

RobustBF: An R Package for Robust Solution to the Behrens-Fisher Problem

by Gamze Güven, Şükrii Acıtaş, Hatice Şamkar and Birdal Şenoğlu

Abstract Welch's two-sample t -test based on least squares (LS) estimators is generally used to test the equality of two normal means when the variances are not equal. However, this test loses its power when the underlying distribution is not normal. In this paper, two different tests are proposed to test the equality of two long-tailed symmetric (LTS) means under heterogeneous variances. Adaptive modified maximum likelihood (AMML) estimators are used in developing the proposed tests since they are highly efficient under LTS distribution. An R package called **RobustBF** is given to show the implementation of these tests. Simulated Type I error rates and powers of the proposed tests are also given and compared with the Welch's t -test based on LS estimators via an extensive Monte Carlo simulation study.

Introduction

Testing the equality of two population means is one of the most encountered problems in applied sciences. Student's t -test, which is uniformly most powerful unbiased, is commonly used under normality and homogeneity of variances assumptions. The well-known Behrens-Fisher (BF) problem arises when the assumption of homogeneity of variances is not met. This problem can be defined as testing the null hypothesis

$$H_0 : \mu_1 = \mu_2 \quad (1)$$

when $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$ ($i = 1, 2$) are independent random samples from $N(\mu_i, \sigma_i^2)$ distribution. Fisher (1939) endorsed Behrens' solution to BF problem by using the fiducial theory. Many researchers studied on this problem. For example, Welch (1938) proposed a test statistic and provided its degrees of freedom approximately. It should be noted that degrees of freedom provided by Welch (1938) can also be obtained by using Satterthwaite approximation; see Satterthwaite (1946). This is why the mentioned degrees of freedom is also known as Welch-Satterthwaite degrees of freedom in the literature. Wang (1971) calculated the Type I error rates of the Welch's two-sample t -test and Aspin-Welch test for different sets of degrees of freedom and nominal significance levels and concluded that Welch's t -test could be used in practice with little loss of accuracy. Davenport and Webster (1975) considered the test suggested by Fairfield Smith (1936) for the BF problem and compared its Type I error rates with those of Mehta and Srinivasan (1970). They concluded that this test is a very practical solution to the BF problem besides being stable in regard to size and having adequate power. Best and Rayner (1987) calculated Wald score and likelihood ratio statistics and showed that the test based on Wald statistics has the same asymptotic properties with the Welch's t -test. Kim and Cohen (1998) presented a review of basic concepts and applications concerning the BF problem under fiducial, Bayesian, and frequentist approaches. Singh et al. (2002) developed a test based on Jackknife estimator of the common population variance and compared the powers of the proposed test with those of Welch's t -test and Cochran and Cox (1957) test. According to results of their study, proposed test is more powerful than Cochran-Cox test for all cases, while it is more preferable than Welch's t -test for some cases. Chang and Pal (2008) developed a computational approach test (CAT) for the BF problem and compared it with Welch's t -test, Cochran-Cox test, Generalized p -value test and Singh-Saxena-Srivastava test under the normal and t -model. They found that Welch's t -test, Cochran-Cox test and CAT are robust under the heavier tailed t -models besides having similar size and power.

When the literature is examined, it can be seen that the Welch's t -test has a very good performance as compared to other tests in case of heteroscedasticity and unequal sample sizes in the context of normality. Power of the Welch's t -test decreases very rapidly when the underlying distribution is long-tailed symmetric (LTS) since the least squares (LS) estimators are not robust to the violation of normality. It is known that non-normal distributions are more common in real life problems. Yuen (1974) proposed a two-sample trimmed t -test and compared the performance of it with the Welch's t -test for both normal and long-tailed samples. Tiku and Singh (1981) proposed a Welch-type statistics based on modified maximum likelihood (MML) estimators and showed that the proposed test is more powerful than Yuen (1974)'s trimmed t -test. In addition, Tiku and Singh (1981) investigated the analogous test based on the robust bisquare estimators BS82 and showed that this test statistic gives misleading Type I errors.

