Jahangirnagar University Journal of Statistical Studies
ISSN 1022-4734

http://www.juniv.edu/department/stat

JUJSS, Vol. 36, 2022, pp. 01-22

The Robust Wald Test for Testing a Subset of Regression Parameters of a Multiple Regression Model with Apriori Information on Another Subset

Shahjahan Khan*
School of Mathematics, Physics and Computing, School of Science and Techonology, Asian University of Bangladesh, Dhak 1341, Bangladesh and University of Southern Queensland, Toowoomba, Qld 4350, Australia.

Rossita M Yunus
Institute of Mathematical Sciences, Faculty of Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Abstract
The classical and M-estimator-based robust Wald tests are introduced to simultaneously test an arbitrary subset of coefficients of a multiple regression model when the remaining coefficients are either (i) unspecified, (ii) specified with certainty or (iii) suspected with uncertainty. Under the three scenarios the classical and robust Wald test statistics for (i) unrestricted (UT), (ii) restricted (RT) and (iii) pre-test (PTT) tests are defined. The aims of the paper are to (i) define the classical and robust Wald UT, RT and PTT statistics, (ii) find the asymptotic distribution of the test statistics (iii) determine the power function of the tests and (iv) compare the performance of the robust Wald UT, RT and PTT to their classical counterparts for large data. A Monte Carlo simulation study is conducted to obtainand compare the empirical power of the tests. The simulation study shows a domination of the PTT over the UT and RT when the suspected values are close to the true values and the robust Wald test is better than the its classical counterpart in terms of size and power under a slight departure from normality assumption. An example with Olympic athlete data is provided for illustration of the proposed method.

Keywords: M-estimator, robust test, nonparametric test, Wald test, regression model, asymptotic power, Monte Carlosimulation and contiguity.
Mathematics Subject Classification: 62E20, 62G35, 62F03, 62J05

1. Introduction

In many disciplines, results from previous studies or knowledge of experts in the field may pro-vide valuable prior information on the value of the underlying parameters of a multiple linear regression model. In general, inclusion of any trustworthy prior information in the estimation of parameter and test of hypotheses may improve the quality of statistical inference. Although the prior information usually comes from trusted sources, there is always an element of uncer-tainty in such information. The idea of the removal of the uncertainty in the non-sample prior information (NSPI) through a preliminary testing (pre-test) has drawn an increasing attention in the statistics literature since Bancroft (1944). However, almost all the initial studies in this area were focusing on improving estimation of parameters rather than hypothesis test. Ahmed and Saleh (1989), Akritas et al. (1984), Khan and Saleh (2001), Khan (2000,

* Corresponding author: shahjahan.Khan@usq.edu.au

2003, 2005), and Saleh (2006), to name a few, are among contributors to the study on developing and improving estimators of various kind of statistical models through preliminary testing. Despite the plethora of studies into the effects of pre-test on estimation, there are only few studies that have involved the

In statistics, it is a general interest to increase the power of any test, and it is more so for the testing after pre-test problem. Statistician shave studied the effect of pre-testing on the final test for several models including the analysis of variance (Paull 1950, Bechhofer 1951, Bozivich et al.1956, Mead et al.1975, among others), one sample and two sample problems (Tamura 1965), simple regression, multiple regression, multivariate regression and parallelism models (Saleh and Sen 1982,1983, Lambert et al. 1985, Yunus and Khan 2010, 2011a, b). In the literature, the statistical tests that were used in these models are based on the rank tests (Tamura 1965, Saleh and Sen 1982, 1983) and the robust score tests (Yunus and Khan 2010, 2011 a, b) for large sample size and the t-test (Khan and Pratikno 2013) for small sample size. The feature that distinguishes this paper from the previous works in the context of multiple regression model is the introduction of classical and robust Wald test in the preliminary testing frame work. Furthermore, it defines and investigates the size and power of classical and robust Wald tests in the context of testing an arbitrary subset of regression parameters where prior information on the other subset is available. It also provides an illustrative example of the proposed test using Olympic athlete data.
Consider the following multiple regression model

$$Y_n = X_n \beta + e_n, \tag{1.1}$$

where $Y_n = (Y_1, Y_2, ..., Y_n)'$ is a vector of n realizations of an observable response variable, $\mathbf{X}_n$ is a known design matrix of order $n \times p, \beta = (\beta_1, \beta_2, ..., \beta_p)'$ is a p-dimensional row vector of unknown regression parameters, and $\mathbf{e}_n = (e_1, e_2, \ldots, e_n)'$ is a vector of n independent and identically distributed errors with a distribution function F.

To formulate the testing of an arbitrary subset of the regression parameter vector, let's partition the p-dimensional parameter vector β' as $\beta' = (\beta_1', \beta_2')$ with $\beta_1' = (\beta_1, ..., \beta_r)$ and $\beta_2' = (\beta_{r+1}, ..., \beta_p)$ two r and t dimensional row vectors such that $r + t = p$. Then, partition X_n' as (X_{n1}', X_{n2}'), where $\mathbf{X}_{n1}$ is a known design matrix of order $n \times r$ and $\mathbf{X}_{n2}$ is another known design matrix of order $n \times t$.

Consider testing the parameters vector $\boldsymbol{\beta}_1$ specified at $\boldsymbol{\beta}_{01}$ when there is uncertain NSPI on $\boldsymbol{\beta}_2$, namely values (i) unspecified (ii) specified and (iii) suspected but not sure. For case (i), we want to test $H_0^*: \beta_1 = \beta_{01}$ against $H_A^*: \beta_1 \neq \beta_{01}$ with $\boldsymbol{\beta}_2$ is treated as a nuisance parameter.

This test is called the unrestricted test (UT). For case (ii), the test for testing $H_A^*: \beta_1 = \beta_{01}$ against $H_A^*: \beta_1 \neq \beta_{01}$ when $\beta_2 = \beta_{02}$, is called the restricted text (RT). For case (iii), testing $H_0^{(1)}: \beta_2 = \beta_{02}$, is recommended to remove the uncertainty of the suspicious values of $\beta_2 = \beta_{02}$ before testing on β_1. The test on $H_0^{(1)}: \beta_2 = \beta_{02}$ against $H_A^{(1)}: \beta_2 \neq \beta_{02}$ is known as a pre-test (PT). If the null hypothesis is this pre-test rejected, the UT is appropriate to test H_0^*, otherwise the RT is used to test H_0^*. The final test for testing for testing H_0^*, following a pre-test on $H_0^{(1)}$, is termed as the pre-test test (PTT). The objective of this study is to determine which of the classical and robust Wald UT, RT and PTT is better in terms of the test power criterion.

The next section briefly reviews the classical and robust Wald tests. Section 3 derives the asymptotic distribution of robust M-estimator sunder the null hypothesis. The robust Wald test statistic for the PT to test $H_0^{(1)}$, and the UT, RT and PTT to finally test H_0^* are introduced in Section 4. In the same section, we also provide the classical counterpart for robust Wald test. In Section 5, the Monte Carlo simulation for the comparison of the power of the tests are performed and the results are presented graphically. An illustrative example on the application of the method is provided in Section 6 with Olympic athlete data. Some concluding remarks are included in Section 7. Interested readers may refer to Appendix B and C for the asymptotic distribution under a sequence of local alternatives required for the derivation of the asymptotic power functions of the proposed robust Wald UT, RT and PTT. R-codes for the simulation study is available upon request.

2. Classical and Robust Wald Tests

Wald test is one of the classical tests that is widely used in statistics and econometrics. Originally proposed by Wald (1943), the test has been used to test parameters of linear models by many authors including Engle (1984). When F represents a multivariate normal distribution with mean vector **0** and covariance matrix $\sigma_0^2 \boldsymbol{I}_n$, $\sigma_0^2 > 0$, the test statistic for the classical Wald test for testing the simple null hypothesis $H_0 = \beta = \beta_0$ against $H_A: \beta \neq \beta_0$ is defined as

$$CW_n = \left(\breve{\beta} - \beta_0\right)' \sum_n^{-1} \left(\breve{\beta} - \beta_0\right), \tag{2.2}$$

where $\breve{\beta} = \left(X_n' X_n\right)^{-1} X_n' Y_n$ is the maximum likelihood estimator of β, and $\hat{\sum_n} = \frac{t}{n} \hat{\sigma}_0^2 \left(X_n' X_n\right)^{-1}$ where $\hat{\sigma}_0^2 = \sum \left(Y_i - X_i'\beta\right)^{2/n}$, in which Y_i is the response variable ion individual i and vector X_i is the ith row of the design matrix X_n, for i = 1, 2, ...,n.

