H_0^{(2)}$. Thus, the proof follows using equation (B.3.9). We find the two statistics are not independent, but rather correlated.

B.4 Multiple Linear Regression Model

The following asymptotic results of Jurečková (1977), Sen (1982) and Jurečková and Sen (1996, p.221) are used in deriving the distribution of the proposed tests. For simplicity, we assume S is known or consider the nonstudentized M-estimator, so we omit condition M1 of Jurečková and Sen (1996, p.217) and let $S_n = S$ in equation (5.5.29) of Jurečková and Sen (1996, p.221). Thus,

- Under $\boldsymbol{\beta}_1 = \boldsymbol{a}, \boldsymbol{\beta}_2 = \boldsymbol{b}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are r and s dimensional column vectors of any real numbers, as n grows large,

$$\sup\{n^{-\frac{1}{2}}|\boldsymbol{M}_{n_1}\{(\boldsymbol{a},\boldsymbol{b})+(\boldsymbol{t}_1,\boldsymbol{t}_2)\} - \boldsymbol{M}_{n_1}(\boldsymbol{a},\boldsymbol{b}) + n\gamma(\boldsymbol{Q}_{11}\boldsymbol{t}_1 + \boldsymbol{Q}_{12}\boldsymbol{t}_2)| : \\ |\boldsymbol{t}_1| \le n^{-\frac{1}{2}}K, |\boldsymbol{t}_2| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} \boldsymbol{0}, \qquad \text{(B.4.1)}$$

$$\sup\{n^{-\frac{1}{2}}|\boldsymbol{M}_{n_2}\{(\boldsymbol{a},\boldsymbol{b})+(\boldsymbol{t}_1,\boldsymbol{t}_2)\} - \boldsymbol{M}_{n_2}(\boldsymbol{a},\boldsymbol{b}) + n\gamma(\boldsymbol{Q}_{21}\boldsymbol{t}_1 + \boldsymbol{Q}_{22}\boldsymbol{t}_2)| : \\ |\boldsymbol{t}_1| \le n^{-\frac{1}{2}}K, |\boldsymbol{t}_2| \le n^{-\frac{1}{2}}K\} \xrightarrow{p} \boldsymbol{0}. \qquad \text{(B.4.2)}$$

- Under $\boldsymbol{\beta}_1 = \boldsymbol{0}, \boldsymbol{\beta}_2 = \boldsymbol{0}$, as n grows large,

$$n^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ \boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0}) \end{pmatrix} \xrightarrow{d} N_p\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix} \right), \qquad \text{(B.4.3)}$$

  where $N_p(\cdot\,,\cdot\,)$ represents a p-variate normal distribution with appropriate parameters and $K \in \Re$.

Proof of part (i) of Theorem 6.2.1: By equations (B.4.1) and (B.4.2), we find

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{\beta}_2) - n^{\frac{1}{2}}\gamma\boldsymbol{Q}_{12}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + \boldsymbol{o}_p(1) \qquad \text{(B.4.4)}$$

and

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{\beta}_2) - n^{\frac{1}{2}}\gamma\boldsymbol{Q}_{22}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + \boldsymbol{o}_p(1) \qquad \text{(B.4.5)}$$

under $H_0^{(1)}$. Then, we obtain

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{\beta}_2) - n^{-\frac{1}{2}}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{\beta}_2) + \boldsymbol{o}_p(1) \qquad \text{(B.4.6)}$$

by equations (6.2.5), (B.4.4) and (B.4.5) after some simple algebra.

Further, the distribution of $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2)$ under $H_0^{(1)}$ is the same as the distribution of $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) - n^{-\frac{1}{2}}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0})$ under $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}, \boldsymbol{\beta}_2 =$

$\mathbf{0}$ using equation (B.4.6) and the fact that the distribution of $M_{n_1}(a,b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ when $\theta = 0, \beta = 0$, and similar to $\boldsymbol{M}_{n_2}(0,0)$ (c.f. Saleh, 2006, p.332).

Therefore, utilizing equation (B.4.3), under $H_0^{(1)} : \boldsymbol{\beta}_1 = \mathbf{0}$ as $n \to \infty$, the proof of part (i) of Theorem 6.2.1 is completed.

The proof for part (ii) of Theorem 6.2.1 is obtained in the same way as in part (i).

Proof of part (i) of Theorem 6.4.1 Under $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$, $\boldsymbol{\beta}_2 = \mathbf{0}$, with relation to (B.4.1) and (B.4.2),

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\mathbf{0}) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\mathbf{0},\mathbf{0}) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ n^{\frac{1}{2}}\gamma\boldsymbol{Q}_{21}\tilde{\boldsymbol{\beta}}_1 \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \quad \text{(B.4.7)}$$

Note also that under H_0,

$$n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\tilde{\boldsymbol{\beta}}_1,\mathbf{0}) = n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) - n^{\frac{1}{2}}\gamma\boldsymbol{Q}_{11}\tilde{\boldsymbol{\beta}}_1 + \boldsymbol{o}_p(1) \quad \text{(B.4.8)}$$

and definition (6.2.6) reduce equation (B.4.8) to

$$n^{-\frac{1}{2}}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) = n^{\frac{1}{2}}\gamma\boldsymbol{Q}_{21}\tilde{\boldsymbol{\beta}}_1 + \boldsymbol{o}_p(1). \quad \text{(B.4.9)}$$

Therefore, under H_0, the equation (B.4.7) becomes

$$\begin{aligned} &\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\mathbf{0}) \end{bmatrix} - \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\mathbf{0},\mathbf{0}) - n^{\frac{1}{2}}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \end{bmatrix} \\ &= \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\mathbf{0}) \end{bmatrix} - \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\mathbf{0},\mathbf{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\mathbf{0},\mathbf{0}) \end{bmatrix} \\ &\xrightarrow{p} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \end{aligned} \quad \text{(B.4.10)}$$

Now utilizing the contiguity of probability measures (see Hájek et al., 1999, Ch.7) under $\{K_n\}$ to those under H_0, the equation (B.4.10) implies that

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}'_{n_1}(\mathbf{0},\mathbf{0}) & n^{-\frac{1}{2}}\boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\beta}}_1,\mathbf{0}) \end{bmatrix}'$$

under $\{K_n\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0}) \end{bmatrix}$$

under H_0. But the asymptotic distribution of the above random vector under $\{K_n\}$ is the same as

$$\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(-n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, -n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(-n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, -n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \end{bmatrix}$$

under H_0 due to the fact that the distribution of $M_{n_1}(a, b)$ under $\theta = a, \beta = b$ is the same as that of $M_{n_1}(\theta - a, \beta - b)$ under $\theta = 0, \beta = 0$, and similar to $\boldsymbol{M}_{n_2}(0, 0)$ (c.f. Saleh, 2006, p.332).

Note that under H_0, with relation to (B.4.1) and (B.4.2),

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \end{bmatrix} = \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0}) \end{bmatrix}$$
$$+ \begin{bmatrix} \gamma(\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{12}\boldsymbol{\lambda}_2) \\ \gamma(\boldsymbol{Q}_{21}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{22}\boldsymbol{\lambda}_2) \end{bmatrix} + \begin{bmatrix} \boldsymbol{o}_p \\ \boldsymbol{o}_p \end{bmatrix}.$$

Hence, by equation (B.4.3), under H_0,

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \end{bmatrix} \xrightarrow{d} N_p$$
$$\left(\begin{pmatrix} \gamma(\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{12}\boldsymbol{\lambda}_2) \\ \gamma(\boldsymbol{Q}_{21}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{22}\boldsymbol{\lambda}_2) \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix} \right). \quad \text{(B.4.11)}$$

Thus, as $n \to \infty$, the distribution of

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}'_{n_1}(\boldsymbol{0},\boldsymbol{0}) & n^{-\frac{1}{2}}\boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0}) \end{bmatrix}'$$

under $\{K_n\}$ is p-variate normal with mean vector

$$\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{pmatrix} \gamma(\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{12}\boldsymbol{\lambda}_2) \\ \gamma(\boldsymbol{Q}_{21}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{22}\boldsymbol{\lambda}_2) \end{pmatrix} = \begin{bmatrix} \gamma(\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{12}\boldsymbol{\lambda}_2) \\ \gamma(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\lambda}_2 \end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \sigma_0^2 \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix} \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix}' = \sigma_0^2 \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2^{\star} \end{pmatrix}.$$

Since the two statistics $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0})$ and $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0})$ are uncorrelated, asymptotically, they are independently distributed normal vectors.

