Increasing power of the test through pre-test - a robust method

Rossita M Yunus* and Shahjahan Khan
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Q 4350, AUSTRALIA
Emails: *yunus@usq.edu.au and* khans@usq.edu.au

Abstract

Robust test procedures are developed for testing the intercept of a simple regression model when the slope is (i) completely unspecified, (ii) specified to a fixed value or (iii) suspected to be a fixed value. Defining (i) unrestricted (UT), (ii) restricted (RT) and (iii) pre-test test (PTT) functions for the intercept parameter under the three choices of the slope, tests are formulated using the M-estimation methodology. The asymptotic distributions of the test statistics and their asymptotic power functions are derived. The analytical and graphical comparisons of the tests reveal that the PTT achieves a reasonable dominance over the other tests.

Keywords: pre-test, asymptotic size, asymptotic power, M-estimation, contiguity, regression model.

1 Introduction

In recent years many researchers have contributed to the estimation of one parameter in the presence of uncertain prior information on the value of another parameter. In general, inclusion of non-sample prior information improves the quality of inference. In spite of plethora of work in the area of improved estimation using non-sample prior information (c.f. Saleh, 2006, p.2), very little attention has been paid on the testing of parameters in the presence of uncertain prior information. It may be a natural expectation that testing of one parameter after pre-testing on another would improve the performance of the ultimate test in the sense of higher power and lower size of the ultimate test.

---

*on leave from Institute of Mathematical Sciences, Faculty of Sciences, University of Malaya, Malaysia.

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

$$X_i = \theta + \beta c_i + e_i, \tag{1.1}$$

where the errors e_i's are identically and independently distributed from an unspecified symmetric at 0 and continuous distribution function, $F_i,\ i = 1, \ldots, n$, the c_i's are known real constants of the explanatory variable and θ and β are the unknown intercept and slope parameters respectively.

Gilchrist (1984) used the simple linear regression model to predict the road distance, X_i, by the linear distance, c_i, between any two destinations. In the regression equation, the intercept parameter, θ, represents the distance by road when the linear distance is zero. The slope parameter, β, represents the rate of change in the road distance for one unit change in the linear distance. In the absence of any knowledge on the value of the slope parameter, the researchers may face three different scenarios when the primary interest is to test the significance of the intercept parameter. The slope may be (i) completely unspecified, or (ii) zero or (iii) suspected at zero. In case (i), the commonly used test is applicable for testing θ; in case (ii), testing of θ has no bearings on β as $\beta = 0$; and in case (iii), the uncertainty in the suspected value of β is first removed by performing a pre-test (PT) on β, then use the PT outcome for testing θ. Let ϕ_n^{UT} be the test function for the unrestricted test (UT) on $H_0^\star : \theta = 0$ against $H_A^\star : \theta > 0$ when β is unspecified. For case (ii), let ϕ_n^{RT} be the test function for the restricted test (RT) on $H_0^\star : \theta = 0$ against $H_A^\star : \theta > 0$ when β is 0 (specified). Similarly, let ϕ_n^{PTT} be the test function for the pre-test test (PTT) on $H_0^\star : \theta = 0$ against $H_A^\star : \theta > 0$ following a pre-test on the slope. The test function for the pre-test (PT) on the slope to test $H_0^{(1)} : \beta = 0$ against $H_A^{(1)} : \beta > 0$ is denoted by ϕ_n^{PT}.

The properties of unrestricted estimator (UE), restricted estimator (RE) and pre-test estimator (PTE) have been investigated by many authors (Khan and Saleh, 1997 & 2001, Khan, Hoque and Saleh, 2002). Most of the studies are based on normal or t-models and the results are non-robust. In the studies, the PTE (a linear combination of UE and RE) possesses a small quadratic risk when the distance parameter are large and too close to zero, that makes it the best choice over the other two estimators. Instead of least squares (LS) and maximum likelihood (ML) estimators, the properties of UE, RE and PTE are also studied in the framework of general robust estimators, explicitly, M-estimators. As such, a robust estimator namely the preliminary test M-estimator (PTME) are proposed for linear models (Sen and Saleh, 1987). In this paper, three tests correspond to the UE, RE and PTE are defined. They are unrestricted test (UT), restricted test (RT) and pre-test test (PTT).

The properties of the pre-test as well as the power of the test followed by pre-test have been studied in parametric cases (Bechhofer, 1951, Bozivich, Bancroft and Hartley, 1956). The performance of the ultimate test after a pre-test is also investigated by Tamura (1965) but for one sample and two sample non-parametric problems. After almost two decades, the effect of pre-test (on slope) on the size and power of the ultimate test (on the intercept) were investigated for rank-based nonparametric tests by Saleh and Sen (1982). However, there are some limited discussions on the power of the PTT provided in the paper. To author's knowledge, no research has been done in the investigation of the performance (size and power) of the ultimate test following a pre-test in linear models that is formulated using the M-test, defined along the line of M-estimation methodology. It is expected that the M-test formulated in the M-estimation inherits the robustness properties of the estimation method, so the test is less sensitive to departures from model assumptions. Since the M-estimation method is more popular compared to the other robust methods, the study of the performance of the power function of the ultimate test derived using the M-estimation method is an important addition to the statistical literature.