In this study, a robust version of Welch's t -test for the BF problem is proposed when the underlying distribution is LTS. A second test using fiducial model, which is a special case of functional model given by Dawid and Stone (1982) is also proposed; see Fisher (1933, 1935) for more information about

the fiducial approach. The reason of including robust version of fiducial based test into this study is to see its performance in the context of BF problem and also to make comprehensive comparisons with its rivals (i.e., robust version of Welch's t -test and the traditional Welch's t -test). Both of the proposed tests are based on adaptive modified maximum likelihood (AMML) estimators, see [Tiku and Sürücü \(2009\)](#) and [Dönmez \(2010\)](#). To the best of our knowledge, this is the first study using AMML estimators for testing the equality of two LTS means under heterogeneous variances. These estimators are efficient and easy to compute for LTS samples, see [Tiku and Sürücü \(2009\)](#).

The R packages [stats](#) by R Core Team (1970) and [asht](#) by Fay (2020) include Welch's t -test based on LS estimators and BF test under normality, respectively. [WRS2](#) by Mair and Wilcox (2021) contains Yuen's test based on the trimmed sample means. Different from the earlier studies, we provide an R package **RobustBF** computing the values of the proposed test statistics and/or the corresponding p -values.

The rest of this study is organized as follows. Firstly, AMML estimators are given. Secondly, the robust Welch test and robust test based on the fiducial approach are developed. Thirdly, extensive Monte Carlo simulation study is conducted to compare the performances of the proposed tests with the traditional Welch's t -test based on LS estimators. The proposed tests are applied to a real data set via **RobustBF** package. This paper is finalized some concluding remarks.

AMML Estimators

Assume that $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$ ($i = 1, 2$) be independent random samples from $LTS(p, \mu_i, \sigma_i)$ distribution

$$f(y) = \frac{1}{\sqrt{k}\beta(1/2, p-1/2)\sigma} \left(1 + \frac{(y-\mu)^2}{k\sigma^2}\right)^{-p}, \quad -\infty < y < \infty; \quad -\infty < \mu < \infty; \quad \sigma > 0; \quad p \geq 2, \quad (2)$$

where μ is the location parameter, σ is the scale parameter, p is the shape parameter and $k = 2p - 3$ ([Tiku and Kumra, 1985](#)). It should be noted that $E(y) = \mu$, $V(y) = \sigma^2$ and $t = \sqrt{(v/k)}(y/\sigma)$ has Student's t distribution with $v = 2p - 1$ degrees of freedom.

The log-likelihood ($\ln L$) function is given by

$$\ln L = -N \ln(\sqrt{k}\beta(1/2, p-1/2)) - \sum_{i=1}^2 n_i \ln(\sigma_i) - p \sum_{i=1}^2 \sum_{j=1}^{n_i} \ln \left(1 + \frac{(y_{ij} - \mu_i)^2}{k\sigma_i^2}\right) \quad (3)$$

where $N = n_1 + n_2$. Then the likelihood equations are obtained as follows

$$\frac{\partial \ln L}{\partial \mu_i} = \frac{2p}{k\sigma_i} \sum_{j=1}^{n_i} g(z_{ij}) = 0 \quad (4)$$

$$\frac{\partial \ln L}{\partial \sigma_i} = -\frac{n_i}{\sigma_i} + \frac{2p}{k\sigma_i} \sum_{j=1}^{n_i} z_{ij} g(z_{ij}) = 0 \quad (5)$$

where

$$g(z_{ij}) = \frac{z_{ij}}{1 + (1/k)z_{ij}^2} \quad \text{and} \quad z_{ij} = \frac{y_{ij} - \mu_i}{\sigma_i}. \quad (6)$$