Proof of part (ii) of Theorem 6.4.1: Under $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0},\ \boldsymbol{\beta}_2 = \boldsymbol{0}$, using equations (B.4.1), (B.4.2), (6.2.5) and (B.4.9),

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}'_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0}) \end{bmatrix} - \begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0}) \end{bmatrix}$$
$$\xrightarrow{p} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}. \tag{B.4.12}$$

Now utilizing the contiguity of probability measures under $\{K_n\}$ to those under H_0, the equation (B.4.12) implies that

$$\begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}'_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2) & n^{-\frac{1}{2}}\boldsymbol{M}'_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0}) \end{bmatrix}'$$

under $\{K_n\}$ is asymptotically equivalent to the random vector

$$\begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\boldsymbol{0}) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\boldsymbol{0},\boldsymbol{0}) \end{bmatrix}$$

under H_0. But the asymptotic distribution of the above random vector under $\{K_n\}$ is the same as

$$\begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1} \\ -\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{I}_s \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(-n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, -n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \\ n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(-n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, -n^{-\frac{1}{2}}\boldsymbol{\lambda}_2) \end{bmatrix}$$

under H_0. Then, equation (6.4.3) follows from equation (B.4.11) after some algebra. Since $n^{-\frac{1}{2}}\boldsymbol{M}_{n_1}(\boldsymbol{0},\tilde{\boldsymbol{\beta}}_2)$ and $n^{-\frac{1}{2}}\boldsymbol{M}_{n_2}(\tilde{\boldsymbol{\beta}}_1,\boldsymbol{0})$ are not independent, but rather correlated.