This paper proposes the M-tests for the UT, RT and PTT. The M-test is originally proposed by Sen (1982) using the score function in the M-estimation methodology and it is introduced for testing the significance of the slope only. 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 this paper. Although the asymptotic results of Jurečková and Sen (1996) are used in deriving the distribution of the proposed tests, these results are used in a different model in the context of testing after pre-test. Along with some preliminary notions, the method of M-estimation is presented in Section 2. In Section 3, three statistical tests concerning testing on the intercept, namely, the UT, RT and PTT are proposed for the three different cases mention earlier. Further, the asymptotic distributions of the test statistics and the asymptotic power functions are derived in Section 4. Section 5 is devoted to the analytical results comparing the asymptotic power functions of the UT, RT and PTT while the investigation of the power functions through an illustrative example is presented in Section 6. An application to real data is provided in Section 7. The final Section presents discussions and concluding remarks.

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 of the centered and scaled of observation, X_i,

$$\sum_{i=1}^{n} \rho\left(\frac{X_i - \theta - \beta c_i}{S_n}\right). \tag{2.1}$$

M-estimator of θ and β can also be defined as the solutions of the system of equations,

$$\sum_{i=1}^{n} \frac{\partial \rho}{\partial \theta} = \sum_{i=1}^{n} \psi\left(\frac{X_i - \theta - \beta c_i}{S_n}\right) = 0, \quad \sum_{i=1}^{n} \frac{\partial \rho}{\partial \beta} = \sum_{i=1}^{n} c_i \psi\left(\frac{X_i - \theta - \beta c_i}{S_n}\right) = 0. \tag{2.2}$$

Here $\psi(\cdot)$ is known as the score function and 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). Several choices of $\psi(\cdot)$ function are given in literature. In this paper, we consider the Huber ψ-function,

$$\psi_H(U_i) = \begin{cases} U_i & |U_i| \le k, \\ k \text{ sign}(U_i) & |U_i| > k, \end{cases} \tag{2.3}$$

where $U_i = \frac{X_i - \theta - \beta c_i}{S_n}$ and k is known as the tuning constant because it can be chosen to fine tune the estimator so that it has a specified asymptotic efficiency for a chosen distribution, F. Note, the Huber ψ-function is a continuous, piecewise linear function and satisfies the properties of ψ_c given in equation (5.5.8) of Jurečková and Sen (1996, p.218). ψ-function Also, the ML estimates for θ and β are obtained using $\psi_{ML} = U_i$ for any $U_i \in \Re$.

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

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

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

$$\tilde{\beta} = [\underbrace{\sup\{b : M_{n_2}(0, b) > 0\}}_{b_1} + \underbrace{\inf\{b : M_{n_2}(0, b) < 0\}}_{b_2}]/2. \tag{2.4}$$

Any value $b_1 < b < b_2$ can serve as the estimate of $M_{n_2}(0, b)$. Note that $M_{n_2}(0, b)$ is decreasing if b is increasing (Jurečková and Sen, 1996, p.85).

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

$$\tilde{\theta} = [\sup\{a : M_{n_1}(a, 0) > 0\} + \inf\{a : M_{n_1}(a, 0) < 0\}]/2. \tag{2.5}$$

From Sen (1982), asymptotically,

$$n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta}, 0) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2}) \tag{2.6}$$

under $H_0^{(1)} : \beta = 0$, where

$$\sigma_0^2 = \int \psi^2 \left(\frac{X_i - \theta - \beta c_i}{S}\right) dF(X_i - \theta - \beta c_i) \tag{2.7}$$

is the second moment of $\psi(\cdot)$ while the first moment is zero due to the symmetrically distributed at 0 of error e_i. If $\psi(\cdot)$ is $\psi_{ML}(\cdot)$ and $F \sim N(0, \sigma^2)$, then $S = \sigma$ and $\sigma_0^2 = 1$. Also, $C^{\star 2} = \lim_{n\to\infty} n^{-1} C_n^{\star 2}$, $C_n^{\star 2} = \sum_{i=1}^{n} c_i^2 - n\bar{c}_n^2$, $\bar{c}_n = n^{-1} \sum_{i=1}^{n} c_i$ and $\lim_{n\to\infty} \bar{c}_n = \bar{c}$.

Let ${S_n^{(3)}}^2 = n^{-1}\sum_{i=1}^n \psi^2\left(\frac{X_i-\tilde{\theta}}{S_n}\right)$. The consistency of ${S_n^{(3)}}^2$ as an estimator of σ_0^2 follows from Jurečková and Sen (1981) (c.f. Sen, 1982). 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.

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

$$n^{-\frac{1}{2}}M_{n_1}(0,\tilde{\beta}) \xrightarrow{d} N(0, \sigma_0^2 C^{\star 2}/\{C^{\star 2} + \bar{c}^2\}) \tag{2.8}$$

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

3 The UT, RT, PT and PTT

In this Section, the UT, RT, PT and PTT are introduced using the notations and asymptotic distribution results from the previous Section.

3.1 The unrestricted test (UT)

If β is unspecified, the designated test function is ϕ_n^{UT} to test the null hypothesis $H_0^\star : \theta = 0$ against the alternative hypothesis $H_A^\star : \theta > 0$. 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 (2.4). It follows from equation (2.8) that under $H_0^\star$,

$$T_n^{UT}/\sqrt{C_n^{(1)}{S_n^{(1)}}^2} \xrightarrow{d} N(0,1) \tag{3.1}$$

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

$$P[T_n^{UT} > \ell_{n,\alpha_1}^{UT} | H_0^\star : \theta = 0] = \alpha_1, \tag{3.2}$$

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

$$\Phi(\tau_{\alpha_i}) = 1 - \alpha_i, \quad \text{for } 0 < \alpha_i < 1, \quad i = 1, 2, 3. \tag{3.3}$$

Using (3.1), (3.2) and (3.3), we observe that as $n \to \infty$,

$$n^{-\frac{1}{2}}\ell_{n,\alpha_1}^{UT}/\sqrt{{S_n^{(1)}}^2 C_n^{(1)}/n} \xrightarrow{p} \tau_{\alpha_1} = n^{-\frac{1}{2}}\ell_{n,\alpha_1}^{UT}/\sqrt{\sigma_0^2 C^{\star 2}/(C^{\star 2} + \bar{c}^2)} \text{ (say)}. \tag{3.4}$$

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.