By solving the above likelihood equations (4) and (5) simultaneously, the maximum likelihood (ML) estimators of the parameters μ_i and σ_i are obtained. However, these equations involve nonlinear functions of the parameters, and so ML estimators cannot be obtained explicitly. Hence, numerical methods can be used to solve these equations. Numerical methods may cause some convergence problems like non-convergence of iterations, convergence to wrong root or multiple roots ([Puthenpura and Sinha, 1986](#); [Vaughan, 1992](#)). MML methodology proposed by [Tiku \(1967, 1968\)](#) overcomes these mentioned problems by providing explicit solutions to likelihood equations. In MML methodology, firstly, the standardized statistics are ordered in ascending way, i.e., $z_{i(1)} \leq z_{i(2)} \leq \dots \leq z_{i(n_i)}$. Then likelihood equations in (4) and (5) are rewritten in terms of $z_{i(j)}$ and $g(z_{i(j)})$ ($i = 1, 2; j = 1, 2, \dots, n_i$) as shown in (7) and (8) since summation is invariant to ordering, i.e., $\sum_{j=1}^{n_i} z_{i(j)} = \sum_{j=1}^{n_i} z_{ij}$.

$$\frac{\partial \ln L}{\partial \mu_i} = \frac{2p}{k\sigma_i} \sum_{j=1}^{n_i} g(z_{i(j)}) = 0 \quad (7)$$

$$\frac{\partial \ln L}{\partial \sigma_i} = -\frac{n_i}{\sigma_i} + \frac{2p}{k\sigma_i} \sum_{j=1}^{n_i} z_{i(j)} g(z_{i(j)}) = 0. \quad (8)$$

Here, $z_{i(j)} = \frac{y_{i(j)} - \mu_i}{\sigma_i}$ and $g(z_{i(j)}) = \frac{z_{i(j)}}{1 + (1/k)z_{i(j)}^2}$. The nonlinear function $g(z_{i(j)})$ is linearized utilizing the first two terms of the Taylor series expansion around the expected values of the ordered statistics $E(z_{i(j)}) = t_{i(j)}$ as follows

$$g(z_{i(j)}) \cong \alpha_{ij} + \beta_{ij} z_{i(j)} \quad (9)$$

where

$$\alpha_{ij} = \frac{(2/k)t_{i(j)}^3}{\left(1 + (1/k)t_{i(j)}^2\right)^2} \quad \text{and} \quad \beta_{ij} = \frac{1 - (1/k)t_{i(j)}^2}{\left(1 + (1/k)t_{i(j)}^2\right)^2}. \quad (10)$$

Since $t_{i(j)}$ values cannot be obtained exactly, approximate values of $t_{i(j)}$ which do not affect the efficiencies of the resulting estimators are used

$$\int_{-\infty}^{t_{i(j)}} f(z) dz = \frac{j}{n_i + 1}, i = 1, 2; j = 1, 2, \dots, n_i. \quad (11)$$

Secondly, modified likelihood equations are obtained by inserting the approximation (9) into Eqs. (7) and (8)

$$\frac{\partial \ln L^*}{\partial \mu_i} = \frac{2p}{k\sigma_i} \sum_{j=1}^{n_i} (\alpha_{ij} + \beta_{ij} z_{i(j)}) = 0 \quad (12)$$

$$\frac{\partial \ln L^*}{\partial \sigma_i} = -\frac{n_i}{\sigma_i} + \frac{2p}{k\sigma_i} \sum_{j=1}^{n_i} z_{i(j)} (\alpha_{ij} + \beta_{ij} z_{i(j)}) = 0. \quad (13)$$

Finally, MML estimators of μ_i and σ_i are found by solving Eqs. (12) and (13). They are given as follows

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} \beta_{ij} y_{i(j)}}{m_i} \quad \text{and} \quad \hat{\sigma}_i = \frac{B_i + \sqrt{B_i^2 + 4n_i C_i}}{2\sqrt{n_i(n_i - 1)}} \quad (14)$$

where

$$B_i = \frac{2p}{k} \sum_{j=1}^{n_i} \alpha_{ij} (y_{i(j)} - \hat{\mu}_i), \quad C_i = \frac{2p}{k} \sum_{j=1}^{n_i} \beta_{ij} (y_{i(j)} - \hat{\mu}_i)^2 \quad \text{and} \quad m_i = \sum_{j=1}^{n_i} \beta_{ij} \quad (15)$$

see [Tiku and Suresh \(1992\)](#). The asymptotic properties of the MML estimators $\hat{\mu}_i$ and $\hat{\sigma}_i$ can be demonstrated with the help of the following theorems.