Appendix C

R codes

C.1 Simple Regression Model

Listing C.1: Code for the power functions of the tests for the simple regression model

```
#f2z is an R-code that list down the power functions of the
 +UT, RT and PTT. These power functions will be plotted
 +against lambda2.
f2z <-function(alpha3, lambda1, alpha1, alpha2, bar.c,rho,
 +gamma, sigma0.sq, C.star.sq) {
a <-rep(NA,20)
for(j in 0:20){
 a[j+1] <- j*0.5
 }
lambda2 <- a
m <-length(lambda2)
pi.st <- 0
pi.1 <- 0
pi.2 <- 0
for (i in 1:m){
     pi.one <- 1 - pnorm(qnorm(1-alpha1) - gamma*(lambda1 +
      +lambda2[i]*bar.c)/sqrt(sigma0.sq)) #RT
     pi.two <- 1 - pnorm(qnorm(1-alpha2) -gamma*lambda1*sqrt(C
      +.star.sq/((C.star.sq + bar.c^2)*sigma0.sq)))#UT
     p2<-pnorm(qnorm(1-alpha3)-gamma*lambda2[i]*sqrt(C.star.sq
      +/sigma0.sq))#1-PT
     p3 <- pmvnorm(mean = rep(0,2), sigma =matrix(c(1,rho,rho
      +,1),ncol=2), lower =  c(qnorm(1-alpha2)-gamma*lambda2[i
      +]*sqrt(C.star.sq/sigma0.sq), qnorm(1-alpha3)-gamma*
```

```
+lambda1*sqrt(C.star.sq/((C.star.sq + bar.c^2)*sigma0.sq
+))),upper = rep(Inf, 2))
    p1<-pi.one #RT
    pi.star <- p1*p2+p3
    pi.st <- append(pi.st, pi.star)
    pi.1 <- append(pi.1, pi.one)#RT
    pi.2 <- append(pi.2, pi.two)
    }
list(m=m,lambda2=lambda2, pi.st=pi.st, pi.1=pi.1, pi.2=pi.2)
}
#srs is an R-code that give the rank statistics. It is used
+to estimate gamma in the thesis.
srs<-function(tee,U)
sum(sign(U-tee)*rank(abs(U-tee))/(length(U)+1))
gen.pow.sim<-function(m,rn,lambda1,n,a){
intr<-rep(1,n)
alpha <- 0.05
pi.ptth1<-array(NA,c(rn,m)) #power fn Huber c=1.04 with
+contaminant
pi.uth1<-array(NA,c(rn,m))
pi.rth1<-array(NA,c(rn,m))
pi.ptth3<-array(NA,c(rn,m)) #power fn Huber c=1.64 with
+contaminant
pi.uth3<-array(NA,c(rn,m))
pi.rth3<-array(NA,c(rn,m))
pi.pttc1<-array(NA,c(rn,m)) #power fn LS with no contaminant
pi.utc1<-array(NA,c(rn,m))
pi.rtc1<-array(NA,c(rn,m))
pi.ptth2<-array(NA,c(rn,m)) #power fn Huber c=1.28 with
+contaminant
pi.uth2<-array(NA,c(rn,m))
pi.rth2<-array(NA,c(rn,m))
pi.pttc2<-array(NA,c(rn,m)) #power fn LS with contaminant
pi.utc2<-array(NA,c(rn,m))
pi.rtc2<-array(NA,c(rn,m))
for (i in 1:rn){
    error <-rnorm(n,0,1)# The 0% contamination data
    ci <- sample(a,n,replace=F)
    bar.c <- mean(ci) #bar.c is positive here
    C.star.sq <- (sum(ci^2)-n*(bar.c)^2)/n
```

```
rho <- - bar.c / sqrt(C.star.sq + bar.c^2)
Xi <- 2+ 3*ci + error #let theta=2 and beta=3.
lsq<-lsfit(ci,Xi) #LS
#mady<-mad(error) # median(abs(error - median(error)))
 +/0.6745
res.ls<- lsq$res   # = Xi-lsq$coef[1]-lsq$coef[2]*ci
mady<-mad(res.ls)
std.res.ls<-res.ls/mady
sigma0.sq <- sum(std.res.ls^2)/n  #that's the formula eq
 +.7
pal1 <- f2z(alpha, lambda1, alpha, alpha, bar.c,rho,1/
 +mady,sigma0.sq, C.star.sq)
pi.pttc1[i,]<-pal1$pi.st[2:(m+1)]
lambda2 <-pal1$lambda2[1:m]
pi.rtc1[i,] <-pal1$pi.1[2:(m+1)]
pi.utc1[i,]<-pal1$pi.2[2:(m+1)]
m <-pal1$m
# The 10% contamination data
error.con<-error
chose.obs <- sample(c(1:n), (10/100)*n) # chose
 +observations.
e.out1 <- c(runif(5, min=3.5, max=5),runif(5, min=-5, max
 +=-3.5))
e.out<-sample(e.out1,replace=F)
error.con[chose.obs] <- e.out
Xi.con<-2+3*ci+error.con #the contaminated response
lsq.con<-lsfit(ci,Xi.con) #LS
res.ls<- lsq.con$res   # (= Xi-lsq$coef[1]-lsq$coef[2]*ci)
mady<-mad(res.ls)
std.res.ls<-res.ls/mady
sigma0.sq <- sum(std.res.ls^2)/n  #that's the formula eq
 +.7
pal1 <- f2z(alpha, lambda1, alpha, alpha, bar.c,rho,1/
 +mady, sigma0.sq, C.star.sq)
pi.pttc2[i,]<-pal1$pi.st[2:(m+1)]
pi.rtc2[i,] <-pal1$pi.1[2:(m+1)]
pi.utc2[i,]<-pal1$pi.2[2:(m+1)]

tune.c<-1.04 # Huber c=1.04, qnorm(0.85)
```

```
hr.con<-rlm(matrix(c(intr,ci),ncol=2),Xi.con,k=tune.c) #
 +default tuning constant is 1.345.
mady<-mad(hr.con$res)
std.res.hr<-hr.con$res/mady
psi.res<-psi.huber(std.res.hr, k = tune.c, deriv = 0)*std
 +.res.hr
psiprime.res<-psi.huber(std.res.hr, k = tune.c, deriv =
 +1)# 0 or 1
sigma0.sq <- mean(psi.res^2) #sigma0.sq=sum psi(res/Sn)
ur<-uniroot(srs,c(-10,10),tol=0.0001,U=psiprime.res/mady)
 + #U=psi'/Sn
gamma<-ur$root
pal1 <- f2z(alpha, lambda1, alpha, alpha, bar.c,rho,gamma
 +, sigma0.sq, C.star.sq)
pi.ptth1[i,]<-pal1$pi.st[2:(m+1)]
pi.rth1[i,] <-pal1$pi.1[2:(m+1)]
pi.uth1[i,]<-pal1$pi.2[2:(m+1)]

tune.c<-1.28 #Huber c=1.28 qnorm(0.90)
hr.con<-rlm(matrix(c(intr,ci),ncol=2),Xi.con,k=tune.c)
mady<-mad(hr.con$res)
std.res.hr<-hr.con$res/mady
psi.res<-psi.huber(std.res.hr, k = tune.c, deriv = 0)*std
 +.res.hr
psiprime.res<-psi.huber(std.res.hr, k = tune.c, deriv =
 +1)
sigma0.sq <- mean(psi.res^2)
ur<-uniroot(srs,c(-10,10),tol=0.0001,U=psiprime.res/mady)
gamma<-ur$root
pal1 <- f2z(alpha,lambda1, alpha, alpha, bar.c,rho,gamma,
 + sigma0.sq, C.star.sq)
pi.ptth2[i,]<-pal1$pi.st[2:(m+1)]
pi.rth2[i,] <-pal1$pi.1[2:(m+1)]
pi.uth2[i,]<-pal1$pi.2[2:(m+1)]

tune.c<-1.64 # Huber c=1.64, qnorm(0.95)
hr.con<-rlm(matrix(c(intr,ci),ncol=2),Xi.con,k=tune.c)
mady<-mad(hr.con$res)
std.res.hr<-hr.con$res/mady
```

```
    psi.res<-psi.huber(std.res.hr, k = tune.c, deriv = 0)*std
     +.res.hr
    psiprime.res<-psi.huber(std.res.hr, k = tune.c, deriv =
     +1)
    sigma0.sq <- mean(psi.res^2)
    ur<-uniroot(srs,c(-10,10),tol=0.0001,U=psiprime.res/mady)
    gamma<-ur$root
    pal1 <- f2z(alpha,lambda1, alpha, alpha, bar.c,rho,gamma,
     + sigma0.sq, C.star.sq)
    pi.ptth3[i,]<-pal1$pi.st[2:(m+1)]
    pi.rth3[i,] <-pal1$pi.1[2:(m+1)]
    pi.uth3[i,]<-pal1$pi.2[2:(m+1)]
    }
pi.pttH1<-colMeans(pi.ptth1)
pi.rtH1 <-colMeans(pi.rth1)
pi.utH1<-colMeans(pi.uth1)
pi.pttC1<-colMeans(pi.pttc1)
pi.rtC1<-colMeans(pi.rtc1)
pi.utC1<-colMeans(pi.utc1)
pi.pttC2<-colMeans(pi.pttc2)
pi.rtC2<-colMeans(pi.rtc2)
pi.utC2<-colMeans(pi.utc2)
pi.pttH3<-colMeans(pi.ptth3)
pi.rtH3<-colMeans(pi.rth3)
pi.utH3<-colMeans(pi.uth3)
pi.pttH2<-colMeans(pi.ptth2)
pi.rtH2<-colMeans(pi.rth2)
pi.utH2<-colMeans(pi.uth2)
list(pi.pttH1=pi.pttH1,pi.rtH1=pi.rtH1,pi.utH1=pi.utH1,
pi.pttH2=pi.pttH2,pi.rtH2=pi.rtH2,pi.utH2=pi.utH2,
pi.pttH3=pi.pttH3,pi.rtH3=pi.rtH3,pi.utH3=pi.utH3,
pi.pttC1=pi.pttC1,pi.rtC1=pi.rtC1,pi.utC1=pi.utC1,
pi.pttC2=pi.pttC2,pi.rtC2=pi.rtC2,pi.utC2=pi.utC2, lambda2=
 +lambda2)
}
library(mvtnorm )
n<-100 #number of observation
m<-21
library(MASS)
lambda1 <- 2
```

```
rn<-3000 #number of simulation to run
a<-c(rep(-1,n/2),rep(0,n/2))
fit3<-gen.pow.sim(m,rn,lambda1,n,a)
#graphs for Fig 3.3 (a) and (b)
lambda2<-fit3$lambda2
plot(lambda2,fit3$pi.utC1[1:m], ylab="size of the UT",xlab =
 +" ", xlim = c(0, 10),ylim = c(0, 1), cex.lab = 1.4, pch=4,
 +col="red",type="n")
mtext(expression(lambda[2]), side=1, line = 3 , at = 5, cex
 +=1.5)
points(lambda2, fit3$pi.utH1[1:m], col=1,pch=22)
points(lambda2, fit3$pi.utH2[1:m], col=2,pch=4,lwd=2)
points(lambda2, fit3$pi.utH3[1:m], col=3,pch=15)
lines(lambda2, fit3$pi.utH1[1:m],lty=1, lwd=2, col=1)
lines(lambda2, fit3$pi.utH2[1:m],lty=1, lwd=2, col=2)
lines(lambda2, fit3$pi.utH3[1:m],lty=1, lwd=2, col=3)
lines(lambda2, fit3$pi.utC1[1:m],lty=1, lwd=2, col=4)
lines(lambda2, fit3$pi.utC2[1:m],lty=2, lwd=2, col=6)
legend(0,1.0001 , cex=1.10,c("MLE, uncontaminated",
"MLE, contaminated",
"Huber, contaminated,k=1.04 ",
"Huber, contaminated,k=1.28 ","Huber, contaminated,k=1.64
 + "),
col = c(4,6,1,2,3), text.col= "black",
lty = c(1,2,1,1,1), lwd=c(2,2,2),pch = c(-1,-1,22,4,15),
 +merge = TRUE, bg='white')
#red 2, blue 4, green 3, black 1
title(expression("(a) Size of the UT for "*lambda[1]*"=0, "*
 +bar(c)>0),cex.main=1.5)
#title(expression("(b) Power of the UT for "*lambda[1]*"=2,
 +"*bar(c)>0),cex.main=1.5)
#save in 3000rf1a.eps
```

C.2 Bivariate Noncentral Chi-square Distribution

Listing C.2: R Code for the pdf of the bivariate noncentral chi-square distribution

```
nbcsq.den<-function(rho,theta1,theta2,d1,d2,m){
```

```
# rho is correlation coefficient
# theta1 \& theta2 are noncentrality parameters
# d1 \& d2 are critical values
# m is degree of freedom
ml<-80 # take the sum up to 80 for j
mt1<-80 # take the sum up to 80 for lambda1 \& lambda2.
B<-array(NA,ml+1)
gl<-((1-rho^2)^((m)/2))*gamma((m/2)+c(0:ml))*(rho^(2*c(0:ml))
 +)/(factorial(c(0:ml))*gamma(m/2))
for(k in 0:ml){
     pCj<-(dgamma(d1/((1-rho^2)*2),m/2+c(0:mt1)+k)/(2*(1-rho
      +^2)))*((theta1/2)^c(0:mt1))*exp(-theta1/2)/factorial(c
      +(0:mt1))
     pDj<-(dgamma(d2/((1-rho^2)*2),m/2+c(0:mt1)+k)/(2*(1-rho
      +^2)))*((theta2/2)^c(0:mt1))*exp(-theta2/2)/factorial(c
      +(0:mt1))
     #Note, replace dgamma with pgamma gives the cdf
     A<-matrix(c(pCj*sqrt(gl[k+1])),ncol=1)
     D<-matrix(c(pDj*sqrt(gl[k+1])),ncol=1)
     B[k+1]<-sum(A%*%t(D))
     }
prob <-sum(B)
list(prob=prob)
}
```