3.2 The restricted test (RT)

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

$$n^{-\frac{1}{2}} T_n^{RT} / \sqrt{{S_n^{(2)}}^2} \xrightarrow{d} N(0, 1), \tag{3.5}$$

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

$$P[T_n^{RT} > \ell_{n,\alpha_2}^{RT} | H_0 : \theta = 0, \beta = 0] = \alpha_2, \tag{3.6}$$

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), (3.5) and (3.6), we obtain

$$n^{-\frac{1}{2}} \ell_{n,\alpha_2}^{RT} / \sqrt{{S_n^{(2)}}^2} \xrightarrow{p} \tau_{\alpha_2} = n^{-\frac{1}{2}} \ell_{n,\alpha_2}^{RT} / \sqrt{\sigma_0^2} \text{ (say)} \tag{3.7}$$

as $n \to \infty$. 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 The pre-test test (PTT)

In this section, test on slope is proposed first and followed by the construction of the ultimate test for testing the intercept.

The pre-test (PT)

For the pre-test on the slope, the test function, ϕ_n^{PT} is designed to test the null hypothesis $H_0^{(1)} : \beta = 0$ against $H_A^{(1)} : \beta > 0$. 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 (2.5)). Under $H_0^{(1)}$, if follows from equation (2.6) that

$$T_n^{PT} / \sqrt{C_n^{(3)} {S_n^{(3)}}^2} \xrightarrow{d} N(0, 1) \tag{3.8}$$

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

$$P[T_n^{PT} > \ell_{n,\alpha_3}^{PT} | H_0^{(1)} : \beta = 0] = \alpha_3. \tag{3.9}$$

Also by (3.3), (3.8) and (3.9), as $n \to \infty$,

$$n^{-\frac{1}{2}} \ell_{n,\alpha_3}^{PT} / \sqrt{{S_n^{(3)}}^2 C_n^{\star 2}/n} \xrightarrow{p} \tau_{\alpha_3} = n^{-\frac{1}{2}} \ell_{n,\alpha_3}^{PT} / \sqrt{\sigma_0^2 C^{\star 2}} \text{ (say)}, \tag{3.10}$$

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

The pre-test test (PTT)

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

$$\phi_n^{PTT} = I\left[(T_n^{PT} \le \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{UT})\right] \tag{3.11}$$

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

$$\begin{aligned}\Pi_n^{PTT}(\theta) &= E(\phi_n^{PTT}|\theta) \\ &= P[T_n^{PT} \le \ell_{n,\alpha_3}^{PT}, T_n^{RT} > \ell_{n,\alpha_2}^{RT}|\theta] + P[T_n^{PT} > \ell_{n,\alpha_3}^{PT}, T_n^{UT} > \ell_{n,\alpha_1}^{UT}|\theta]. \end{aligned} \tag{3.12}$$

In general, the power function of the PTT depends on $\alpha_1, \alpha_2, \alpha_3, \theta, n$ as well as β. Note that the size of the ultimate test α_n^{PTT} is a special case of the power of the test 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).

4 Asymptotic properties of UT, RT and PTT

In this section, the asymptotic joint distributions of $\left[T_n^{UT}, T_n^{PT}\right]$ and $\left[T_n^{RT}, T_n^{PT}\right]$ are derived under the local alternative K_n (defined below). Then, the asymptotic distribution for the UT, RT and PTT are used to obtain the power function of the UT, RT and PTT under K_n.

Theorem 4.1 *Let $\{K_n\}$ be a sequence of alternative hypotheses, where*

$$K_n : (\theta, \beta) = (n^{-\frac{1}{2}}\lambda_1, n^{-\frac{1}{2}}\lambda_2), \tag{4.1}$$

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

(i)

$$n^{-1/2}\begin{bmatrix} T_n^{RT} \\ T_n^{PT} \end{bmatrix} \sim N_2\left[\begin{pmatrix} \gamma(\lambda_1 + \lambda_2\bar{c}) \\ \gamma\lambda_2 C^{\star 2} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{pmatrix}\right], \tag{4.2}$$

(ii)

$$n^{-1/2}\begin{bmatrix} T_n^{UT} \\ T_n^{PT} \end{bmatrix} \sim N_2\left[\begin{pmatrix} \frac{\gamma\lambda_1 C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ \gamma\lambda_2 C^{\star 2} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} \frac{C^{\star 2}}{C^{\star 2}+\bar{c}^2} & -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} \\ -\frac{\bar{c}C^{\star 2}}{C^{\star 2}+\bar{c}^2} & C^{\star 2} \end{pmatrix}\right], \tag{4.3}$$

where $\gamma = \frac{1}{S}\int \psi'\left(\frac{X_i - \theta - \beta c_i}{S}\right) dF(X_i - \theta - \beta c_i)$ and ψ' is the derivative of ψ-function.

The proof of Theorem 4.1 is in the Appendix. form (1982)

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

$$d(q_1, q_2; \zeta) = \frac{1}{2\pi(1-\zeta^2)^{1/2}} \int_{q_1}^{\infty}\int_{q_2}^{\infty} \exp\left\{\frac{-(x^2 + y^2 - 2\zeta xy)}{2(1-\zeta^2)}\right\} dxdy, \tag{4.4}$$

where q_1, q_2 are real numbers and $-1 < \zeta < 1$ is the correlation coefficient.