The reason for preferring the Wald test over other alternative tests is its simple formulation. The Wald test is defined using the estimated coefficient and the variance of the estimator. The Wald test is easier to implement than its competitors such as the score test (Rao 1948), as it does not require the computation of the score function and the inversion of the information matrix (Carolan and Rayner 2000). The Wald and score tests for simple null hypothesis are asymptotically equivalent for large sample size (cf. Atkinson and Lawrance 1989, Rayner and Best1989, p.42). The performance of the tests can vary significantly for small samples. Un-like the likelihood ratio test (Neyman and Pearson 1928), the Wald test does not require the computation of the maximum likelihood function of the parameters both under the null and alternative hypotheses (Sen et al. 2010, pp.77). The likelihood ratio test has widely been used to test for the significance of a subset of parameters of a multiple regression model. According to Engle (1984, pp.792), the Wald (specifically the Wald UT) and likelihood ratio tests are asymptotically equivalent but the Wald test is computationally easier.

Most real-world data sets contain outliers, and thus do not follow the commonly assumed normal distribution. The classical Wald test defined in equation (2.2), however, is highly sen-sitive to model mis-specification and presence of outlying observations. Several versions of robust Wald test for linear model appear in Carroll and Ruppert (1988, pp. 214), Heritier and Ronchetti (1994), Jurečkováand Sen (1996,pp.419) and Silvapulle (1992). Recently, Basuet al. (2018) used minimum density power divergence estimator to define Wald test for multiple linear regression model. Using the same estimator Basuetal. (2017) defined Wald type test for the logistic regression model.

In this paper, we use M-estimator to define robust Wald test. This method is used to define alternative Wald tests, namely the UT, RT and PTT when uncertain non-sample prior information is available, and compared with their classical counterparts. A robust version of Wald test that is given in Jurečková and Sen (1996, pp.419) is defined below

$$RW_n = \frac{\tilde{\gamma}^2}{\tilde{\sigma}^2}\left(\tilde{\beta}-\beta_0\right)'\left(X_n'X_n\right)\left(\tilde{\beta}-\beta_0\right), \tag{2.3}$$

where $\tilde{\gamma} = \sum \left({}_nS_n\right)^{-1}\psi'\left(\frac{Y_i - X_i'\tilde{\beta}}{S_n}\right)$ and $\tilde{\sigma}^2 = n^{-1}\sum \psi^2\left(\frac{Y_i - X_i'\tilde{\beta}}{S_n}\right)$ are respectively estimates of $\gamma = \frac{1}{s}\int_{-\infty}^{\infty}\psi'(u/s)\,dF(u)$ and $\sigma^2 = \int_{-\infty}^{\infty}\psi^2(u/s)\,dF(u) < \infty$. Note here $\tilde{\beta}$ is a robust M-estimator of β, studentized by S_n, (cf. Jurečková and Sen,1996, pp. 216), and is the solution to

$M_n(\beta) = \sum X_i'\psi\left(\frac{Y_i - X_i'\beta}{S_n}\right) = 0$, where Y_i is the responses variable on individual i, vector X_i is the ith row of the designed matrix X_n, the function ψ is a nondecreasing and skew symmetric score function, in the sense of Huber (1981). Here S_n is a scale statistic for estimating s = s(F), the scale parameter of distribution F. S_n is the scaled median absolute deviation (MAD) of $\left(Y_i - X_i'\tilde{\beta}\right)$. The classical Wald test is a special case of the robust Wald test. In this case $s = \sigma_0^2$ and taking $\psi(u/s) = u/s$, where $u = Y - X'\beta$, gives $\gamma = 1$ and $\sigma^2 = \sigma_0^2$.

3. Asymptotic Distribution of Robust M-estimator

To find the power function of the tests defined in Section 4, the following asymptotic distributions of the robust M-estimators are required. According to Jurečková and Sen (1996, pp.216), if $\sigma^2 = \int_{-\infty}^{\infty} \psi^2(u/s)dF(u) < \infty$, $\gamma = \frac{1}{s}\int_{-\infty}^{\infty} \psi'(u/s)dF(u)$ then

$$\sqrt{n}\left(\tilde{\beta} - \beta\right) \to N\left(0, \frac{\sigma^2}{\gamma^2} Q^{-1}\right), \tag{3.4}$$

where $Q = \lim_{n\to\infty} \frac{1}{n} Q_n$ with $Q_n = X_n' X_n$.

To facilitate the derivation of the joint distribution, consider the following matrix partitioning:

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \lim_{n\to\infty} \frac{1}{n} Q_n = \lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} Q_{n11} & Q_{n12} \\ Q_{n21} & Q_{n22} \end{pmatrix}, \tag{3.5}$$

where $Q_{njk} = X_{nj}' X_{nk}$ for j, k = 1, 2. The following theorem provides the asymptotic joint distributions of the robust M-estimators under the null hypothesis

Theorem 3.1 Under $H_0 : \beta_1 = \beta_{01}, \beta_2 = \beta_{02}$, asymptotically,

(i)

$$\sqrt{n}\begin{pmatrix} \tilde{\beta}_1 - & \beta_{01} \\ \tilde{\beta}_2 - & \beta_{02} \end{pmatrix} \xrightarrow{d} N_p\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{\sigma^2}{\gamma^2}\begin{pmatrix} Q_1^{*-1} & Q_{12}^{*} \\ Q_{21}^{*} & Q_2^{*-1} \end{pmatrix}\right], \tag{3.6}$$

(ii)

$$\sqrt{n}\begin{pmatrix} \tilde{\beta}_1 - & \beta_{01} \\ \tilde{\beta}_2 - & \beta_{02} \end{pmatrix} \xrightarrow{d} N_p\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{\sigma^2}{\gamma^2}\begin{pmatrix} Q_{11}^{*-1} & 0 \\ 0 & Q_2^{*-1} \end{pmatrix}\right], \tag{3.7}$$

where $\tilde{\beta}_1$ is the first r rows of $\tilde{\beta}_1$ $\tilde{\beta}_2$ is the last t rows of $\tilde{\beta}$ and $\tilde{\beta}_1$ is the restricted robust M-estimator of β_1, the solution of $M_{n1}\left(\beta_1, \beta_{02}\right) = 0$, where

$$M_{nj}\left(a,b\right) = \sum_{i=1}^{n} X'_{ji}\psi\left(\frac{Y_i - X'_{1i}a - X'_{2i}b}{S_n}\right) \tag{3.8}$$

for j = 1, 2 with Sn is a scale statistic of $s, a \in \Re_r$ and $b \in \Re_t$, Note here $Q^*_{12} = -Q^{-1}_{11}Q_{12}Q^{*-1}_{2}, Q^*_{21} = -Q^{-1}_{22}Q^{-1}_{21}Q^{*-1}_{1}, Q^*_1 = Q_{11} - Q_{12}Q^{-1}_{22}$ and $Q^*_2 = Q_{22} - Q_{21}Q^{-1}_{11}Q_{12}$.

The proof of Theorem 3.1 is given in appendix A. The asymptotic distributions of the robust M-estimators given in this section are used to obtain the power functions of robust Wald UT RT, PT and PTT defined in the next section.

4. The Classical and Robust Wald UT, RT, PT and PTT

In this section, the test statistics of the classical and robust Wald UT, RT, PT and PTT are defined.

4.1 The Classical and Robust Wald UT

In this section, the test statistics of the classical and robust Wald UT are defined.