Listing C.3: R Code for the plot of pdf of the bivariate noncentral chi-square distribution

```
gen.pdf <-function(theta1, theta2, rho,m,p){
# theta1 and theta2 are noncentrality parameters
# rho is correlation coefficient
# m is degree of freedom
# p is max value on the axis, e.g 100 will plot pdf for range
 + 0 to 100
dd1<-c(0:p) # 3D axis range
fp<-array(NA,c(p,p))
for(j in 1: p){
     d2<-dd1[j]
                  for(i in 1:p){
                  d1<-dd1[i]
                  f1<-nbcsq.den(rho, theta1, theta2,d1,d2,m)
```

```
                fp[i,j]<-f1$prob
                }
        }
    p.val<-c(fp)
    list(pdf.val=p.val)
    }
p1<-100 #range of x and y axes are from 0 to 100
dd1<-c(0:p)
d3<-array(NA, c(p,p))
for (i in 1:p){
    d3[i,]<-c(rep(dd1[i],p))
    d4[,i]<-c(rep(dd1[i],p))
    }
axis.y<-c(d3) #this gives values from 0:p-1 is repeated p
 +times.
axis.x<-c(d4)
#this gives p repeated 0, repeated 1, ..., p repeated p-1
f1<-gen.pdf(10,30,0.25,3,p1)
axis.z <-f1$pdf.val
library(lattice)
wireframe(axis.z~axis.y*axis.x,scales = list(arrows=FALSE),
 +xlab=expression(y[1]),ylab=expression(y[2]),zlab="pdf",
 +shade = TRUE,aspect = c(61/87, 0.4),light.source = c
 +(10,0,10),main=expression(theta[1]*paste("=10, ")*theta[2]*
 +paste("=30, ")* rho*paste(" =0.25, ")*m*paste(" =3")))
```

Listing C.4: R Code for cdf of the noncentral bivariate chi-square distribution

```
nbcsq.dist.c2<-function(rho, theta1, theta2,d1,d2,m,n){
ml<-80 #just let ml1 = ml2 = ml
mt1<-80
gl<-((1-rho^2)^((m)/2))*gamma((m/2)+c(0:ml))*(rho^(2*c(0:ml))
 +)/(factorial(c(0:ml))*gamma(m/2))
g2<-((1-rho^2)^((n)/2))*gamma((n/2)+c(0:ml))*(rho^(2*c(0:ml))
 +)/(factorial(c(0:ml))*gamma(n/2))
sAj<-rep(NA,(ml+1))
for(l in 0:ml){
    pCj<-(pgamma(d1/((1-rho^2)*2),m/2+c(0:mt1)+l))*((theta1
     +/2)^c(0:mt1))*exp(-theta1/2)/factorial(c(0:mt1))
    Aj<-matrix(c(pCj*gl[l+1]),ncol=1)
    sAj[l+1]<-sum(Aj)
```

```
    }
sumA<-sum(sAj)
sDk<-rep(NA,(ml+1))
for(l in 0:ml){
    pDk<-(pgamma(d2/((1-rho^2)*2),n/2+c(0:mt1)+l))*((theta2
     +/2)^c(0:mt1))*exp(-theta2/2)/factorial(c(0:mt1))
    Dk<-matrix(c(pDk*g2[l+1]),ncol=1)
    sDk[l+1]<-sum(Dk)
    }
sumD<-sum(sDk)
prob<-sumA*sumD
list(prob=prob)
}

nhaf<-function(rho, theta1,d1,m){
ml <-80
mt1<-80
B<-array(NA,ml+1)
gl<-((1-rho^2)^((m)/2))*gamma((m/2)+c(0:ml))*(rho^(2*c(0:ml))
 +)/(factorial(c(0:ml))*gamma(m/2))
for(l in 0:ml){
    pCj<-(pgamma(d1/((1-rho^2)*2),m/2+c(0:mt1)+l))*((theta1
     +/2)^c(0:mt1))*exp(-theta1/2)/factorial(c(0:mt1))
    A<-matrix(c(pCj*gl[l+1]),ncol=1)
    B[l+1]<-sum(A)
    }
prob <-sum(B)
list(prob=prob)
}
```