The asymptotic power functions for the UT and RT under $\{K_n\}$ are respectively

$$\Pi^{UT}(\lambda_1, \lambda_2) = 1 - \Phi(\tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}\ /\sigma_0) \text{ and} \quad (4.5)$$

$$\Pi^{RT}(\lambda_1, \lambda_2) = 1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0) \quad (4.6)$$

using equations (3.3), (3.4), (3.7), (4.2) and (4.3). Note that $\alpha_1 = \Pi^{UT}(0, \lambda_2) = \alpha^{UT}$ while $\alpha_2 = \Pi^{RT}(0, 0)$ and $\alpha^{RT} = \Pi^{RT}(0, \lambda_2)$.

It follows from equations (3.3), (3.4), (3.7), (3.10), (4.2), (4.3) and (4.4) that the asymptotic power function for the PTT, under $\{K_n\}$, is

$$\Pi^{PTT}(\lambda_1, \lambda_2) = \Phi(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0)[1 - \Phi(\tau_{\alpha_2} - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0)] + d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_{\alpha_1} - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}/\sigma_0; -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}\ ). \quad (4.7)$$

5 Asymptotic comparison

This section gives analytic asymptotic comparison of the power functions of the UT, RT and PTT.

- Case I: When $\bar{c} = 0$, $\alpha_1 = \alpha_2 = \alpha$,

  $$\Pi^{UT}(\lambda_1, \lambda_2) = \Pi^{RT}(\lambda_1, \lambda_2) = \Pi^{PTT}(\lambda_1, \lambda_2) = 1 - \Phi(\tau_\alpha - \gamma\lambda_1/\sigma_0) \quad (5.1)$$

  using equations (4.5), (4.6) and (4.7), i.e. the power functions for the UT, RT and PTT are the same.

- Case II: When $\bar{c} > 0$, $\alpha_1 = \alpha_2 = \alpha$, we find

  Result (i): $\Pi^{RT}(\lambda_1, \lambda_2) > \Pi^{PTT}(\lambda_1, \lambda_2)$ from equations (4.5) and (4.7),

  Result (ii): $\Pi^{RT}(\lambda_1, \lambda_2) > \Pi^{UT}(\lambda_1, \lambda_2)$ from equations (4.5) and (4.6) and

  Result (iii): $\Pi^{UT}(\lambda_1, \lambda_2) \lesseqgtr \Pi^{PTT}(\lambda_1, \lambda_2)$ if $B \lesseqgtr |A|$ where $A = \Phi(\tau_\alpha - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0) - \Phi(\tau_\alpha - \gamma\lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}\ /\sigma_0)$ and $B = d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0, \tau_\alpha - \gamma(\lambda_1 + \lambda_2\bar{c})/\sigma_0; 0)$ $-d(\tau_{\alpha_3} - \gamma\lambda_2 C^\star/\sigma_0,\ \tau_\alpha - \gamma\lambda_1\ \sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}\ /\sigma_0; -\bar{c}/\sqrt{C^{\star 2} + \bar{c}^2}\ )$ from equation (4.7).

- Case III: When $\bar{c} < 0$, $\alpha_1 = \alpha_2 = \alpha$ and $\lambda_1 + \lambda_2\bar{c} < \lambda_1\sqrt{C^{\star 2}/(C^{\star 2} + \bar{c}^2)}$, we find

  Result (iv): $\Pi^{RT}(\lambda_1, \lambda_2) < \Pi^{PTT}(\lambda_1, \lambda_2)$ from equations (4.5) and (4.7), and

  Result (v): $\Pi^{RT}(\lambda_1, \lambda_2) < \Pi^{UT}(\lambda_1, \lambda_2)$ from equations (4.5) and (4.6).

The analytical results in this section is accompanied with an illustrative example in investigating the comparison of the power of the tests discussed in the next section. The power of the tests at any value of the intercept other than $\theta = 0$ is also considered in the example to study the behavior of the power functions corresponding to the probabilities of Type I and Type II errors.

6 Power Comparison - Simulated data

The asymptotic power functions for the UT, RT and PTT are compared in this section and are supported by the analytical results given in Section 5 for the three cases I, II and III.

The Monte Carlo method is used for this simulated example. The random errors, e_i's of the simple linear regression model are generated from the normal distribution with mean 0 and variance 1. Then, set $\theta = 2$ and $\beta = 3$. The sample size is $n = 100$. Three sets of values: 0 and 1 with 50% for each for the first set, -1 and 1 with 50% for each for the second set and -1 and 0 with 50% for each for the third set are considered as the values of the regressor $c_i,\ i = 1, 2, \ldots, 100$. These values guarantee $\bar{c} > 0$, $\bar{c} = 0$ and $\bar{c} < 0$ respectively to the sets of regressors. Under normality assumption i.e. $e_i \sim N(0, \sigma^2)$ with $\sigma^2 = 1$, we prefer $\psi(U_i)$ to be $\psi_{ML}(U_i) = U_i$ together with $\psi'_{ML}(U_i) = 1$ for any U_i and $S_n = MAD/0.6745$. Then, the power functions of the UT, RT and PTT for this ML ψ-function are obtained using equations (4.5), (4.6) and (4.7) and are plotted as solid lines in Figure 1.