(i) The Robust Wald UT (RWUT)

If β_2 is unspecified, the proposed test statistic for testing $H^*_0 : \beta_1 = \beta_{01}$ against $H^*_A : \beta_1 \neq \beta_{01}$ is

$$RW^{UT}_n = \tilde{\gamma}^2\left(\tilde{\beta}_1 - \beta_{01}\right)' Q^*_{n1}\left(\tilde{\beta}_1 - \beta_{01}\right) / \tilde{\sigma}^2, \tag{4.1}$$

where $Q^*_{n1} = Q_{n11} - Q_{n12}Q^{-1}_{n22}Q_{n21}$, in which $Q_{njk} = X'_{nj}X_{nk}$ for j, k = 1, 2. It follows from equation (3.6) that RW^{UT}_n follows a χ^2_r (chi-squared distribution with r degrees of freedom) under H^*_0 $as\ n \to \infty$.

(ii) The Classical Wald UT (CWUT)

The classical Wald UT is the nonrobust counterpart of the robust Wald UT and it is given as follows:

$$CW^{UT}_n = \left(\tilde{\beta}_1 - \beta_{01}\right)' Q^*_{n1}\left(\tilde{\beta}_1 - \beta_{01}\right) / \hat{\sigma}^2_0 \tag{4.2}$$

with $\tilde{\beta}_1$ is the first r rows of $\tilde{\beta}$. Note also that CW^{UT}_n follows as χ^2_r under H^*_0 as $n \to \infty$.

4.2 The Classical and Robust Wald RT

In this section, the test statistics of the classical and robust Wald RT are defined.

(i) The robust Wald RT (RWRT)

If $\beta_2 = \beta_{02}$ (specified), we find $Y_n = X_{n1}\beta_1 + X_{n2}\beta_{02} + e_n$, and the proposed test statistics for testing $H_0^*: \beta_1 = \beta_{01}$ against $H_A^*: \beta_1 \neq \beta_{01}$ is

$$RW_n^{RT} = \hat{\gamma}^2 \left(\hat{\beta}_1 - \beta_{01}\right)' Q_{n11} \left(\hat{\beta}_1 - \beta_{01}\right) / \tilde{\sigma}_2^2, \tag{4.3}$$

where $\tilde{\sigma}_2^2 = n^{-1} \times \sum \psi^2 \left(\frac{Y_i - X'_{1i}\hat{\beta}_1 - X'_2\beta_{02}}{S_n^{(2)}} \right)$ with $S_n^{(2)}$ is the scaled MAD of $\left(Y_i - X'_{1i}\hat{\beta}_1 - X'_{2i}\beta_{02}\right)$ and $\hat{\gamma} = \frac{1}{{}_nS_n^2} \times \sum \psi' \left(\frac{Y_i - X'_{1i}\hat{\beta}_1 - X'_{2i}\beta_{02}}{S_n^{(2)}} \right)$. It follows from equation (3.7) that $RW_n^{RT} \xrightarrow{d} \chi_r^2$ under $H_0^*: \beta_1 = \beta_{01}$ when $\beta_2 = \beta_{01}$ for large n.

(ii) The Classical Wald RT (CWRT)
The classical Wald RT is the nonrobust counterpart of the robust Wald RT and it is given as follows:

$$CW_n^{RT} = \left(\breve{\beta}_1^{\dagger} - \beta_{01}\right)' Q_{n11} \left(\breve{\beta}_1^{\dagger} - \beta_{01}\right) / \hat{\sigma}_2^2 \tag{4.4}$$

with $\breve{\beta}_1^{\dagger} = (X'_{n1}X_{n1})^{-1} X'_{n1}Y_n^{**}$, where $Y_n^{**} = Y_n - X'_{n2}\beta_{02}$ and $\hat{\sigma}_2^2 = \sum \left(Y_i - X'_{1i}\breve{\beta}_1^{\dagger} - X'_{i2}\beta_{02}\right)^2 / n$. The CW_n^{RT} follows a χ_r^2 under H_0^* as $n \to \infty$.

4.3 The Classical and Robust Wald PT

The test statistics of the classical and robust Wald PT are defined below.

(i) The robust Wald PT (RWPT)
For the preliminary test on the β_2, the proposed test statistic for testing $H_0^{(1)}: \beta_2 = \beta_{02}$ against $H_A^{(1)}: \beta_2 \neq \beta_{02}$ is given by

$$RW_n^{PT} = \tilde{\gamma}^2 \left(\tilde{\beta}_2 - \beta_{02}\right)' Q_{n2}^* \left(\tilde{\beta}_2 - \beta_{02}\right) / \tilde{\sigma}^2 \tag{4.5}$$

where $Q_{n2}^* = Q_{n22} - Q_{n21}Q_{n11}^{-1}Q_{n12}$. It follows from equation (3.7) that $RW_n^{PT} \xrightarrow{d} \chi_t^2$ under $H_0^{(1)}$.

(ii) The Classical Wald PT (CWPT)

The classical Wald PT is the nonrobust counterpart of the robust Wald PT and it is given follow:

$$CW_n^{UT} = \left(\breve{\beta}_2 - \beta_{02}\right)' Q_{n2}^* \left(\breve{\beta}_2 - \beta_{02}\right) / \hat{\sigma}_0^2, \qquad (4.6)$$

with $\breve{\beta}_2$ is the last t rows of $\breve{\beta}$. Then, $CW_n^{PT} \xrightarrow{d} \chi_t^2$ under $H_0^{(1)}$.

4.4 The Classical and Robust Wald PTT

If the null hypothesis of this pre-test is rejected, the UT is appropriate to test H_0^*, otherwise the RT is used to test H_0^*. The final test for testing $H_0^{(1)}$, following a pre-test on $H^{(1)}$, is termed as the pre-test test (PTT). The test statistic for testing H_0^* following a pre-test on $\boldsymbol{\beta}_2$ is a choice between RT and UT. The UT is used if $H_0^{(1)}: \beta = \beta_0$ is rejected and the RT is used if $H^{(1)}$ is accepted.

The asymptotic power functions of the classical and robust Wald UT, RT and PT are based on the univariate χ^2 probability distribution, but the asymptotic power function for the final PTT involves two bivariate χ^2 distributions (see Appendix B and C for details). Here the PT and RT are independent and PT and UT are correlated. So, the correlated bivariate non central χ^2 distribution is used to find the asymptotic power function of the PTT. Thus, the computation of the power of the PTT using the asymptotic power function involves the bivariate noncentral χ^2 probability integral. Instead of using directly the asymptotic power function formula to compute the power of the PTT, a Monte Carlosimulation method was used in this study.

5. Power Comparison using a Monte Carlo Simulation

To compare the performance of the tests, the analytical comparison is unrealistic. Instead a Monte Carlo simulation method is used to compare the power of the tests. In this section, the power of the classical and robust Wald UT, RT and PTT are obtained using a computer generated Monte Carlo experiment. The objectives of this section are to determine which of the UT, RT and PTT is better, and to compare the proposed robust Wald UT, RT and PTT to the classical Wald UT, RT and PTT, each under normality and a slight change to normality. A multiple liner regression model with three parameter $y_i = \theta_1 + \theta_2 x_{1i} + \beta x_{2i} + \varepsilon_i$ for i = 1, 2, ...,n was considered in the simulation. Here, take n = 100. The error terms ε_i, i = 1, 2, ...,n are generated randomly from (i) normal with mean 0 and variance 1, N (0, 1) (ii) 10% wild: First ε_i is generated from normal distribution with mean 0 and variance 1, then choose randomly 5% of the generated ε_i and multiply them by a scalar 10, and another 5% choose randomly 5% is multiply by a scalar -10. The observed values of the regression x_{1i} where generated randomly from a uniform distribution with minimum and maximum values of 0 and 1, and those of x_{2i} were from normal distribution with mean 1 and variance 1. We let $\theta_1 = \theta_{10} + n^{-\frac{1}{2}}\delta_1, \theta_2 = \theta_{20} + n^{-\frac{1}{2}}\delta_1; \quad \beta = \beta_0 + n^{-\frac{1}{2}}\delta_2, \quad with\ \delta_1, \delta_2 \geq 0, \theta_{10}, \theta_{20} \in \Re$ and then generate a random sample for selected values of $\theta_1, \theta_2\ and\ \beta$.