C.3 Multivariate Simple Regression Model

Listing C.5: Code for the power functions of the tests for multivariate simple regression model

```
pow.fn<-function(psi1,psi2,psi1.pr,psi2.pr,d1,la,a1,b1,rho){
# psi1, psi1.pr is score fn and its derivative error
# d1 is the critical value
# la is the number of
# a1 is (a,a) in Fig 1 of Chap 4
```

```
# b1 is (b,b)
# rho the correlation coef
p<-2 # Only for multivariate simple model with p=2
 delta11 <- (1/n)*sum(psi1*psi1)
 delta12 <- (1/n)*sum(psi1*psi2)
 delta21 <- (1/n)*sum(psi2*psi1)
 delta22 <- (1/n)*sum(psi2*psi2)
gamma1<-(1/n)*sum(psi1.pr)
gamma2<-(1/n)*sum(psi2.pr)
TTT<-matrix(c(delta11/(gamma1*gamma1),delta21/(gamma2*gamma1)
 +,delta12/(gamma1*gamma2),delta22/(gamma2*gamma2)),ncol=2,
 +byrow=T)
#So, all ncp use the same estimate of T which depends on the
 +distribution of data.
pow.ut<-array(NA,c(la))
pow.rt<-array(NA,c(la))
pow.ptt<-array(NA,c(la))
 for(i in 1:la){
     thetaUT<-t(matrix(rep(a1,2),ncol=1))%*%solve(TTT)%*%
      +matrix(rep(a1,2),ncol=1)*(C.star.sq/(C.star.sq+bar.c*
      +bar.c))
     thetaRT<-t(matrix(rep(a1,2)+rep(bar.c*b1[i],2),ncol=1))
      +%*%solve(TTT)%*%matrix(rep(a1,2)+rep(bar.c*b1[i],2),
      +ncol=1)
     thetaPT<-t(matrix(rep(b1[i],p),ncol=1))%*%solve(TTT)%*%
      +matrix(rep(b1[i],p),ncol=1)*C.star.sq
     pow.ut[i] <- 1 - pchisq(d1,p,thetaUT)
     pow.rt[i] <- 1 - pchisq(d1,p,thetaRT)
     p1 <- pchisq(d1,p,thetaPT)
     f.dis<-nbcsq.dist.c2(rho,c(thetaUT),c(thetaPT),d1,d1,p,p)
     acf <- nhaf(rho, c(thetaPT),d1,p)
     bcf <- nhaf(rho, c(thetaUT),d1,p)
     pow.ptt[i]<- p1*pow.rt[i]+1+f.dis$prob-acf$prob-bcf$prob
     }
 list(pi.ut=pow.ut, pi.rt=pow.rt,pi.ptt=pow.ptt)
 }

 gen.power.smvt<-function(a1,b1,sm,d1,rho,n,error1.rand,error2
 +.rand){
#a1 is (a,a), b1=(b,b)
```

```
#n is number of observations
#error1.rand is errors of size n for every simulation (sm)
la<-length(b1)
p<-2
pi.ut.ls<-array(NA,c(la,sm)) #power fn of UT with LS psi fn
pi.rt.ls<-array(NA,c(la,sm)) # power fn of RT with LS psi fn
pi.ptt.ls<-array(NA,c(la,sm)) # power fn of PTT with LS psi
 +fn
pi.ut.h<-array(NA,c(la,sm))  #..with Huber psi fn
pi.rt.h<-array(NA,c(la,sm))  #..with Huber psi fn
pi.ptt.h<-array(NA,c(la,sm))  #..with Huber psi fn
pi.ut.t<-array(NA,c(la,sm))  #..with Tukey psi fn
pi.rt.t<-array(NA,c(la,sm))  #..with Tukey psi fn
pi.ptt.t<-array(NA,c(la,sm))  #..with Tukey psi fn
for( j in 1:sm){
    error1<-error1.rand[,j]
    error2<-error2.rand[,j]
    psi1<-error1  #LS psi function
    psi2<-error2
    psi1.pr<-rep(1,n) #derivative of LS psi function
    psi2.pr<-rep(1,n)
    f1<-pow.fn(psi1,psi2,psi1.pr,psi2.pr,d1,la,a1,b1,rho)
    pi.ut.ls[,j]<-f1$pi.ut
    pi.rt.ls[,j]<-f1$pi.rt
    pi.ptt.ls[,j]<-f1$pi.ptt
    psi1<-psi.huber(error1,deriv=0)*error1  #Huber psi
     +function
    psi2<-psi.huber(error2,deriv=0)*error2
    psi1.pr<-psi.huber(error1, deriv = 1) #derivative of
     +Huber psi fn
    psi2.pr<-psi.huber(error2, k = 1.345, deriv = 1)
    f1<-pow.fn(psi1,psi2,psi1.pr,psi2.pr,d1,la,a1,b1,rho)
    pi.ut.h[,j]<-f1$pi.ut
    pi.rt.h[,j]<-f1$pi.rt
    pi.ptt.h[,j]<-f1$pi.ptt
    psi1<-psi.bisquare(error1, deriv = 0)*error1 #Tukey psi
     +fn
    psi2<-psi.bisquare(error2, deriv = 0)*error2
    psi1.pr<-psi.bisquare(error1, deriv = 1) #derivative of
     +Tukey psifn
```

```
    psi2.pr<-psi.bisquare(error2, deriv = 1)
    f1<-pow.fn(psi1,psi2,psi1.pr,psi2.pr,d1,la,a1,b1,rho)
    pi.ut.t[,j]<-f1$pi.ut
    pi.rt.t[,j]<-f1$pi.rt
    pi.ptt.t[,j]<-f1$pi.ptt
    }
pi.ut.tm<-rowMeans(pi.ut.t, na.rm = FALSE, dims = 1)
pi.rt.tm<-rowMeans(pi.rt.t, na.rm = FALSE, dims = 1)
pi.ptt.tm<-rowMeans(pi.ptt.t, na.rm = FALSE, dims = 1)
pi.ut.hm<-rowMeans(pi.ut.h, na.rm = FALSE, dims = 1)
pi.rt.hm<-rowMeans(pi.rt.h, na.rm = FALSE, dims = 1)
pi.ptt.hm<-rowMeans(pi.ptt.h, na.rm = FALSE, dims = 1)
pi.ut.lsm<-rowMeans(pi.ut.ls, na.rm = FALSE, dims = 1)
pi.rt.lsm<-rowMeans(pi.rt.ls, na.rm = FALSE, dims = 1)
pi.ptt.lsm<-rowMeans(pi.ptt.ls, na.rm = FALSE, dims = 1)
list(pi.ut.tm=pi.ut.tm, pi.rt.tm=pi.rt.tm,pi.ptt.tm=pi.ptt.tm
 +,pi.ut.hm=pi.ut.hm,pi.rt.hm=pi.rt.hm, pi.ptt.hm=pi.ptt.hm,
 +pi.ut.lsm=pi.ut.lsm, pi.rt.lsm=pi.rt.lsm,pi.ptt.lsm=pi.ptt.
 +lsm)
}

n<-60 #number of observations, i=1,..,n
p<-2 # dimensional of the multivariate simple is 2. j=1,2.
a <-c(rep(0,(n/2)),rep(1,(n/2))) # just to generate ci
ci <- sample(a,n,replace=F) #independent variables
bar.c <- mean(ci) #bar.c is positive here
C.star.sq <- (sum(ci^2)-n*(bar.c)^2)/n
rho <- - bar.c / sqrt(C.star.sq + bar.c^2) #correlation coef
alpha<-0.05
d1<-qchisq(1-alpha,p,0) #critical value
a1<-2 # a in Figure
b1<-c(0:28)*0.5 #b in Figure
sm<-100 # the number of simulation
library(MASS)
error1.rand<-array(NA,c(n,sm))
error2.rand<-array(NA,c(n,sm))
# creating Normal distribution error terms
for(j in 1:sm){
    error1.rand[,j]<-rnorm(n, 0,1)
    error2.rand[,j]<-rnorm(n, 0,1)
```

```
    }
f2<-gen.power.smvt(a1,b1,sm,d1,rho,n, error1.rand,error2.rand
 +)
pi.ut.tn<-f2$pi.ut.tm # Normal data, Tukey, UT
pi.rt.tn<-f2$pi.rt.tm
pi.ptt.tn<-f2$pi.ptt.tm
pi.ut.hn<-f2$pi.ut.hm
pi.rt.hn<-f2$pi.rt.hm
pi.ptt.hn<-f2$pi.ptt.hm
pi.ut.lsn<-f2$pi.ut.lsm
pi.rt.lsn<-f2$pi.rt.lsm
pi.ptt.lsn<-f2$pi.ptt.lsm
# creating 10\% wild error terms
error1.wild<-array(NA,c(n,sm))
error2.wild<-array(NA,c(n,sm))
    for( j in 1:sm){
    error1<-error1.rand[,j]
    chose.obs <- sample(c(1:n), (10/100)*n) #10\% will be
     +wild
    e.out1 <- 10*error1[chose.obs]
    e.out<-sample(e.out1,replace=T)
    error1[chose.obs] <- e.out1
    error2<-error2.rand[,j]
    chose.obs <- sample(c(1:n), (10/100)*n) #sample(c(1:n),
     +1)
    e.out1 <- 10*error2[chose.obs] #c(rnorm(1, 0, 100))
    e.out<-sample(e.out1,replace=T)
    error2[chose.obs] <- e.out1
    error1.wild[,j]<-error1
    error2.wild[,j]<-error2
    }
f2<-gen.power.smvt(a1,b1,sm,d1,rho,n, error1.wild,error2.wild
 +)
pi.ut.tt<-f2$pi.ut.tm  #Wild 10%, Tukey, UT
pi.rt.tt<-f2$pi.rt.tm
pi.ptt.tt<-f2$pi.ptt.tm
pi.ut.ht<-f2$pi.ut.hm
pi.rt.ht<-f2$pi.rt.hm
pi.ptt.ht<-f2$pi.ptt.hm
pi.ut.lst<-f2$pi.ut.lsm
```

```
pi.rt.lst<-f2$pi.rt.lsm
pi.ptt.lst<-f2$pi.ptt.lsm
error1.cau<-array(NA,c(n,sm))
error2.cau<-array(NA,c(n,sm))
for( j in 1:sm){
    error1.cau[,j]<-rcauchy(n, location = 0, scale = 1)
    error2.cau[,j]<-rcauchy(n, location = 0, scale = 1)
    }
f2<-gen.power.smvt(a1,b1,sm,d1,rho,n, error1.cau,error2.cau)
pi.ut.tc<-f2$pi.ut.tm  #Cauchy, Tukey, UT
pi.rt.tc<-f2$pi.rt.tm
pi.ptt.tc<-f2$pi.ptt.tm
pi.ut.hc<-f2$pi.ut.hm
pi.rt.hc<-f2$pi.rt.hm
pi.ptt.hc<-f2$pi.ptt.hm
pi.ut.lsc<-f2$pi.ut.lsm
pi.rt.lsc<-f2$pi.rt.lsm
pi.ptt.lsc<-f2$pi.ptt.lsm
plot(b1,pi.ptt.lsn, ylab="Power of the test ",xlab =
 +expression(b), xlim = c(0, 12),ylim = c(0,1), lwd=2,cex =
 +1, pch=4, col="red", cex.lab=1.3,mgp=c(2.5, 1, 0),type="n")
title(expression(paste("(c) Power of the test when a=2, ")*n*
 +paste("=60, ")*p*paste("=2")*paste(" with normal errors")),
 +cex.main=1.5)
lines(b1, pi.ut.lsn,lty=2, lwd=2, col=2)
points(b1, pi.ut.lsn,cex=1, lwd=2, col=2)
lines(b1, pi.ptt.lsn,lty=1, lwd=2, col=2)
lines(b1, pi.ut.hn,lty=3, lwd=2, col=1)
lines(b1, pi.ptt.hn,lty=2, lwd=2, col=1)
points(b1, pi.ptt.hn,pch=4, lwd=2, col=1)
legend("topright", ncol=2, cex=0.95, c("UT, LS", "PTT, LS",
 +"UT, Huber","PTT, Huber"),
col = c(2,2,1,1), text.col= "black",
lty = c(2,1,3,2), lwd=c(2,2,2,2),
pch = c(1, -1, -1,4), merge = TRUE, bg='white')
```