Theorem 1 $\hat{\mu}_i$ is the minimum variance bound (MVB) estimator and is asymptotically normally distributed with mean μ_i and variance σ_i^2 / M_i ($M_i = 2pm_i/k$).

Theorem 2 $(n_i - 1)\hat{\sigma}_i^2 / \sigma_i^2$ is distributed as chi-square (more accurately a multiple of chi-square) with $(n_i - 1)$ degrees of freedom.

For proofs of theorems see e.g. [Şenoğlu and Tiku \(2001\)](#); [Güven et al. \(2019\)](#).

MML estimators have the same asymptotic properties with the ML estimators and are as efficient as ML estimators even for small samples. They are easy to compute and robust to the outliers.

It should be noted that the shape parameter p is assumed to be known in MML methodology. However, in some real life applications, it may be possible to assume that the data comes from a certain type of distribution, namely LTS distribution, but there is no opportunity to specify the value of the shape parameter. Hence, [Tiku and Sürücü \(2009\)](#) proposed AMML methodology, which is a new version of MML methodology, see [Dönmez \(2010\)](#) and [Acıtaş et al. \(2020, 2021\)](#). This methodology relaxes the assumption of known shape parameter. AMML estimators are computed in two iterations. In the first iteration, initial t_{ij} values are calculated from the sample data as shown below

$$t_{ij} = (y_{ij} - T_{0i}) / S_{0i} \quad i = 1, 2; j = 1, \dots, n_i. \quad (16)$$

Here, T_{0i} and S_{0i} are the initial estimates of μ_i and σ_i and they are calculated as

$$T_{0i} = \text{med} \{y_{ij}\} \quad \text{and} \quad S_{0i} = 1.483 \text{med} \{|y_{ij} - T_{0i}|\} \quad i = 1, 2; j = 1, \dots, n_i, \quad (17)$$

respectively. Using the t_{ij} values in (16), α_{ij} and β_{ij} coefficients are calculated as follows

$$\alpha_{ij} = \frac{(1/k)t_{ij}}{1 + (1/k)t_{ij}^2} \quad \text{and} \quad \beta_{ij} = \frac{1}{1 + (1/k)t_{ij}^2}. \quad (18)$$

Then the AMML estimates of the parameters μ_i and σ_i are obtained using Eq. (14) and α_{ij} and β_{ij} values given in Eq. (18). To distinguish these estimates from the MML estimates, they are represented by $\hat{\mu}_{i(AMML)}$ and $\hat{\sigma}_{i(AMML)}$ in the rest of the paper. In the second iteration, t_{ij} values are revised as follows

$$t_{ij} = (y_{ij} - \hat{\mu}_{i(AMML)}) / \hat{\sigma}_{i(AMML)} \quad i = 1, 2; j = 1, \dots, n_i \quad (19)$$

and recalculate the α_{ij} and β_{ij} values using the equalities in (18) for these t_{ij} values. Then final AMML estimates of μ_i and σ_i are obtained.

It should be noted that in AMML methodology y_{ij} observations are used rather than the ordered $y_{i(j)}$ observations since t_{ij} values are calculated from the sample observations. In addition, the shape parameter p is taken to be 16.5 in the calculations of α_{ij} and β_{ij} coefficients since this value makes AMML estimators efficient for normal and near normal distributions. It also makes them robust to mild outliers. The reason of why we use AMML methodology in the proposed tests is that it provides the same asymptotic properties with MML methodology and as mentioned before, relaxes the assumption of known shape parameter p .

Proposed Test Statistics

In this section, we propose two different tests for testing the equality of two LTS means.