In practice, often the normality assumption is not met due to the presence of contaminants in the collected data. In this example, to create contamination observations, we randomly choose to replace $m(< n)$ of the n responses with some additive contamination, such that the contaminated responses X'_i is $X'_i = \theta + \beta c_i + \delta_i$ with δ_i is 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 (4.5), (4.6) and (4.7) using $\psi_{ML}(\cdot)$ and $\psi_H(\cdot)$ functions (given in equation (2.3)) with $S_n = MAD/0.6745$. Three values of tuning constant for the Huber ψ-function are selected, that are, $k = 1.04, 1.28$ 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 observations will have a value between $-k$ and k (Wilcox, 2005, p.76) while $k = 1.04$ (and 1.64) means there is 0.7 (and 0.90) probability that a random sample observations will have the value in the range of $(-1.04, 1.04)$ (and $(-1.64, 1.64)$). When Huber $\psi_H(\cdot)$ is used, the estimate for σ_0^2 is taken to be $\sum \psi_H(U_i)^2/n$. 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 $a_n(k) = k/(n+1), k = 1, \ldots, n$. Here, $V_i = \psi'_H(U_i)/S_n$ where $\psi'_H(U_i)$ is just the derivative of the Huber ψ-function.

The simulated values of the response and independent variables for the contaminated normal cases are used to obtain the M-estimates of the intercept and slope. The estimated coefficients for each set of data are used to calculate the power functions through the residual, $X_i - \hat{\theta} - \hat{\beta}c_i$ where $\hat{\theta}$ and $\hat{\beta}$ are the M-estimate of the intercept and slope. The simulation is run 3000 times. The average values of the power functions for 3000 data sets are plotted in Figures 1 and 2. The R-package (mvtnorm) is used to evaluate the bivariate normal probability integral for the power function of the PTT.

The power functions for the UT, RT and PTT are plotted against λ_2 at two values of λ_1 in Figures 1 and 2. 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 Type I error is small. Since we also expect to get small value of probability of Type II error, the power of the test at $\lambda_1 = 2$ is considered. An acceptable power function of the test is the one has smaller values when the null hypothesis is true and larger values when λ_1 differs much from $\theta = 0$.

The role of tuning constant of Huber ψ-function as a key parameter that control the efficiency and robustness of the procedure is studied (see Figure 1). Figure 1 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 graphs. Under the normality assumption, we know that the MLE of θ and β are unbiased estimators. The power function of the test that is based on the ML estimates using ψ_{ML} inherits the same good property. The test becomes the most powerful test when the normality assumption is met. However, this normality assumption may not be satisfied in practical situations. Studies show that the ML estimator is non-robust when there is departures from the model assumption or when outlier or contaminant occurs in the data. Figure 1 shows that the power curves obtained using the ML ψ-function for the contaminated data is far from those of the uncontaminated data. The large distance between two curves suggests that the MLE is not robust when there is contamination in the data. On the other hand, there is a tuning constant that fine tune the robustness of the Huber ψ-function based procedure. The power curves obtained using the Huber ψ-function with appropriate selection of tuning constant is closer to the power curves obtained using 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 all plots in Figure 1). This small distance between two curves means even 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 Huber ψ-function with $k = 1.28$ represents the majority of the data and the procedure is robust against some departure from the model assumption.

In Figure 2, the comparison on the performance (size and power) of the UT, RT and PTT are studied for three cases of regressor. All power curves in Figure 2 are obtained using the

Huber ψ-function with tuning constant $k = 1.28$. The first set of regressors is used to plot Figures 2(a) and 2(b). As λ_2 grows larger, size of the RT ($\Pi^{RT}(0, \lambda_2)$) approaches 1. However, size of the PTT ($\Pi^{PTT}(0, \lambda_2)$), 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 ϕ_n^{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 lesser than that of $\Pi^{RT}(0, \lambda_2)$. The size of the UT ($\Pi^{UT}(0, \lambda_2)$) is constant and does not depend on λ_2. The same pattern occurs in Figure 2(b) 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 RT, though the UT remains as the best choice. However, the RT is the best choice but the PTT is preferred to UT if the power of the test at $\lambda_1 = 2$ is considered. Setting $\bar{c} = 0$ in Figures 2(c) and 2(d) imply all power functions remain the same regardless of the value of λ_2 and these constant power functions increase as λ_1 increases. Figures 2(e) and 2(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 λ_2 more than a small positive value, say λ_0. The probability of Type I error for all test functions are fairly small. The size and power of the RT is decreasing to 0 as λ_2 growing larger (Figures 2(e) and 2(f)) suggesting 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.

7 Application to real 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 1. 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 paper, 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 2 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 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

Table 1: 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	0.000	0.834	3.009	0.0040

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.

Table 2: 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	$X_i = \beta c_i + e_i$	$X_i = e_i$	$X_i = \theta + e_i$
Coef	$\tilde{\beta} = 1.289$	None	$\tilde{\theta} = 20.855$
Test 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

The sensitivity of the 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 (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. 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.

8 Concluding Remarks

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

Table 3: Summary statistics of a one sided test for the UT, RT and PTT using the ML and Huber ψ-functions on the modified data.

		UT $\begin{pmatrix} H_0^\star : \theta = 0 \\ X_i' = \beta c_i + e_i \end{pmatrix}$	RT $\begin{pmatrix} H_0^\star : \theta = 0 \\ X_i' = e_i \end{pmatrix}$	PT $\begin{pmatrix} H_0^{(1)} : \beta = 0 \\ X_i' = \theta + e_i \end{pmatrix}$
ML	Coef	$\tilde{\beta} = 1.321$	None	$\tilde{\theta} = 22.855$
	Test 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
Huber	Coef	$\tilde{\beta} = 1.289$	None	$\tilde{\theta} = 22.208$
	Test 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}

of the distance parameter. The UE has a constant MSE that does not depend on the distance parameter. The PTE has smaller MSE than that of the RE for moderate and larger values of the distance parameter. The PTE has smaller MSE than the UE if the value of 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 mean value larger than 0, the size of the RT is small when $\lambda_2 = \sqrt{n}\beta = 0$ or close to 0, but the size grows large and converges to 1 for larger values of λ_2. The UT has a constant size regardless of the value of λ_2. The PTT has 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 smaller size than the UT for $\lambda_2 = 0$ or very close to 0.