C.4 Parallelism Model

Listing C.6: Code for the power functions of the tests for parallelism model

```
f1.par<-function(thetaUT,thetaRT,thetaPT,alpha,p,rho,m){
pi.1<-array(NA,m)
pi.2<-array(NA,m)
pi.st<-array(NA,m)
for(i in 1:m){
     pi.1[i] <- 1 - pchisq(qchisq(1-alpha,p,0),p,thetaUT[i])
     pi.2[i] <- 1 - pchisq(qchisq(1-alpha,p,0),p,thetaRT[i])
     p1 <- pchisq(qchisq(1-alpha,p,0),p,thetaPT[i])
     p2<-pi.2[i]
     d1<-qchisq(1-alpha,p,0)
     f.dis<-nbcsq.dist.c2(rho,c(thetaUT[i]),c(thetaPT[i]),d1,
      +d1,p,p)
     acf <- nhaf(rho, c(thetaPT[i]),d1,1)
     bcf <- nhaf(rho, c(thetaUT[i]),d1,1)
     pi.st[i]<- p1*p2+1+f.dis$prob-acf$prob-bcf$prob
     }
list(pi.st=pi.st, pi.1=pi.1, pi.2=pi.2)
}
rn<-1
n1<-50
n2<-50
n<-100
alpha<- 0.05
m<-29
p<-1
rho<- -0.8164966
bar.c<-0.5
C.star.sq <-0.125
jack1 <-0
jack2 <-0
b<-c(0:28)*0.5
jill1 <-b
jill2 <-b
mat2.ut<-diag(c(((n1/n)*C.star.sq)/(C.star.sq + bar.c^2), ((
 +n2/n)*C.star.sq)/(C.star.sq + bar.c^2)),2,2)
mat2.rt<-diag(c((n1/n),(n2/n)),2,2)
mat2.pt<-diag(c( (n1/n)*C.star.sq,(n2/n)*C.star.sq ),2,2)
ad1<-(jack1*(n1/n)*C.star.sq)/(C.star.sq + bar.c^2)
ad2<-(jack2*(n2/n)*C.star.sq)/(C.star.sq + bar.c^2)
ggl1<-(n1/n)*jack1
```

```
gcl1<-(n1/n)*bar.c
ggl2<-(n2/n)*jack2
gcl2<-(n2/n)*bar.c
gsl1<-(n1/n)*C.star.sq
gsl2<-(n2/n)*C.star.sq
thetaUT<-array(NA,c(29,rn))
thetaRT<-array(NA,c(29,rn))
thetaPT<-array(NA,c(29,rn))
for (i in 1:29){
mat1.ut<-matrix(c(ad1,ad2),nrow=1,ncol=2,byrow=T)
    thetaUT[i,] <-mat1.ut%*%solve(mat2.ut)%*%t(mat1.ut)
    mat1.rt<-matrix(c(ggl1+gcl1*jill1[i],ggl2+gcl2*jill2[i]),
     +nrow=1,ncol=2,byrow=T)
    thetaRT[i,] <- mat1.rt%*%solve(mat2.rt)%*%t(mat1.rt)
    mat1.pt <-matrix(c( gsl1*jill1[i],gsl2*jill2[i]),nrow=1,
     +ncol=2,byrow=T)
    thetaPT[i,] <- mat1.pt%*%solve(mat2.pt)%*%t(mat1.pt)
    }
pi.ptt<-array(NA,c(rn,m))
pi.ut<-array(NA,c(rn,m))
pi.rt<-array(NA,c(rn,m))
for (i in 1:rn){
    pal1<-f1.par(thetaUT[,i],thetaRT[,i],thetaPT[,i],alpha,p,
     +rho,m)
    pi.ptt[i,]<-pal1$pi.st
    pi.ut[i,] <-pal1$pi.1
    pi.rt[i,]<-pal1$pi.2
    }
pi.pttc<-rbind(pi.ptt,pi.ptt)
pi.utc<-rbind(pi.ut,pi.ut)
pi.rtc<-rbind(pi.rt,pi.rt)
pi.Ptt<-rep(NA,29)
pi.Rt<-rep(NA,29)
pi.Ut<-rep(NA,29)
for(j in 1:29){
    pi.Ptt[j]<-mean(pi.ptt[,j])
    pi.Rt[j]<-mean(pi.rt[,j])
    pi.Ut[j]<-mean(pi.ut[,j])
    }
b<-jill2
```

```
plot(b,pi.Ptt, ylab="Size of the test ",xlab = "b", xlim = c
+(0, 13.5),ylim = c(0, 1), lwd=2,cex = 1, pch=4, col="red",
+cex.lab=1.8,mgp=c(2.5, 1, 0))
title(expression(paste("Size of the test when a=0, ")*alpha*
+paste("=0.05")),cex.main=2)
lines(b, pi.Ut,lty=2, lwd=2, col=1)
lines(b, pi.Rt,lty=1, lwd=2, col="blue")
lines(b, pi.Ptt,lty=1, lwd=2, col="red")
legend("topleft", cex=1.5, c(expression(UT), expression(RT),
+expression(PTT)), col = c(1,4,2), text.col= "black",
lty = c(2, 1, 1), lwd=c(2,2,2),pch = c(-1, -1, 4), merge =
+TRUE, bg='white')
```