Robust Welch (RW) Test

In this subsection, we briefly introduce Welch's t -test proposed by Welch (1938) under normal theory and then give robust version of it. Welch's t -test based on LS estimators is defined as $W = \{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\} / \sqrt{\{(s_1^2/n_1) + (s_2^2/n_2)\}}$. It is known that W is approximately distributed as Student's t with degrees of freedom

$$f = \frac{1}{\{c^2/(n_1 - 1) + (1 - c^2)/(n_2 - 1)\}} \quad (20)$$

where $c = (s_1^2/n_1) / \{(s_1^2/n_1) + (s_2^2/n_2)\}$. Here, \bar{x}_i and s_i^2 ($i = 1, 2$) are the sample means and sample variances, respectively. The value of W test can be obtained using `t.test` function available in R.

In this study, we propose the following test statistics based on AMML estimators as a robust alternative to Welch's t -test

$$RW = \frac{(\hat{\mu}_{1(AMML)} - \hat{\mu}_{2(AMML)}) - (\mu_1 - \mu_2)}{\sqrt{(\hat{\sigma}_{1(AMML)}^2/M_1) + (\hat{\sigma}_{2(AMML)}^2/M_2)}}. \quad (21)$$

As we shall see at the end of this section, the null distribution of RW is approximately distributed as Student's t based upon Theorems 1 and 2. The approximate degrees of freedom for this test is obtained using the Satterthwaite (1946) approximation as follows.

Let

$$c_1 = \frac{\sigma_1^2}{(n_1 - 1)M_1}, \quad c_2 = \frac{\sigma_2^2}{(n_2 - 1)M_2} \quad (22)$$

$$Q_1 = \frac{(n_1 - 1)\hat{\sigma}_{1(AMML)}^2}{\sigma_1^2} \quad \text{and} \quad Q_2 = \frac{(n_2 - 1)\hat{\sigma}_{2(AMML)}^2}{\sigma_2^2} \quad (23)$$

where Q_1 and Q_2 are independent chi-square random variables with degrees of freedom $(n_1 - 1)$ and $(n_2 - 1)$, respectively (see Theorem 2). If the linear combination of Q_1 and Q_2 is written as

$$Q = c_1 Q_1 + c_2 Q_2 = \frac{\hat{\sigma}_{1(AMML)}^2}{M_1} + \frac{\hat{\sigma}_{2(AMML)}^2}{M_2} \quad (24)$$

then $\nu Q/E(Q)$ has an approximate χ^2 distribution with the following degrees of freedom

$$\begin{aligned} \nu &= \frac{[c_1 Q_1 + c_2 Q_2]^2}{\left([c_1 Q_1]^2/\nu_1\right) + \left([c_2 Q_2]^2/\nu_2\right)} \\ &= \frac{\left((\hat{\sigma}_{1(AMML)}^2/M_1) + (\hat{\sigma}_{2(AMML)}^2/M_2)\right)^2}{(\hat{\sigma}_{1(AMML)}^2/M_1)^2/(n_1 - 1) + (\hat{\sigma}_{2(AMML)}^2/M_2)^2/(n_2 - 1)}. \end{aligned} \quad (25)$$

Here,

$$\nu_1 = n_1 - 1, \quad \nu_2 = n_2 - 1 \quad \text{and} \quad E(Q) = \frac{\sigma_1^2}{M_1} + \frac{\sigma_2^2}{M_2}. \quad (26)$$

RW in (21) can be rewritten as follows

$$RW = \frac{\left(\hat{\mu}_{1(AMML)} - \hat{\mu}_{2(AMML)} - (\mu_1 - \mu_2)\right) / \sqrt{(\sigma_1^2/M_1) + (\sigma_2^2/M_2)}}{\sqrt{(\hat{\sigma}_{1(AMML)}^2/M_1) + (\hat{\sigma}_{2(AMML)}^2/M_2)} / \sqrt{(\sigma_1^2/M_1) + (\sigma_2^2/M_2)}}. \quad (27)$$

Since this expression is equivalent to

$$\frac{Z}{\sqrt{Q}/\sqrt{E(Q)}} \quad (28)$$

it is obvious that RW is approximately distributed as Student's t with ν degrees of freedom. Here, $Z \sim N(0, 1)$ (see Theorem 1) and $\sqrt{Q}/\sqrt{E(Q)} \sim \sqrt{\chi_\nu^2/\nu}$.