For selected values of δ_1 and δ_2, 5000 simulations were run in which a sample was drawn from normal or 10% wild distributions. Each of the tests was run with in each simulation and $H_0^*:\theta_1=\theta_{10}$, $\theta_2=\theta_{20}$ was either rejected or not rejected at the 5% significance level. Size of the UT, RT and PTT is the probability of rejecting the null hypothesis $H_0^*:\theta_1= \theta_{10}$, $\theta_2= \theta_{20}$ when it is true. First, we generate data set for δ_1=0. Then, each test statistic was computed for the dataset. We then find the proportion of tests rejecting the null hypothesis from 5000 simulated data sets, and then use it to estimate the size of the test for the UT and the RT. On the other hand, the power of the test is the probability of rejecting false H_0^*. We generate a different data set for an arbitrary positive value of δ_1 (eg.2, 4). As δ_1 moves away from 0, it is suspected that the power of test increases. To estimate the probability of rejecting the null hypothesis H_0^*, we find the proportion of rejecting the null hypothesis from 5000 simulated data sets when the true values of θ_1 and θ_2 are not respectively θ_{10} and θ_{20}, that is, when $\delta_1>0$.

For the PTT, the UT is used if $H_0^{(1)}:\beta=\beta_0$ is rejected and the RT is used if $H_0^{(1)}$ is accepted. So, the PTT is either the UT or RT depending on the outcome of the PT. The proportion of rejecting H_0^* from the UT or the RT following the result from the PT among 5000 generated data sets is taken as the probability of rejecting H_0^* for the PTT.

Since the classical and robust Wald UT, RT and PTT are defined based on the knowledge of β, it is of interest to compare the size and power of the tests by plotting them against δ_2, where $\delta_2 = \sqrt{n}(\beta - \beta_0)$. Figure1 (in Appendix D) shows the size and power of the classical and robust Wald UT, RT and PTT, when the distribution of the error term in the multiple linear regression model is $N(0,1)$ and 10% wild, and sample size is n =100 for selected values of δ_1. Our first aim is to determine which of the UT, RT and PTT is better, both in the size and power of the test. Although the classical and robust Wald RT have the largest power in comparison to those of the UT and PTT, it also has the largest size as δ_2 grows larger. The RT is defined when $\beta=\beta_0$, and it is as expected that the size of RT increases as $\delta_2 = \sqrt{n}(\beta - \beta_0)$ increases. On the other hand, the size and power of the UT are constant regardless the value of δ_2. This is because UT treats β as a nuisance parameter. In comparison to RT, the UT has the smallest size, but also the smallest power for small δ_2 because β is not specified at the nullhypothesis in the specification of the UT. The PTT is either the UT or RT depending on the PT. Thus it is a compromise between the two tests. The PTT behaves similar to the RT for small δ_2, that is, when PT accepts $H_0^{(1)}$. On the other hand, it behaves similar to the UT when $H_0^{(1)}$ is rejected using the PT. Thus, for a larger value of δ_2, it behaves similar to the UT. The PTT is better than the RT in terms of size and it is better than the UT in terms of power when δ_2 is small. Although the prior information on the β vector is uncertain, there is a high possibility that its true value is quite close to the suspected value. Therefore, the study on the behaviour of the three tests when δ_2 is small is more realistic.

It is of interest to see the effect of wild observations in the data on the classical and robust Wald tests performance. Figure1(in Appendix D) depicts that the power of classical Wald UT, RT and PTT are about the same as those of the robust Wald UT, RT and PTT when the distribution of the error term is normal (see Figures 1(a), (c) and (e)). However, under a slight change to normality, the classical test lost its power and the robust Wald PTT has shown a remarkable performance in terms size and power compared to the classical Wald tests when the error term is 10% wild (see Figures 1(b), (d) and (f)). The robust Wald PTT has shown some robustness property under a slight change to normality assumption through this simulation example.

The performance of a test depends on the sample size. Figure 2 (in Appendix D) shows the power of the classical and robust Wald UT, RT and PTT for data generated from normal or
10% wild distributions when $\delta_1 = 2$ and for some selected sample sizes n = 40, 60 and 120. The proportion of rejecting H_0^* is plotted against Δ_β, where $\Delta_\beta = \beta - \beta_0 = \delta_2 / \sqrt{n}$. As *n* grows larger, the performance of the tests differ. When the distribution of the error term is normal, the power of both classical and robust Wald UTs are same and constant but increase as *n* grows larger. The power of the classical and robust Wald RT increase as Δ_β increases and are larger for data with a larger sample size than that with a smaller sample size. The power of the PTT is better than that of the UT and is larger for data with a larger sample size when Δ_β is small, i.e. when the suspected value is close to the true β. When the distribution of the error term is10% wild, the classical and robust Wald tests performance differ more as *n* gets smaller. Again, the powers of the tests are larger for data with larger sample sizes. The power of all the classical UT, RT and PTT tests are smaller than those of the robust Wald tests for smaller Δ_β, i.e. when the suspected value is close to the true β. Even for a larger value of Δ_β, i.e. in case when the suspected value is quite far away from the true β, the robust Wald tests have more power than the classical tests for data with a smaller *n* that those with a larger sample size. As sample size grows larger, the robust Wald tests are more insensitive to

wild observations (outliers) and have closer power performance to those tests in the normal case. The classical Wald tests fail to maintain similar power as they have in the normal case when the sample size is small and in the presence of wild observations.

In the next section, the test statistics of the UT, RT and PTT are computed for the Olympic athletes data set.

6. Application on Data

In this section, the proposed robust Wald UT, RT and PT were used on a set of real life data to view the effect of pre-testing (PT) on the final test (PTT) and the effectiveness of the robust Wald test compared to its classical counterpart. For this illustration, we used 2012 Olympic athletes data which can be accessed from the Guardian website,
http://www.theguardian.com/sport/datablog/2012/aug/07/olympic-2012-athletes-age-weight-height.

The data set contains several variables related to the athletes from 205 countries, including place of birth, height, weight, age, type of sport, etc. However, we use the data set on Australia athletes and focus on three variables weight, height and age. The height (in cm) and age (in year) are the independent variables and weight (in kilogram) is the response variable. The regression model is as follows:

$$\text{weight}_i = \beta_0+\beta_1 \,\text{height}_i +\beta_2 \,\text{age}_i +e_i. \tag{6.7}$$

The quantile-quantile normal plot and scatter plot of the standardized residuals (see Figure 3, as in Appendix D) from the least-squares fitted model revealed several outliers in the data. The least squares estimates of β_0, β_1 and β_2 are 138.8, 0.587, and -0.1178, respectively. Table 1 gives the test results for two hypothesis: (i) $H_0^* : \beta_0 = 140;\ \beta_1 = 0.6$ (using RT and UT), and (ii) $H_0^{(1)} : \beta_2 = 0$ (using PT).

In the case of the robust Wald test, the age is not significant at the 5 % level for the PT (P=0.052). Thus, PTT becomes the RT, and we then reject at the 5% level (P<0.001). On the contrary, age becomes significantly important in the case of the classical Wald test at the 5% significance level (P = 0.044), so we use the UT as the final test, and we cannot reject H_0^* (P=0.323).

So the classical Wald PTT cannot reject H_0^* using the classical Wald UT. But under the robust Wald test, the PTT becomes the RT leadings to the rejection of H_0^*. Since the robust Wald PTT has more power than the classical Wald PTT under a slight departure from normality assumption, we decide to reject H_0^* at the 5% significance level.

7. Concluding remarks

In this paper, we have introduced robust Wald UT, RT and PT in the context of multiple regression model in the preliminary testing framework. Theorem 3.1 shows it is clear that there is a correlation between the two components of the test statistics, namely the robust Wald UT and PT, but no correlation between that of the robust Wald RT and PT from the structure of the covariance matrix in equations (3.6) and (3.7). Since robust Wald PTT is either UT or RT depending on the outcome of the PT, its asymptotic distribution is determined from the distribution of the test statistics for the UT, RT and PT. It is found that, the asymptotic distribution of the robust Wald PTT is a bivariate noncentral chi-squared distribution with the noncentrality parameters of the robust Wald UT, RT and PT. The asymptotic joint distribution of the test statistics under the alternative hypothesis is used to obtain the asymptotic power function (See Appendix B and C for details).