Again for a set of realistic values of the regressor, with mean larger than 0, the RT is the best choice for having largest power but the worst choice for having largest size. The size of the UT is constant regardless of the value of the slope (via λ_2). The UT is the best choice for having smallest size but the worst choice for having smallest power. The PTT has smaller size than the RT for moderate and larger values of the slope and has larger power than the UT for smaller and moderate values of the slope. Therefore, the power function of the PTT is found to behave similar to the MSE of the PTE in the sense that though it is not uniformly the best statistical test with the smallest size and the largest power but it protects from the risk of a too large size and a too small power. 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.

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 the slope is small or moderate. Even for large values of the slope the PTT has at least as much power as the UT.

Acknowledgements
The authors thankfully acknowledge valuable suggestions of Professor A K Md E Saleh, Carleton University, Canada that helped improve the content and quality of the results in the paper. Some useful suggestions by an unknown referee are also acknowledged.

A Appendix

Interested readers are referred to Jurečková (1977), Sen (1982) and Jurečková and Sen (1996, p.221) for the following asymptotic properties. For simplicity and assumption that S_n is an unbiased estimator of S, let $S_n = S$ in equation (5.5.29) of Jurečková and Sen (1996, p.221). Under $H_0 : \theta = 0, \beta = 0$, as n grows large,

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

$$\sup\{n^{-\frac{1}{2}} |M_{n_1}(a,b) - M_{n_1}(0,0) + n\gamma(a + b\bar{c})| : |a|, |b| \le n^{-\frac{1}{2}} K\} \xrightarrow{p} 0 \text{ and} \tag{A.2}$$

$$\sup\{n^{-\frac{1}{2}} |M_{n_2}(a,b) - M_{n_2}(0,0) + n\gamma\{a\bar{c} + b(C^{\star 2} + \bar{c}^2)\}| : |a|, |b| \le n^{-\frac{1}{2}} K\} \xrightarrow{p} 0, \tag{A.3}$$

where K is a positive constant and $N_2(\cdot\ ,\cdot\ )$ represents a bivariate normal distribution with appropriate parameters. The above convergence is in probability, means the sequences of random variables converges in probability to a fix value (0).

Proof of part (i) of Theorem 4.1: Under $H_0 : \theta = 0,\ \beta = 0$, with relation to (A.2) and (A.3),

$$n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta}, 0) = n^{-\frac{1}{2}} M_{n_2}(0,0) - n^{\frac{1}{2}} \gamma \tilde{\theta} \bar{c} + o_p(1) \text{ and} \tag{A.4}$$

$$n^{-\frac{1}{2}} M_{n_1}(\tilde{\theta}, 0) = n^{-\frac{1}{2}} M_{n_1}(0,0) - n^{\frac{1}{2}} \gamma \tilde{\theta} + o_p(1). \tag{A.5}$$

Recalling definition (2.5), the equation (A.5) reduces to

$$n^{-\frac{1}{2}} M_{n_1}(0,0) = n^{\frac{1}{2}} \gamma \tilde{\theta} + o_p(1), \tag{A.6}$$

and hence equation (A.4) becomes

$$n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta}, 0) = n^{-\frac{1}{2}} M_{n_2}(0,0) - n^{-\frac{1}{2}} M_{n_1}(0,0)\bar{c} + o_p(1). \tag{A.7}$$

Therefore, under H_0, we find

$$\begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(0,0) \\ n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta},0) \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(0,0) \\ n^{-\frac{1}{2}} M_{n_2}(0,0) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.8}$$

Now utilizing the contiguity of probability measures (see Hájek et al., 1999, Chapter 7) under $\{K_n\}$ to those under H_0, the equation (A.8) implies that $[n^{-\frac{1}{2}} M_{n_1}(0,0), n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta},0)]'$ 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 by 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 similarly for $M_{n_2}(0,0)$ (c.f. Saleh, 2006 p.332).

Note that under $H_0 : \theta = 0, \beta = 0$, it follows from equations (A.1), (A.2) and (A.3), $n^{-\frac{1}{2}}[M_{n_1}(-n^{-\frac{1}{2}}\lambda_1, -n^{-\frac{1}{2}}\lambda_2), M_{n_2}(-n^{-\frac{1}{2}}\lambda_1, -n^{-\frac{1}{2}}\lambda_2)]'$

$$\to N_2\left(\begin{pmatrix} \gamma(\lambda_1 + \lambda_2\bar{c}) \\ \gamma\{\lambda_1\bar{c} + \lambda_2(C^{\star 2} + \bar{c}^2)\} \end{pmatrix}, \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix}\right). \tag{A.9}$$

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

$$\begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix} \sigma_0^2 \begin{pmatrix} 1 & \bar{c} \\ \bar{c} & C^{\star 2} + \bar{c}^2 \end{pmatrix} \begin{bmatrix} 1 & 0 \\ -\bar{c} & 1 \end{bmatrix}' = \sigma_0^2 \begin{bmatrix} 1 & 0 \\ 0 & C^{\star 2} \end{bmatrix}. \tag{A.10}$$

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 4.1: Under $H_0 : \theta = 0, \beta = 0$, using equations (2.4), (A.2), (A.3) and (A.7), as $n \to \infty$,

$$\begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(0,\tilde{\beta}) \\ n^{-\frac{1}{2}} M_{n_2}(\tilde{\theta},0) \end{bmatrix} - \begin{bmatrix} 1 & -\bar{c}/(C^{\star 2} + \bar{c}^2) \\ -\bar{c} & 1 \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} M_{n_1}(0,0) \\ n^{-\frac{1}{2}} M_{n_2}(0,0) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.11}$$

Now by using the contiguity of probability measures under $\{K_n\}$ to those under H_0, the equation (A.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

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

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

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

under H_0. Then, equation (4.3) follows from equation (A.9) after some algebra. Clearly, the two test statistics $n^{-\frac{1}{2}}T_n^{UT}$ and $n^{-\frac{1}{2}}T_n^{PT}$ are not independent, rather correlated.