To verify the null distribution of the RW, the probabilities

$$p_1 = \Pr(|RW| \geq t_{1-\alpha/2, \nu}) \quad (29)$$

are simulated from 10,000 Monte Carlo runs for various combinations of the sample sizes n_1 and n_2 . The results are demonstrated in Table 0. Here, ν is the degrees of freedom for RW.

This table number should refer to the table, i.e., clickable/cross referenced.

Robust Fiducial (RF) Based Test

In this section, fiducial-based test is proposed using the concept of fiducial inference and pivotal model see Fisher (1933, 1935) and Dawid and Stone (1982). Let denote the RW test based on the observed values as

$$RW^* = \frac{(\hat{\mu}_{1(AMML)}^* - \hat{\mu}_{2(AMML)}^*) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{\sigma}_{1(AMML)}^{2*}}{M_1} + \frac{\hat{\sigma}_{2(AMML)}^{2*}}{M_2}}}. \quad (30)$$

First the fiducial distribution of RW^* is derived by using pivotal quantities and fiducial distribution of the parameters of interest. Then the corresponding p -value is obtained. Here, $(\hat{\mu}_{i(AMML)}^*, \hat{\sigma}_{i(AMML)}^{2*})$ are the observed values of $(\hat{\mu}_{i(AMML)}, \hat{\sigma}_{i(AMML)}^2)$ ($i = 1, 2$). Let

$$Z_i = \frac{\hat{\mu}_{i(AMML)} - \mu_i}{\sigma_i / \sqrt{M_i}} \quad (31)$$

and

$$Q_i = \frac{(n_i - 1)\hat{\sigma}_{i(AMML)}^2}{\sigma_i^2} \quad (32)$$

are mutually independent pivotal quantities. They have asymptotically $N(0, 1)$ and $\chi_{(n_i-1)}^2$ distributions, respectively (see Theorems 1 and 2). Using pivotal quantities Z_i and Q_i , data generating equations are obtained as given below

$$\hat{\mu}_{i(AMML)} = \mu_i + (\sigma_i / \sqrt{M_i}) Z_i \quad (33)$$

and

$$\hat{\sigma}_{i(AMML)}^2 = \sigma_i^2 Q_i / (n_i - 1). \quad (34)$$

Given $(\hat{\mu}_{i(AMML)}^*, \hat{\sigma}_{i(AMML)}^{2*})$, Eqs. (33) and (34) are expressed as follows

$$\hat{\mu}_{i(AMML)}^* = \mu_i + (\sigma_i / \sqrt{M_i}) z_i \quad (35)$$

and

$$\hat{\sigma}_{i(AMML)}^{2*} = \sigma_i^2 q_i / (n_i - 1). \quad (36)$$

Here, (z_i, q_i) are the observed values of (Z_i, Q_i) . Eqs. (35) and (36) have the unique solutions as given below

$$\mu_i = \hat{\mu}_{i(AMML)}^* - \frac{z_i}{\sqrt{q_i/(n_i - 1)}} \frac{\hat{\sigma}_{i(AMML)}^*}{\sqrt{M_i}} \quad (37)$$

and

$$\sigma_i^2 = \frac{(n_i - 1) \hat{\sigma}_{i(AMML)}^{2*}}{q_i}. \quad (38)$$

Since $\frac{Z_i}{\sqrt{Q_i/(n_i-1)}}$ is distributed as a t_i variable with $(n_i - 1)$ degrees of freedom, the fiducial distribution of μ_i is the same as that of

$$T_{\mu_i}^* = \hat{\mu}_{i(AMML)}^* - \frac{t_i \hat{\sigma}_{i(AMML)}^*}{\sqrt{M_i}} \quad (39)$$

for given $(\hat{\mu}_{i(AMML)}^*, \hat{\sigma}_{i(AMML)}^{2*})$. Therefore, the fiducial distribution of RW^* in (30) is derived by utilizing the fiducial distribution of μ_i as follows