For the purpose of comparing the power of the competing tests, the Monte Carlo experiment is used. The asymptotic power of the PTT is obtained as the proportion of the rejection of the UT or the RT in N replicated samples, following the result from the PT. From the simulation, both the classical and robust Wald tests are preferred when normality assumption is satisfied, as both tests have similar performance in terms of size and power of the test. However, in the presence of contamination or wild observations in the data, the classical Wald test lose its power. On the other hand, the power of the robust Wald test is not affected by the wild observations in the data. Thus, the robust Wald test has shown a remarkable robustness property under a slight departure from normality assumption.

The robust Wald RT has the highest power compared to those of the UT and PTT, but it also has the largest size. On the contrary, the robust Wald UT's size is the smallest, but its power is also the smaller expect when $\lambda_1 = n^{\frac{1}{2}}(\beta_1 - \beta_{01})$ or $\lambda_2 = n^{\frac{1}{2}}(\beta_1 - \beta_{02})$ is large. So, both robust Wald UT and RT fail to attain the lowest size and the highest power criteria. The robust Wald PTT's size is smaller than that of the robust Wald RT. Its power is also higher than the robust Wald UT, with the exception of very large values of $\boldsymbol{\lambda}_1$ or $\boldsymbol{\lambda}_2$. Consequently, if the prior information on the value of $\boldsymbol{\beta}_2$ is not far from its true value, that is, $\boldsymbol{\lambda}_2$ is near $\mathbf{0}$ (small or moderate) difference the robust Wald PTT's size is smaller than that of the RT, and its power is higher than that of the UT. Thus, the robust Wald PTT is a better choice among the three tests regardless the normality assumption is violated or not. Since the prior information is given by experienced experts in the field or from previous studies, the value of $\boldsymbol{\lambda}_2$ should not be far from $\mathbf{0}$, even though it may not be $\mathbf{0}$, and therefore the robust Wald PTT is preferable over those of the UT and RT.

Acknowledgments

The second author was partially supported by University of Malaya under Grant no. BKS077-2017.

References

Ahmed, S.E., Saleh, A.K.M.E. (1989). Estimation strategy using a preliminary test in some normal models. Soochow. J. of Math. 14: 135-165.

Akritas, M.G., Saleh, A.K.M.E., Sen, P.K. (1984). Nonparametric estimation of intercepts after a preliminary test on parallelism of several regression lines. In: Sen P.K. ed., *Biostatis- tics: Statistics in Biomedical, Public Health and Environmental Sciences*. North-Holland, New York: The Bernard G. Greenberg Vol.

Atkinson, A.C., Lawrance, A.J. (1989). A comparison of asymptotically equivalent test statistics for regression transformation. Biometrika 76: 2, 223–229.

Bancroft, T.A. (1944). On Biases in Estimation Due to the Use of Preliminary Tests of Significance. Ann. Math. Statist. 15: 190 - 204.

Basu, A., Ghosh A., Mandal, A., Martin, N., and Pardo, L. (2017). A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimators. Electron J Stat 11: 2741 - 2772.

Basu, A., Ghosh A., Martin, N., and Pardo, L. (2018). Robust Wald-type tests for non-homogeneous observations based on the minimum density power divergence estimator. Metrika 81: 493 - 522.

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

Carolan, A.M., Rayner, J.C.W. (2000). Wald tests of location for symmetric nonnormal data. Biometrical Journal 42: 6, 777-792.

Carroll, R.J., Ruppert, D. (1988). *Transformation and Weighting in Regression*. US: Chapman& Hall/CRC.

Engle, R.F. (1984). Wald, Likelihood Ratio and Lagrange Multiplier Tests in Economics. In Griliches Z. and Intriligator M.D. eds., *Handbook of Econometrics, Volume II*, Elsevier, pp. 775-826.

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

Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

Heritier, S., Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. Journal of the American Statistical Association 89: 897 - 904.

Joarder, A.H. (2006). Product moments of bivariate Wishart distribution. J. Prob. and Stat Sci. 4: 233 - 243.

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

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

Khan, S. (2000). Improved estimation of the mean vector for Student-t model. Communicationsin 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: 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: 79-94.

Khan S., Pratikno B. (2013). Testing base load with sample prior information on process load. Statistical Papers 54: 605-617.

Lambert, A., Saleh A.K.Md.E., Sen P.K. (1985). Test of homogeneity of intercepts after a preliminary test on parallelism of several regression lines: from parametric to asymptotic. Comm. Statist. 14: 2243-2259.

Mead, R., Bancroft, T.A., Han, C.P.(1975). Power of analysis of variance test procedures for incompletely specified fixed models. Ann. Statist. 3: 797-808.

Muirhead, R.J. (1982). *Aspects of multivariate statistical theory*. New York: John Wiley & Sons.

Neyman, J., Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical ineferences, Part I. Biometrika 20: 175-240.

Paull, A.E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. Ann. Math. Statist. 21: 539 - 556.

Rayner, J.C.W., Best, D.J. (1989). *Smooth Tests of Goodness of fit*. New York: Oxford University Press.

Rao, C.R. (1948). Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation. Proceedings of the Cambridge Philosophical Society 44: 50-57.

Saleh, A.K.Md.E., Sen, P.K. (1982). Non-parametric tests for location after a preliminary test on regression. Communications in Statistics: Theory and Methods. 11: 639 - 652.

Saleh, A.K.Md.E. Sen, P.K. (1983). Nonparametric tests of location after a preliminary test on regression in the multivariate case. Communications in Statistics: Theory and Methods 12: 1855-1872.

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

Sen, P.K., Singer, J.M., Pedroso, de Lima A.C. (2010). *From Finite Sample to Asymptotic Methods in Statistics*. New York: Cambridge University Press.

Silvapulle, M.J. (1992). Robust Wald-type tests of one-sided hypotheses in the linear model. Journal of the American Statistical Association. 87:156 - 161.

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

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the numberof observations is large. Trans. Am. Math. Soc. 54: 426-482.

Yunus, R.M., Khan, S. (2010). Increasing power of robust test through pre-testing in multiple regression model. Pakistan Journal of Statistics. Special Volume on 25th Anniversary, edby Ahmed, S. E. 26: 1, 151-170.

Yunus, R.M., Khan, S. (2011a). Increasing the power of the test through pre-test - a robust method. Communications in Statistics - Theory and Methods 40: 581 - 597.

Yunus, R.M., Khan, S. (2011b). M-tests for multivariate regression model. Journal of Non-parametric Statistics 23: 201-218.

Yunus, R.M., Khan, S. (2011c). The bivariate noncentral chi-square distribution - A compound distribution approach. Journal of Applied Mathematics and Computation 217: 6237-6247.

A Appendix A

Proof of Theorem 3.1(i)

Following Theorem 5.5.1 of Jurečková and Sen (1996), we find that $\tilde{\boldsymbol{\beta}}$ admits the asymptotic representation

$$\begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix} = (n\gamma)^{-1} \begin{pmatrix} \mathbf{Q}_{n11} & \mathbf{Q}_{n12} \\ \mathbf{Q}_{n21} & \mathbf{Q}_{n22} \end{pmatrix}^{-1} \sum_{i=1}^{n} \mathbf{X}_i \psi(u_i/s) + R_n \tag{1}$$

with $R_n = O_p(n^{-1})$ and

$$\begin{pmatrix} \mathbf{Q}_{n11} & \mathbf{Q}_{n12} \\ \mathbf{Q}_{n21} & \mathbf{Q}_{n22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{Q}_{n1}^{\star\;-1} & -\mathbf{Q}_{n11}^{-1}\mathbf{Q}_{n12}\mathbf{Q}_{n2}^{\star\;-1} \\ -\mathbf{Q}_{n22}^{-1}\mathbf{Q}_{n21}\mathbf{Q}_{n1}^{\star\;-1} & \mathbf{Q}_{n2}^{\star\;-1} \end{pmatrix},$$

where $\mathbf{Q}_{n1}^{\star} = \mathbf{Q}_{n11} - \mathbf{Q}_{n12}\mathbf{Q}_{n22}^{-1}\mathbf{Q}_{n21}$ and $\mathbf{Q}_{n2}^{\star} = \mathbf{Q}_{n22} - \mathbf{Q}_{n21}\mathbf{Q}_{n11}^{-1}\mathbf{Q}_{n12}$.