C.5 Multiple Linear Regression Model

Listing C.7: Code of power functions of the tests for the multiple linear regression

```
mulp3d<-function(a,b,p.rho,sigma0.sq,gamm,Q11,Q12,Q21,Q22,Q1s
+,Q2s,Q12s,Q21s,alpha1,alpha2,alpha3,d1,d2){
la<-length(a)
pi.1 <-array(NA,la)
pi.2 <-array(NA,la)
pi.stn0<-array(NA,la)
pi.stn<-array(NA,la)
thetaUT<-array(NA,la)
thetaPT<-array(NA,la)
thetaRT<-array(NA,la)
Bb <-matrix(rep(b,2),ncol=1)
for(i in 1:la){
Aa <-matrix(rep(a[i],1),ncol=1)
#choose up to la so that ncp for UT <50 and choose a up to 10
+ only.
thetaUT[i] <- (gamm^2/sigma0.sq)*(t(Aa)%*%Q1s%*%Aa)
thetaRT[i] <- (gamm^2/sigma0.sq)*(t(Aa)%*%Q11%*%Aa + t(Aa)%*%
+Q12%*%Bb + t(Bb)%*%Q21%*%Aa+t(Bb)%*%Q21%*%solve(Q11)%*%Q12
+%*%Bb)
thetaPT[i] <- (gamm^2/sigma0.sq)*(t(Bb)%*%Q2s%*%Bb)
pi.1[i] <- 1 - pchisq(qchisq(1-alpha1,1,0),1,thetaUT[i])
pi.2[i] <- 1 - pchisq(qchisq(1-alpha2,1,0),1,thetaRT[i])
```

```
p1 <- pchisq(qchisq(1-alpha3,2,0),2,thetaPT[i])
p2 <- 1-pchisq(qchisq(1-alpha2,1,0),1,thetaRT[i])
f.dis<-nbcsq.dist.c2(p.rho,c(thetaUT[i]),c(thetaPT[i]),d1,d2
 +,1,2)
acf <- nhaf(p.rho, c(thetaPT[i]),d2,2)
bcf <- nhaf(p.rho, c(thetaUT[i]),d1,1)
pi.stn[i]<- p1*p2+1+f.dis$prob-acf$prob-bcf$prob
}
pi.1
pi.2
pi.stn
list(pi.1=pi.1,pi.2=pi.2, pi.stn=pi.stn, thetaRT=thetaRT,
 +thetaUT=thetaUT,thetaPT=thetaPT)
}
#Start Here
library(MASS)
n<-100
error <-rnorm(n, 0,1)
c1<-rep(1,n)
c2<-rnorm(n,0,2)#choose 1, power fn do not approach 1 as a
 +goes larger.
c3<-rnorm(n,0,2)
a<-c(rep(0,n/2),rep(1,n/2))
c2 <- sample(a,n,replace=F)
c3 <- sample(a,n,replace=F)
X <- 1+c2+c3+error
c1i <-matrix(c(c1),ncol=1)
c2i <-matrix(c(c2,c3),ncol=2,byrow=F) #an n by 2 dime
Q11<-(t(c1i)%*%c1i)/n
Q22<-(t(c2i)%*%c2i)/n
Q12<-(t(c1i)%*%c2i)/n
Q21<-(t(c2i)%*%c1i)/n
Q<-matrix(c(c(Q11),c(Q12),c(Q21[1,]),c(Q22[1,]),c(Q21[2,]),c(
 +Q22[2,])),byrow=T,ncol=3)
Q1s<-Q11-Q12%*%solve(Q22)%*%Q21
Q2s<-Q22-Q21%*%solve(Q11)%*%Q12
Q12s<-Q12%*%solve(Q22)%*%Q21%*%solve(Q11)%*%Q12-Q12
Q21s<-Q21%*%solve(Q11)%*%Q12%*%solve(Q22)%*%Q21-Q21
A<-matrix(c(c(Q1s),c(Q12s),c(Q21s[1,]),c(Q2s[1,]),c(Q21s[2,])
 +,c(Q2s[2,])),byrow=T,ncol=3)
```

```
rho<-array(NA, c(3,3))
for(i in 1:2){
        for(j in (i+1):3){
        rho[i,j]<-A[i,j]/(sqrt(A[i,i])*sqrt(A[j,j]))}}
p.rho<-sqrt(sum(rho[1,2]^2+rho[1,3]^2+rho[2,3]^2)/3)

b<-0 #let's choose b=0,1,2,4
a<-c(0:40)*0.25
alpha1<-0.05
alpha2<-0.05
alpha3<-0.05
d1<-qchisq(1-alpha1,1,0) #r=1
d2<-qchisq(1-alpha3,2,0) #s=2
ci<-cbind(c1,c2,c3)
error.con<-error
chose.obs <- sample(c(1:n), (10/100)*n) # I am goin to chose
 +observations to be changed.
#e.out1 <- c(rnorm((10/100)*(n/2), 0, 10), rnorm((10/100)*(n
 +/2), 0,10))
e.out1 <- c(runif(5, min=3.5, max=5),runif(5, min=-5, max
 +=-3.5))
#e.out1 <- c(runif((10/100)*(n/2), min=-2.5, max=-1.5),runif
 +((10/100)*(n/2), min=1.5, max=2.5))
#e.out1 <- c(runif((10/100)*(n/2), min=-15, max=-10),runif
 +((10/100)*(n/2), min=10, max=15))
e.out<-sample(e.out1,replace=F) #replce false so, every of e.
 +ou1 will be in e.out.
error.con[chose.obs] <- e.out
Xi.con<-c1+c2+c3+error.con #the contaminated response

tune.c<-1.28 #qnorm(0.9)
hr.con<-rlm(ci,Xi.con,k=tune.c) #the tuning constant that
 +they use is 1.345.
hr.con$coef
mady<-mad(hr.con$res)#how come this is forgotten..
res.hr<-(Xi.con-hr.con$coef[1]*c1-hr.con$coef[2]*c2-hr.
 +con$coef[3]*c3)/mady
res<-res.hr
zeros <- rep(0,n)
psiprime.res <- rep(1,n)
```

```
psi.res <- res
band.out <- abs(res) > tune.c
psi.res[band.out] <- (tune.c*sign(res))[band.out]
psiprime.res[band.out]<- zeros[band.out]
gamma <- mean(psiprime.res)
sigma0.sq <- mean(psi.res^2)
srs<-function(tee,U)
sum(sign(U-tee)*rank(abs(U-tee))/(length(U)+1))
ur<-uniroot(srs,c(-10,10),tol=0.0001,U=psiprime.res/mady) #
 +psiprime.res is either 0 or 1.
gamma<-ur$root
#the power functions
pm<-mulp3d(a,b,p.rho,sigma0.sq,gamma,Q11,Q12,Q21,Q22,Q1s,Q2s,
 +Q12s,Q21s,alpha1,alpha2,alpha3,d1,d2)
pi.ptthc2<-pm$pi.stn
pi.rthc2<-pm$pi.2
pi.uthc2<-pm$pi.1

#To plot Figure 6.1(a).
plot(a,pi.rthc2, ylab="power of the test",xlab = expression(
 +lambda[1]), xlim = c(0, 10),ylim = c(0, 1), cex.lab=1.4,col
 +=1, type="n")
title(expression("(a) Power of the test for "*lambda
 +[2]*"=[0,0]"^T),cex.main=1.7)
lines(a,pi.rthc2,lty=1, lwd=2, col=4)
lines(a,pi.uthc2,lty=2, lwd=2, col=1)
lines(a,pi.ptthc2,lty=1, lwd=2, col=2)
points(a,pi.ptthc2,pch=16,lwd=1,col=2)
legend("bottomright", cex=1.4, c(expression(Pi^UT),
 +expression(Pi^RT),expression(Pi^PTT)),, col = c(1,4,2),
 +text.col= "black",
            lty = c(2,1,1), lwd=c(2,2,2),pch = c(-1, -1, 16),
             +merge = TRUE, bg='white')
```