References

Abraham, B. and Ledolter, J. (2006). *Introduction to Regression Modeling.* US: Thomson Brooks/Cole.

Bechhofer, R.E. (1951). The effect of preliminary test of significance on the size and power of certain tests of univariate linear hypotheses, Ph.D. Thesis (unpublished), Columbia Univ.

Bozivich, H., Bancroft, T.A. and Hartley, H. (1956). Power of analysis of variance test procedures for certain incompletely specified models. Ann. Math. Statist. 27:1017-1043.

Gilchrist, W. (1984). *Statistical Modelling.* Chichester, UK: Wiley.

Hájek, J., Šidák, Z., and Sen, P.K. (1999). *Theory of Rank Tests.* New York: Academia Press.

Wiley. Jurečková, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. Ann. Statist. 5:464–72.

Jurečková, J. and Sen, P.K. (1981). Sequential procedures based on M-estimators with discontinuous score functions. J. Statist. Plan. Infer. 5:253-66.

Jurečková, J. and Sen, P.K. (1996). *Robust Statistical Procedures Asymptotics and Interrelations.* New York: Wiley.

Khan, S. and Saleh, A.K.Md.E. (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal, 39, 131-147.

Khan, S., and Saleh, A.K.Md.E. (2001). On the comparison of pre-test and shrinkage estimators for the univariate normal mean. Stat. papers. 42: 451-473.

Khan, S., Hoque, Z., and Saleh, A.K.Md.E. (2002). Estimation of the slope parameter for linear regression model with uncertain prior information. J. Stat. Res. 36: 55-74.

Montgomery, D.C., Peck, E.A. and Vining G.G. (2001). *Introduction to linear regression analysis* (3rd ED.). New York: Wiley.

Saleh, A.K.Md.E. and Sen, P.K. (1982). Non-parametric tests for location after a preliminary test on regression. Communications in Statistics: Theory and Method 11:639-651.

Saleh, A.K.Md.E. (2006). *Theory of Preliminary test and Stein-type estimation with applications.* New Jersey: Wiley.

Sen. P.K. (1982). On M-tests in linear models. Biometrika 69:245-248.

Sen, P.K. and Saleh, A.K.Md.E. (1987). On preliminary test and schrinkage M-estimation in linear models. Ann. Statist. 15:1580-1592.

Tamura, R. (1965). Nonparametric inferences with a preliminary test. Bull. Math. Stat. 11:38-61.

Wilcox, R.R. (2005). *Introduction to Robust Estimation and Hypothesis Testing* (2nd ED.). US: Academia Press.