Since $(\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\mathbf{Q}_{2}^{\star -1})' = \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{1}^{\star -1}$ and $\mathrm{Var}\,(\mathbf{X}_i\psi(u_i/s)) = n\sigma^2\mathbf{Q}_n$, the proof of part (i) follows after some algebra.

Proof of Theorem 3.1(ii)

It follows from (1) that

$$\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = \frac{1}{n\gamma}\left\{\mathbf{Q}_{n1}^{\star\;-1}\sum \mathbf{X}_{1i}\psi\left(\frac{u_i}{s}\right) - \mathbf{Q}_{n11}^{-1}\mathbf{Q}_{n12}\mathbf{Q}_{n2}^{\star\;-1}\sum \mathbf{X}_{2i}\psi\left(\frac{u_i}{s}\right)\right\} + R_n \tag{2}$$

and

$$\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 = \frac{1}{n\gamma}\left\{-\mathbf{Q}_{n22}^{-1}\mathbf{Q}_{n21}\mathbf{Q}_{n1}^{\star\;-1}\sum \mathbf{X}_{1i}\psi\left(\frac{u_i}{s}\right) + \mathbf{Q}_{n2}^{\star\;-1}\sum \mathbf{X}_{2i}\psi\left(\frac{u_i}{s}\right)\right\} + R_n. \tag{3}$$

On the other hand, we admits that

$$\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = \frac{1}{n\gamma}\left\{\mathbf{Q}_{n11}^{-1}\sum \mathbf{X}_{1i}\psi(u_i/S)\right\} + R_n \tag{4}$$

when $\boldsymbol{\beta}_2 = \mathbf{0}$. Then, (2), (3) and (4) imply that

$$\mathbf{Q}_{n22}^{-1}\mathbf{Q}_{n21}\mathbf{Q}_{n1}^{\star\;-1}\mathbf{Q}_{n11}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = -(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + \frac{1}{n\gamma}\mathbf{Q}_{n2}^{\star\;-1}\sum \mathbf{X}_{2i}\psi\left(\frac{u_i}{s}\right) + R_n \tag{5}$$

and

$$\mathbf{Q}_{n1}^{\star\;-1}\mathbf{Q}_{n11}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = (\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + \frac{1}{n\gamma}\mathbf{Q}_{n11}^{-1}\mathbf{Q}_{n12}\mathbf{Q}_{n2}^{\star\;-1}\sum \mathbf{X}_{2i}\psi\left(\frac{u_i}{s}\right) + R_n \tag{6}$$

after some algebra. Then solving equations (5) and (6) take us to

$$(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = (\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + \mathbf{Q}_{n11}^{-1}\mathbf{Q}_{n12}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + R_n. \tag{7}$$

It follows from equation (7) that

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_r & \mathbf{Q}_{n11}^{-1}\mathbf{Q}_{n12} \\ 0 & \mathbf{I}_s \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix} + R_n. \tag{8}$$

Thus the asymptotic distribution of $\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix}$ under H_0 is bivariate normal with zero mean vector and covariance matrix

$$\frac{\sigma^2}{\gamma^2}\begin{pmatrix} \mathbf{I}_r & \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12} \\ 0 & \mathbf{I}_s \end{pmatrix}\begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{I}_r & \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12} \\ 0 & \mathbf{I}_s \end{pmatrix}' = \frac{\sigma^2}{\gamma^2}\begin{pmatrix} \mathbf{Q}_{11}^{\;-1} & 0 \\ 0 & \mathbf{Q}_{2}^{\star -1} \end{pmatrix}. \tag{9}$$

B Appendix B

In this section, two theorems related to the distributions of the M-estimators and test statistics are derived under a sequence of local alternatives.

Theorem B.1 *Let $\{K_n\}$ be a sequence of local alternatives, where*

$$K_n : (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\boldsymbol{\beta}_{01} + n^{-\frac{1}{2}}\boldsymbol{\lambda}_1, \boldsymbol{\beta}_{02} + n^{-\frac{1}{2}}\boldsymbol{\lambda}_2), \tag{10}$$

with $\boldsymbol{\lambda}_1 = n^{\frac{1}{2}}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{01}) > \mathbf{0}$ and $\boldsymbol{\lambda}_2 = n^{\frac{1}{2}}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{02}) > \mathbf{0}$ are (fixed) real numbers. Under $\{K_n\}$, asymptotically,

$$\sqrt{n}\begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01} \\ \tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{02} \end{pmatrix} \xrightarrow{d} N_p\left[\begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix}, \frac{\sigma^2}{\gamma^2}\begin{pmatrix} \boldsymbol{Q}_1^{\star-1} & \boldsymbol{Q}_{12}^{\star} \\ \boldsymbol{Q}_{21}^{\star} & \boldsymbol{Q}_2^{\star-1} \end{pmatrix}\right], \tag{11}$$

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01} \\ \tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{02} \end{pmatrix} \xrightarrow{d} N_p\left[\begin{pmatrix} \boldsymbol{\lambda}_1 + \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2 \\ \boldsymbol{\lambda}_2 \end{pmatrix}, \frac{\sigma^2}{\gamma^2}\begin{pmatrix} \boldsymbol{Q}_{11}{}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2^{\star-1} \end{pmatrix}\right]. \tag{12}$$

Proof The proof of this Theorem is obtained directly from equations (3.6) and (3.7) using the contiguity probability measures (Hájek et al. 1999). ■

Theorem B.2 *Under $\{K_n\}$, asymptotically (RW_n^{RT}, RW_n^{PT}) are independently distributed as bivariate noncentral chi-squared distribution with (r, t) degrees of freedom and (RW_n^{UT}, RW_n^{PT}) are distributed as correlated bivariate noncentral chi-squared distribution with (r, t) degrees of freedom and noncentrality parameters,*

$$\begin{aligned}
\theta^{UT} &= \gamma^2(\boldsymbol{\lambda}_1'\boldsymbol{Q}_1^{\star}\boldsymbol{\lambda}_1)/\sigma^2, & (13)\\
\theta^{RT} &= \gamma^2(\boldsymbol{\lambda}_1'\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_1'\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_2'\boldsymbol{Q}_{21}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2)/\sigma^2, & (14)\\
\theta^{PT} &= \gamma^2(\boldsymbol{\lambda}_2'\boldsymbol{Q}_2^{\star}\boldsymbol{\lambda}_2)/\sigma^2. & (15)
\end{aligned}$$

Proof From Theorem B.1, we find that

$$n^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01}) \xrightarrow{d} N_r\left(\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2, (\sigma^2/\gamma^2)\boldsymbol{Q}_{11}^{-1}\right) \tag{16}$$

and using Theorem 1.4.1 of Muirhead (1982)

$$RW_n^{RT} = \hat{\gamma}^2(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})'\mathbf{Q}_{n11}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})/\hat{\sigma}^2 \tag{17}$$

is a chi-squared distribution with r degrees of freedom and noncentrality parameter

$$\theta^{RT} = \gamma^2(\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2)'\boldsymbol{Q}_{11}(\boldsymbol{\lambda}_1 + \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2)/\sigma^2 \tag{18}$$

which is simplified as $\theta^{RT} = \frac{\gamma^2}{\sigma^2}(\boldsymbol{\lambda}_1'\boldsymbol{Q}_{11}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_1'\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_2'\boldsymbol{Q}_{21}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\lambda}_2)$ after some algebra. In the same manner, the other two noncentrality parameters θ^{UT} and θ^{PT} are obtained.

There is no correlation between RW_n^{RT} and RW_n^{PT} because the covariance matrix between $\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})$ and $\sqrt{n}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{02})$ is a zero matrix.

Note that for any two variables Z_1 and Z_2 that follow a bivariate normal with mean 0 and covariance matrix $\begin{pmatrix} \sigma_{z1}^2 & \rho_z\sigma_{z1}\sigma_{z2} \\ \rho_z\sigma_{z1}\sigma_{z2} & \sigma_{z2}^2 \end{pmatrix}$, the correlation coefficient between $U = Z_1^2/\sigma_{z1}^2$ and $V = Z_2^2/\sigma_{z2}^2$ is ρ_z^2 (cf. Joarder 2006).