$$T_{RF} = \frac{\left((t_1 \hat{\sigma}_{1(AMML)}^*) / \sqrt{M_1} \right) - \left((t_2 \hat{\sigma}_{2(AMML)}^*) / \sqrt{M_2} \right)}{\sqrt{(\hat{\sigma}_{1(AMML)}^{2*}) / M_1 + (\hat{\sigma}_{2(AMML)}^{2*}) / M_2}} \quad (40)$$

where $t_1 \sim t_{(n_1-1)}$ and $t_2 \sim t_{(n_2-1)}$. Since

$$RW_0^* = \frac{(\hat{\mu}_{1(AMML)}^* - \hat{\mu}_{2(AMML)}^*)}{\sqrt{(\hat{\sigma}_{1(AMML)}^{2*}) / M_1 + (\hat{\sigma}_{2(AMML)}^{2*}) / M_2}} \quad (41)$$

is the observed value of T_{RF} under $H_0 : \mu_1 = \mu_2$, the corresponding p -value is given by

$$p = \Pr(T_{RF} \geq RW_0^*). \quad (42)$$

An algorithm for calculating the fiducial p -value in Eq.(42) via Monte Carlo simulation study is given as follows

Algorithm 1

- Step 1** For the given data, compute $\hat{\mu}_{i(AMML)}^*, \hat{\sigma}_{i(AMML)}^{2*}$ ($i = 1, 2$) and then RW_0^* utilizing Eq. (41).
- Step 2** Generate $t_i \sim t_{(n_i - 1)}$, ($i = 1, 2$).
- Step 3** Compute T_{RF}^2 utilizing Eq. (40).
- Step 4** Let $F_l = 1$ if $T_{RF}^2 > RW_0^{2*}$, else $F_l = 0$
- Step 5** Repeat the steps 2-4 K times.
- Step 6** Compute the simulated p -value using $p = \frac{1}{K} \sum_{j=1}^K F_j$.

It should be noted that the squares of T_{RF} and RW_0^* in Steps 3 and 4 are taken since the alternative hypothesis is two-sided, i.e., $H_1 : \mu_1 - \mu_2 \neq 0$ see Li et al. (2011)

Monte Carlo Simulation

In this section, Type I error rates and powers of the proposed tests (RW and RF) are compared with those of W test under the specified nominal level $\alpha = 0.05$. The plan of the simulation study is outlined as follows:

We use the following population distributions while generating samples.

Population 1	Population 2
(a) $Cauchy(0, 1)$	$Cauchy(0, 1)$
(b) $5 \times Cauchy(0, 1)$	$Cauchy(0, 1)$
(c) $Normal(0, 3^2) / Uniform(0, 1)$	$Normal(0, 1) / Uniform(0, 1)$
(d) $0.8Normal(0, 4^2) + 0.2 \frac{Normal(0, 4^2)}{Uniform[0, 1]}$	$0.8Normal(0, 1) + 0.2 \frac{Normal(0, 1)}{Uniform[0, 1]}$
(e) $3t_2$	t_2
(f) $2t_5$	t_5
(g) $Logistic(0, 3)$	$Logistic(0, 1)$
(h) $Laplace(0, 1)$	$Logistic(0, \sqrt{6})$

Here, t_a : Student's t distribution with a degrees of freedom.

10,000 different samples are considered for each of size $n_i (i = 1, 2)$. Sample sizes are taken as $(n_1, n_2) = (6, 6), (6, 10), (10, 10), (10, 15), (10, 30), (20, 20), (20, 30), (20, 50), (30, 50)$ and $(50, 50)$ while comparing the Type I error rates and powers of the tests. Simulations are conducted in R software.

To compute the Type I error rates of the RW , RF and W tests, firstly, samples are generated under the null hypothesis $H_0: \mu_1 = \mu_2$ for given (n_1, n_2) . Then AMML and LS estimates of the parameters are calculated. The probability in Eq. (29) gives the Type I rates of the RW test. It should be noted that this probability shows that how close the distribution of RW test is to Student's t with degrees of freedom ν . RF is carried out using Algorithm 1 with $K=5,000$. The fiducial p -value for the RF is computed in the final step of the mentioned algorithm. This procedure is repeated for each of the 10,000 samples. The proportion of the 10,000 p -values that are less than the nominal level $\alpha = 0.05$ gives Type I error rates of the RF .