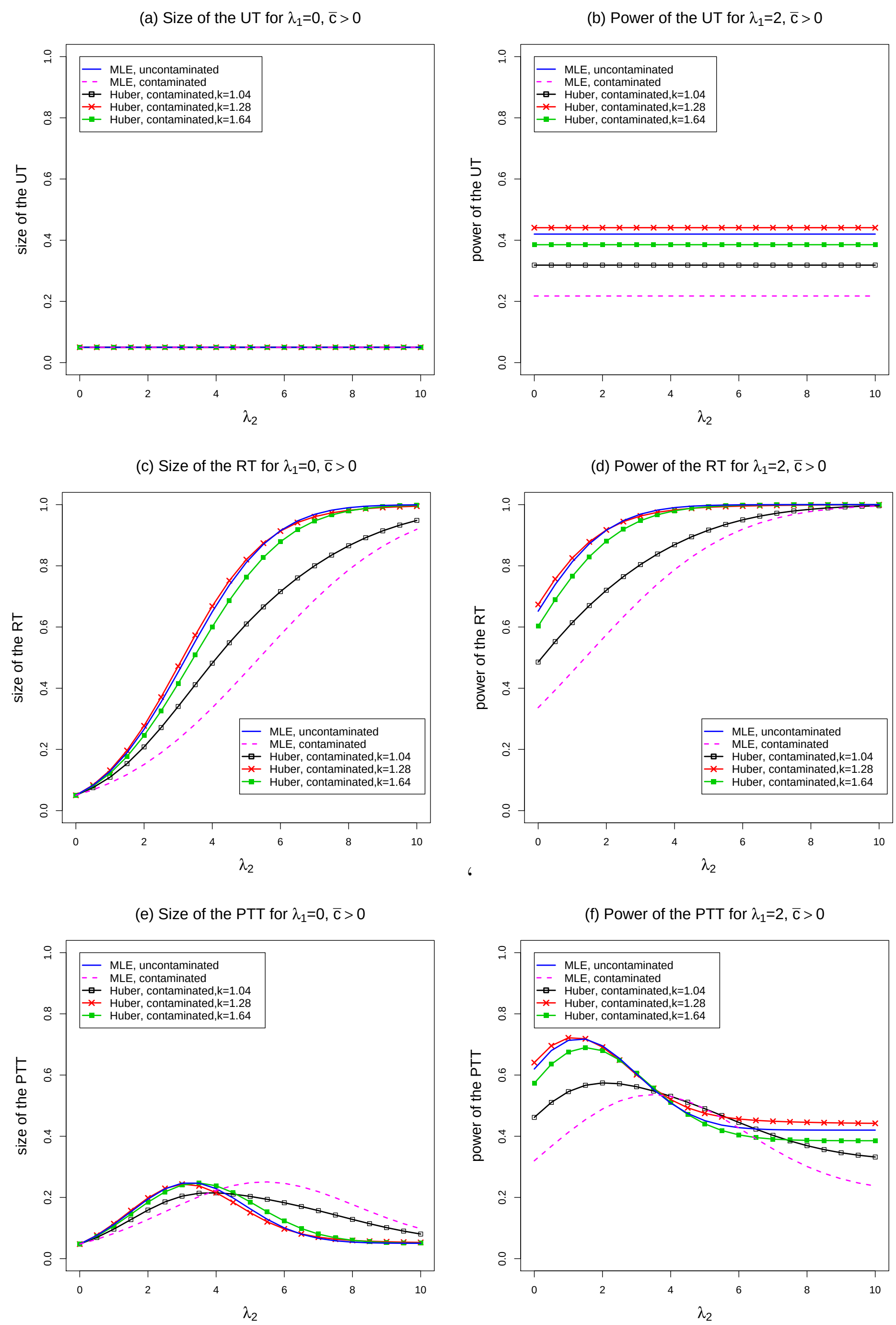

Figure 1: Graphs of power functions as a function of λ_2 for selected values of λ_1, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$ and $n = 100$.

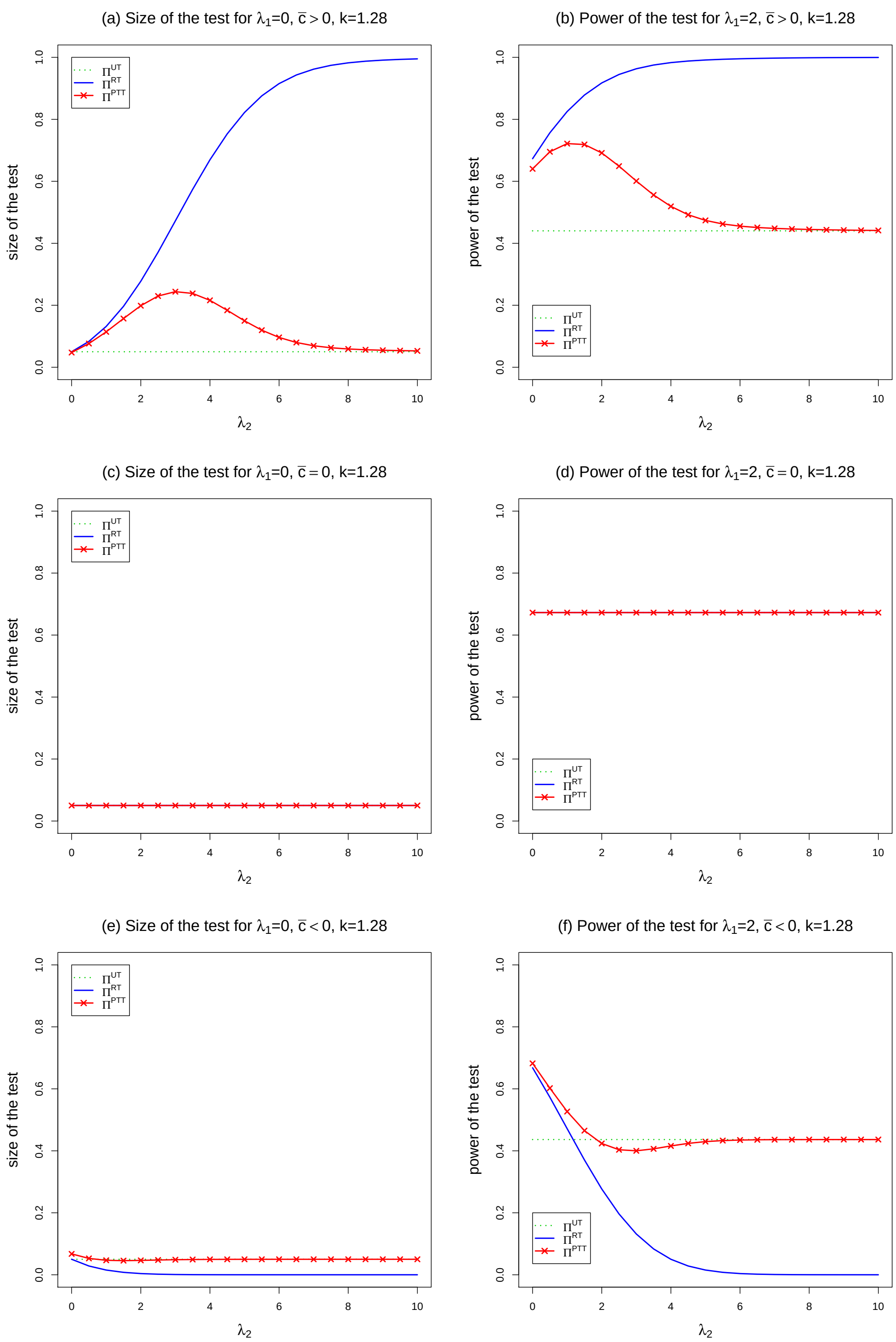

Figure 2: Graphs of power functions as a function of λ_2 for selected values of λ_1 and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05$.