From Theorem 3.1, the covariance matrix of two vectors $\sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})$ and $\sqrt{n}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{02})$ is a nonzero matrix of size r by s. The (i, j)th element of this covariance matrix is the covariance of the ith element of vector $\sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01})$ and the jth element of vector $\sqrt{n}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{02})$. Denote $\rho_k^{\diamond}$ as the correlation coefficient for any two different elements of the augmented vector $(\sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01}), \sqrt{n}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{02}))$, following Joarder (2006), the correlation between RW_n^{UT} and RW_n^{PT} is $\sum_{k=1}^{p} \rho_k^{\diamond 2}/p$. ■

The asymptotic distributions of test statistics given in this section are used to derive the asymptotic power functions of the test statistics in the following section.

C Appendix C

Using results in Theorem B.2, under $\{K_n\}$, the asymptotic power function for the UT, RT and PT is

$$\Pi^{UT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \lim_{n\to\infty} \Pi_n^{UT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \lim_{n\to\infty} P(RW_n^{UT} > \ell_{n,\alpha_1}^{UT}|K_n) = 1 - G_r(\chi_{r,\alpha_1}^2; \theta^{UT}), \quad (19)$$

$$\Pi^{RT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \lim_{n\to\infty} \Pi_n^{RT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \lim_{n\to\infty} P(RW_n^{RT} > \ell_{n,\alpha_2}^{RT}|K_n) = 1 - G_r(\chi_{r,\alpha_2}^2; \theta^{RT}), \quad (20)$$

$$\Pi^{PT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \lim_{n\to\infty} \Pi_n^{PT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \lim_{n\to\infty} P(RW_n^{PT} > \ell_{n,\alpha_3}^{PT}|K_n) = 1 - G_s(\chi_{s,\alpha_3}^2; \theta^{PT}), \quad (21)$$

respectively, with $G_k(\chi_{k,\alpha}^2; \theta)$ is the cumulative distribution function of the noncentral chi-squared distribution with k degrees of freedom and noncentrality parameter θ, and level of significance α. Here, $\chi_{k,\alpha}^2$ is the upper $100\alpha\%$ critical value of a central chi-squared distribution and $\ell_{n,\alpha_1}^{UT} \to \chi_{r,\alpha_1}^2$ under $H_0^\star$, $\ell_{n,\alpha_2}^{RT} \to \chi_{r,\alpha_2}^2$ under $H_0^\star$ when $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{02}$, and $\ell_{n,\alpha_3}^{PT} \to \chi_{s,\alpha_3}^2$ under $H_0^{(1)}$. Since the PTT is a choice between RT and UT, define the power function of the PTT as

$$\begin{aligned} & \Pi_n^{PTT}(\boldsymbol{\beta}_1) \\ = \; & E(I\left[(RW_n^{PT} < \ell_{n,\alpha_3}^{PT}, RW_n^{RT} > \ell_{n,\alpha_2}^{RT}) \text{ or } (RW_n^{PT} > \ell_{n,\alpha_3}^{PT}, RW_n^{UT} > \ell_{n,\alpha_1}^{UT})\right] |\boldsymbol{\beta}_1) \end{aligned} \quad (22)$$

where ℓ_{n,α_3}^{PT} is the critical value of RW_n^{PT} at the α_3 level of significance and $I(A)$ is an indicator function of the set A which takes value 1 if A occurs, otherwise it is 0. The size of the PTT is obtained by substituting $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{01}$ in equation (22). For testing $H_0^\star$ following a pre-test on $\boldsymbol{\beta}_2$, using equation (22) and the results of Theorem B.2, the asymptotic power function for the PTT under $\{K_n\}$ is given by

$$\begin{aligned} & \Pi^{PTT}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \\ = \; & \lim_{n\to\infty} P(RW_n^{PT} \le \ell_{n,\alpha_3}^{PT}, RW_n^{RT} > \ell_{n,\alpha_2}^{RT}|K_n) + \lim_{n\to\infty} P(RW_n^{PT} > \ell_{n,\alpha_3}^{PT}, RW_n^{UT} > \ell_{n,\alpha_1}^{UT}|K_n) \\ = \; & G_s(\chi_{s,\alpha_3}^2; \theta^{PT})\{1 - G_r(\chi_{r,\alpha_2}^2; \theta^{RT})\} + \int_{\chi_{r,\alpha_1}^2}^{\infty} \int_{\chi_{s,\alpha_3}^2}^{\infty} \phi^\star(w_1, w_2) dw_1 dw_2, \end{aligned} \quad (23)$$

where, $\phi^\star(\cdot)$ is the density function of a bivariate noncentral chi-squared distribution with probability integral given by

$$\begin{aligned}
&\int_{\chi^2_{r,\alpha_1}}^{\infty}\int_{\chi^2_{t,\alpha_3}}^{\infty} \phi^\star(w_1,w_2)dw_1dw_2 \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\delta_1=0}^{\infty}\sum_{\delta_2=0}^{\infty}(1-\rho^{\star 2})^{(r+t)/2}\frac{\Gamma(\frac{r}{2}+j)}{\Gamma(\frac{r}{2})j!}\frac{\Gamma(\frac{t}{2}+k)}{\Gamma(\frac{t}{2})k!}\rho^{\star 2(j+k)} \\
&\times\left[1-\gamma^\star\left(\frac{r}{2}+j+\delta_1,\frac{\chi^2_{r,\alpha_1}}{2(1-\rho^{\star 2})}\right)\right]\left[1-\gamma^\star\left(\frac{t}{2}+k+\delta_2,\frac{\chi^2_{t,\alpha_3}}{2(1-\rho^{\star 2})}\right)\right] \\
&\times\frac{e^{-\theta^{UT}/2}(\theta^{UT}/2)^{\delta_1}}{\delta_1!}\frac{e^{-\theta^{PT}/2}(\theta^{PT}/2)^{\delta_2}}{\delta_2!},
\end{aligned} \tag{24}$$

with (r,t) degrees of freedom, noncentrality parameters, θ^{UT} and θ^{PT}, correlation coefficient, $\rho^{\star 2}$, with $-1<\rho^\star<1$ and the incomplete gamma function, $\gamma^\star(v,d)=\int_0^d x^{v-1}e^{-x}/\Gamma(v)dx$. See Yunus and Khan (2011c) for details of the bivariate probability integral. Let $\rho^{\star 2}=\sum_{k=1}^{p}\frac{1}{p}\rho_k^{\diamond 2}$, the mean correlation, where $\rho_k^\diamond$ is the correlation coefficient for any two different elements of the augmented vector $[\ n^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_{01}),\ \ n^{-\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_{02})\ ]$ in equation (3.6).

Table 1: Test results from weight-height-age relationship of Australian Olympic athletes

	robust Wald test			classical Wald test		
Hypothesis	Test	χ^2	p-value	Test	χ^2	p-value
$H_0^\star:\beta_0=140,$	UT	5.35	0.069	UT	2.27	0.323
$\beta_1=0.6,$	RT	216.91	<0.001	RT	223.73	<0.001
$H_0^{(1)}:\beta_2=0$	PT	3.78	0.052	PT	4.07	0.044

D Appendix D - Figures

1. Figure 1. Size and power of the classical and robust Wald UT, RT and PTT for simulated data sized $n=100$ with normal or 10% wild errors and for selected values of δ_1.
2. Figure 2. Power of the classical and robust Wald UT, RT and PTT for simulated data sized $n=40,60,100$ with normal or 10% wild errors when $\delta_1=2$.
3. Figure 3. Diagnostics from weight-height-age linear relationship for n =399 Australian Olympic athletes.