Listing C.8: Code of power functions of the tests - comparing chi-square and normal tests.

```
mulp1<-function(a,b,p.rho,sigma0.sq,gamma,Q11,Q12,Q21,Q22,Q1s
 +,Q2s,Q12s,Q21s,alpha1,alpha2,alpha3,d1,d2){
la<-length(a)
pi.1 <-array(NA,la)
```

```
pi.2 <-array(NA,la)
pi.bnc<-array(NA,la)
pi.stnr<-array(NA,la)
thetaUT<-array(NA,la)
thetaPT<-array(NA,la)
thetaRT<-array(NA,la)
p<-1
Bb <-matrix(rep(b,1),ncol=1)
for(i in 1:la){
    Aa <-matrix(rep(a[i],1),ncol=1)
    #choose up to la so that ncp for UT <50 and choose a up
     +to 10 only.
    thetaUT[i] <- (gamma^2/sigma0.sq)*(t(Aa)%*%Q1s%*%Aa)
    thetaRT[i] <- (gamma^2/sigma0.sq)*(t(Aa)%*%Q11%*%Aa + t(
     +Aa)%*%Q12%*%Bb + t(Bb)%*%Q21%*%Aa+t(Bb)%*%Q21%*%solve(
     +Q11)%*%Q12%*%Bb)
    thetaPT[i] <- (gamma^2/sigma0.sq)*(t(Bb)%*%Q2s%*%Bb)
    pi.1[i] <- 1 - pchisq(qchisq(1-alpha1,p,0),p,thetaUT[i])
    pi.2[i] <- 1 - pchisq(qchisq(1-alpha2,p,0),p,thetaRT[i])
    p1 <- pchisq(qchisq(1-alpha3,p,0),p,thetaPT[i])
    p2 <- 1-pchisq(qchisq(1-alpha2,p,0),p,thetaRT[i])
    f.dis<-nbcsq.dist.c2(p.rho,c(thetaUT[i]),c(thetaPT[i]),d1
     +,d2,p,p)
    acf <- nhaf(p.rho, c(thetaPT[i]),d2,p)
    bcf <- nhaf(p.rho, c(thetaUT[i]),d1,p)
    pi.bnc[i]<- p1*p2+1+f.dis$prob-acf$prob-bcf$prob #The
     +proposed bncs
    mu1<-sqrt(thetaUT[i])
    mu2<-sqrt(thetaPT[i])

    #steux roux approximation
    b1<-d1/(mu1^2+1)
    b2<-d2/(mu2^2+1)
    rho.stsq<-((p.rho + mu1*mu2)^2)/((1+mu1^2)*(1+mu2^2))
    f1<-nbcsq.dist.c2(sqrt(rho.stsq),0,0,b1,b2,p,p) #with no
     +ncp
    f2<-nhaf(sqrt(rho.stsq),0,b1,p)#with no ncp
    f3<-nhaf(sqrt(rho.stsq),0,b2,p)#with no ncp
    pi.stnr[i]<- p1*p2+1+f1$prob-f2$prob-f3$prob  #SteynRoux
     +Approximation
```

```
    }
list(pi.1=pi.1,pi.2=pi.2, pi.stnr=pi.stnr, pi.bnc=pi.bnc,
 +thetaRT=thetaRT,thetaUT=thetaUT,thetaPT=thetaPT)
}

library(MASS)
n<-100
a<-c(rep(0,n/2),rep(1,n/2))
error <-rnorm(n, 0,1)
c1<-rep(1,n)
c2 <- sample(a,n,replace=F)
X <- 2+3*c2+error
c1i <-matrix(c(c1),ncol=1)
c2i <-matrix(c(c2),ncol=1,byrow=F) #an n by 2 dime
Q11<-(t(c1i)%*%c1i)/n
Q22<-(t(c2i)%*%c2i)/n
Q12<-(t(c1i)%*%c2i)/n
Q21<-(t(c2i)%*%c1i)/n
Q<-matrix(c(Q11,Q12,Q21,Q22),byrow=T,ncol=2)
Q1s<-Q11-Q12%*%solve(Q22)%*%Q21
Q2s<-Q22-Q21%*%solve(Q11)%*%Q12
Q12s<-Q12%*%solve(Q22)%*%Q21%*%solve(Q11)%*%Q12-Q12
Q21s<-Q21%*%solve(Q11)%*%Q12%*%solve(Q22)%*%Q21-Q21
A<-matrix(c(Q1s,Q12s,Q21s,Q2s),byrow=T,ncol=2)
rho<-array(NA, c(2,2))
for(i in 1:1){
        for(j in (i+1):2){
        rho[i,j]<-A[i,j]/(sqrt(A[i,i])*sqrt(A[j,j]))
        }
    }
p.rho<-rho[1,2]
p<-1
b<-6 #let's choose a=0,1,2,4
a<-c(0:40)*0.25
alpha1<-0.05
alpha2<-0.05
alpha3<-0.05
d1<-qchisq(1-alpha1,p,0)
d2<-qchisq(1-alpha3,p,0)
ci<-cbind(c1,c2)
```

```
error.con<-error
chose.obs <- sample(c(1:n), (10/100)*n)
e.out1 <- c(runif(5, min=3.5, max=5),runif(5, min=-5, max
+=-3.5))
e.out<-sample(e.out1,replace=F)
error.con[chose.obs] <- e.out
Xi.con<-c1+c2+error.con #the contaminated response
tune.c<-1.28 # qnorm(0.9)
hr.con<-rlm(ci,Xi.con,k=tune.c) #default tuning constant is
+1.345.
mady<-mad(hr.con$res)
std.res.hr<-hr.con$res/mady
psi.res<-psi.huber(std.res.hr, k = tune.c, deriv = 0)*std.res
+.hr
psiprime.res<-psi.huber(std.res.hr, k = tune.c, deriv = 1)
sigma0.sq <- mean(psi.res^2)
srs<-function(tee,U)
sum(sign(U-tee)*rank(abs(U-tee))/(length(U)+1))
ur<-uniroot(srs,c(-10,10),tol=0.0001,U=psiprime.res/mady)
gamma<-ur$root

#Figure 6.3(e)
pm<-mulp1(a,b,p.rho,sigma0.sq,gamma,Q11,Q12,Q21,Q22,Q1s,Q2s,
+Q12s,Q21s,alpha1,alpha2,alpha3,d1,d2)
pi.ptthc2<-pm$pi.bnc
pi.rthc2<-pm$pi.2
pi.uthc2<-pm$pi.1

plot(a,pi.rthc2, ylab="power of the test",cex.lab=1.3, xlab =
+ expression(lambda[1]), xlim = c(0, 10),ylim = c(0, 1), col
+=1, type="n")
title(expression("(e) Power of the test for "*lambda[2]*"=6")
+,cex.main=1.5)
lines(a,pi.rthc2,lty=1, lwd=2, col=4)
lines(a,pi.uthc2,lty=2, lwd=2, col=1)
lines(a,pi.ptthc2,lty=1, lwd=2, col=2)
points(a,pi.ptthc2,pch=4,lwd=2,col=2,cex=0.6)
legend("bottomright", cex=1.3, c(expression(Pi^UT),
+expression(Pi^RT),expression(Pi^PTT)),, col = c(1,4,2),
+text.col= "black", lty = c(2,1,1), lwd=c(2,2,2,2,2),pch = c
```

```
+(-1, -1, 4), merge = TRUE, bg='white')
```

Appendix D

Publications

Journal Articles

(1) Yunus, R.M. and Khan, S. (2010). Increasing power of the test through pre-test - a robust method, *Communication in Statistics Theory and Methods*, to appear.

(2) Yunus, R M and Khan, S. (2010). Increasing power of robust test through pre-testing in multiple regression model, *Pakistan Journal of Statistics*, Special Volume on 25th Anniversary, ed by S.E. Ahmed, Vol. 26, No.1.

(3) Yunus, R.M. and Khan, S. (2010). M-test for intercept after pre-testing on slope, *Journal Statistical Modeling and Analytics*, to appear.

(4) Osland, E., Yunus, R., Khan, S. and Memon, M. A. (2010). Early versus traditional postoperative feeding in resectional gastrointestinal surgical patients: A meta analysis, *Journal of Parenteral and Enteral Nutrition*, to appear.

(5) Osland, E., Yunus, R M., Khan, S. and Memon, M A. (2009). Letter to the Editors: Early enteral nutrition within 24 h of intestinal surgery versus later commencement of feeding: A systematic review and meta-analysis, *Journal of gastrointestinal surgery*, **13**(6), 1163-1165.

(6) Matthew, P., Mukhtar, A., Yunus, R.M., Khan, S., Memon, B., Memon, M.A (2009). Meta analysis of randomized clinical trials comparing open and laparoscopic anti-reflux, *American Journal of Gastroenterology*, **104**: 1548 - 1561.

(7) Memon, M.A., Khan, S., Yunus, R.M., Barr, R., Memon, B. (2008). Meta-Analysis Of Laparoscopic And Open Distal Gastrectomy For Gastric Carcinoma, *Surg. Endosc.*, **22**(8):1781-9.

Conference Proceedings

(1) Khan, S. and Yunus, R.M. M-test of two parallel regression lines under uncertain prior information, In Proc: 10th Islamic Countries Conference on Statistical Sciences (ICCS-X), 20-23 Dec 2009, Cairo, Egypt.

(2) Yunus, R.M. and Khan, S. Increasing power of robust test through pre-testing in multivariate simple linear regression model. In Proc: The CD-ROM of the 3rd International Conference on Mathematics and Statistics (ICoMS-3), 5-6 Aug 2008, Bogor, Indonesia.

(3) Yunus, R.M. and Khan, S. Test for intercept after pre-testing on slope - a robust method. In Proc: 9th Islamic Countries Conference on Statistical Sciences (ICCS-IX): Statistics in the Contemporary World - Theories, Methods and Applications, 12-14 Dec 2007, Concord Hotel, Shah Alam, Malaysia.

Articles Submitted

(1) Yunus, R.M and Khan, S. Robust tests for multivariate simple model (submitted to Journal of Nonparametric Statistics).

(2) Yunus, R.M and Khan, S. The bivariate noncentral chi-square distribution - a compound distribution approach (submitted to Journal of Applied Mathematics & Computing).