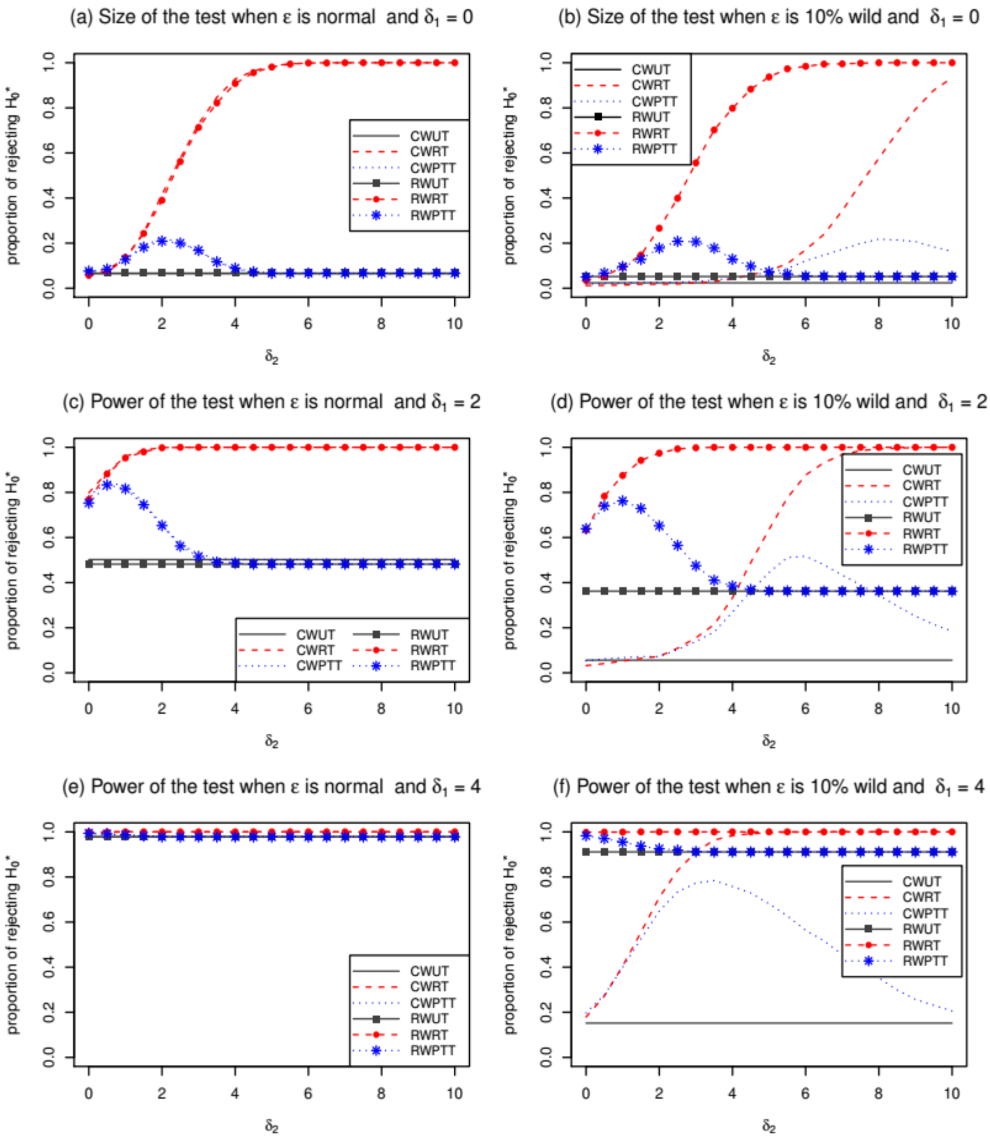

Figure 1: Size and power of the classical and robust Wald UT, RT and PTT for simulated data sized $n = 100$ with normal or 10% wild errors and for selected values of δ_1.

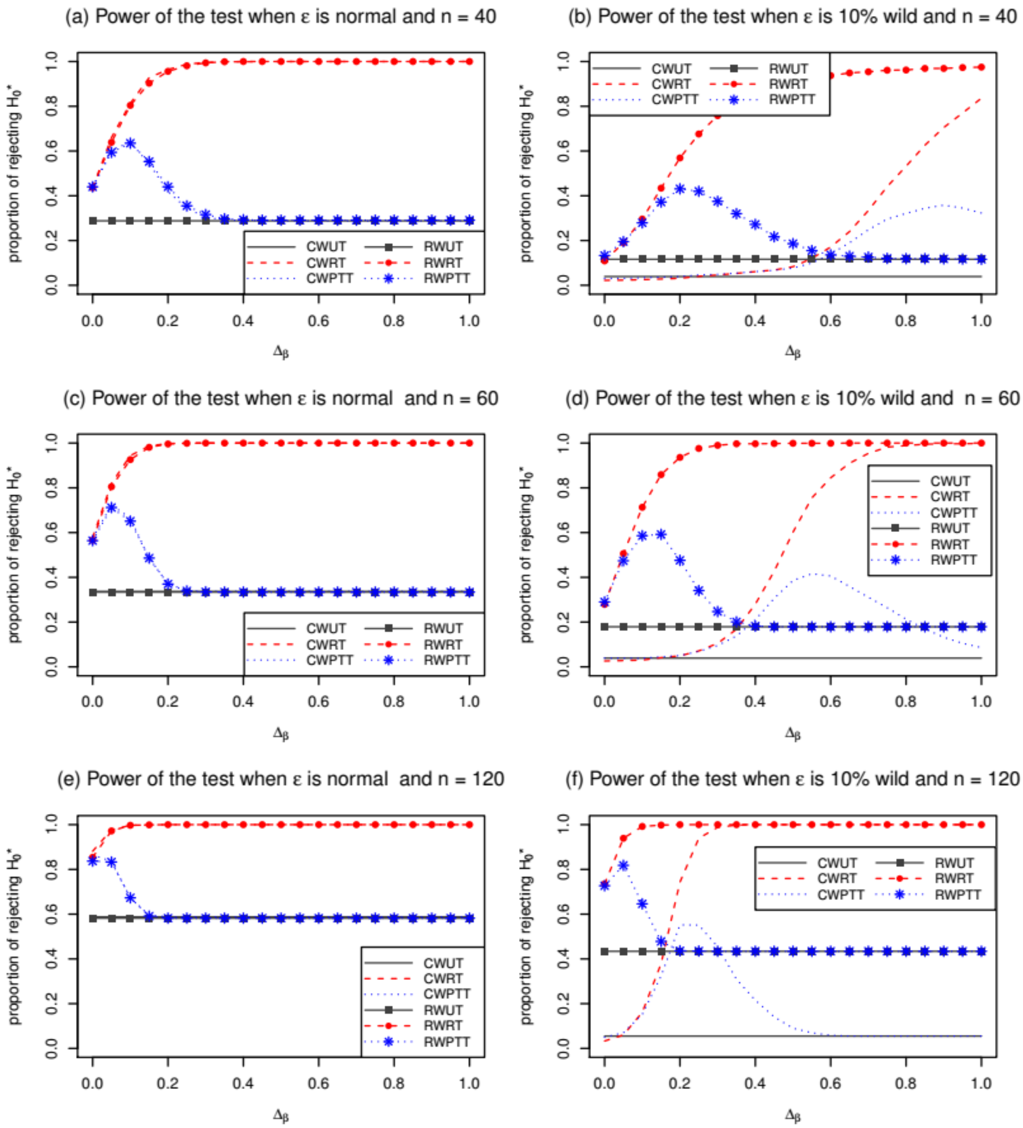

Figure 2: Power of the classical and robust Wald UT, RT and PTT for simulated data sized $n = 40, 60, 100$ with normal or 10% wild errors when $\delta_1 = 2$.

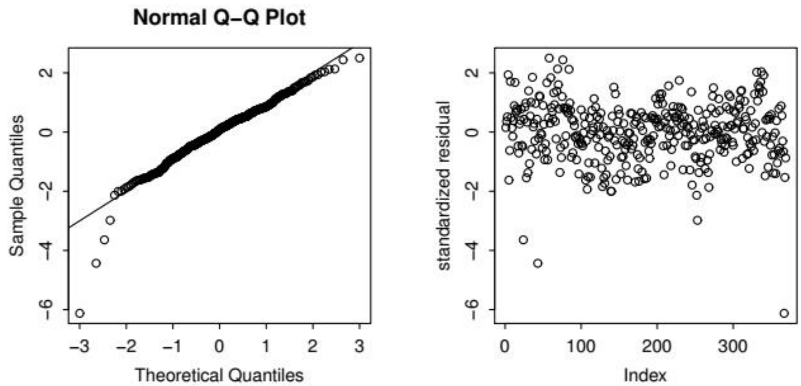

Figure 3: Diagnostics from weight-height-age linear relationship for n =399 Australian Olympic athletes.