To compute the power of the tests, similar steps are followed but a constant d is added to the observations in the first population. Any test can be considered powerful if it achieves maximum power and adheres to the prescribed significance level.

Results

The results of the Monte Carlo simulation study are given in Tables 1-9. The Type I error rates and power of the tests are given in Table 1 and Tables 2-9, respectively.

Numerical results of Table 1 can be summarized for Models (a)-(h) as follows

- Models (a)-(c): Type I error rates of the RW , RF and W tests are smaller than the nominal level of $\alpha = 0.05$. However, Type I error rates of the RW test are much closer to the nominal level than those of RF and W . RW test is followed by RF . W test is conservative regardless of the sample sizes.
- Models (d)-(e): Type I error rates of the RW test are closest to the nominal level of $\alpha = 0.05$ for all sample sizes even for small ones. RW test is followed by RF . W test is conservative compared to the RW and RF .
- Models (f)-(h): Type I error rates of the RW , RF and W tests are close to the nominal level of $\alpha = 0.05$. However, RW has very good performance in terms of Type I error rates even for small samples. Type I error rates of the RF test are slightly smaller than the nominal level for small sample sizes in Model (h).

The Numerical results of Tables 2-9 can be summarized as follows. It should be noted that the first line of the Tables 2-9, that is $d = 0.00$ presents simulated Type I error rates of the tests.

- Model (a): RW test appears to be more powerful than the RF for small to moderate sample sizes. However, as sample sizes increase, powers of RW and RF tests get closer to each other. These two tests outperform the W test for all sample sizes. The power of the W test is decreasing with increasing sample sizes.
- Models (b)-(e): RW and RF tests exhibit similar power properties and they have the most power for all sample sizes. W test has the least power. However, W test shows the worst performance in Model (b) when sample sizes are large and equal, i.e., $(n_1, n_2) = (50, 50)$.
- Models (f)-(g): RW , RF and W tests have similar power properties for small to moderate sample sizes. However, as sample sizes increase, RW and RF tests exhibit better performance than W test.

- Model (h): ^{the} ~~the~~ RW test has the most power followed by ^{the} ~~the~~ RF test has the least power especially for moderate to large sample sizes.

Overall, ^{the} ~~the~~ RW test can be recommended for testing the equality of two LTS means under the assumption of heterogeneous variances since it has the best performance with respect to ^{the} ~~the~~ size and power. Although the performance of RF test is not as good as RW test, it has better performance than the traditional Welch's ^{the} ~~the~~ ^{italic} ~~test~~ test.

Using RobustBF package

In the **RobustBF** package, we show the implementation of the proposed tests (RW and RF) based on AMML estimators and W test based on LS estimators using the data representing the values of $10(y - 2.0)$ (y is the pollution level (measurement of lead) in water samples from two lakes). It has been shown that long-tailed symmetric distribution provides plausible model for the mentioned data, see [Tiku and Akkaya \(2004\)](#) and also reference therein.

To run **RobustBF** package, we first install the package and then load it by typing:

```
> install.packages("RobustBF")
> library(RobustBF)
```

respectively. Next, the pollution level data are inputted for each lakes (Lake 1 and Lake 2) in terms of the vectors as shown below

```
y1 <- c(-1.48, 1.25, -0.51, 0.46, 0.60, -4.27, 0.63, -0.14, -0.38, 1.28,
        0.93, 0.51, 1.11, -0.17, -0.79, -1.02, -0.91, 0.10, 0.41, 1.11)
y2 <- c(1.32, 1.81, -0.54, 2.68, 2.27, 2.70, 0.78, -4.62, 1.88, 0.86,
        2.86, 0.47, -0.42, 0.16, 0.69, 0.78, 1.72, 1.57, 2.14, 1.62)
```

The value of the RW test, its degrees of freedom with the corresponding p -value and AMML estimates of the location parameters ($\hat{\mu}_{1(AMML)}$, $\hat{\mu}_{2(AMML)}$) and AMML estimates of the scale parameters ($\hat{\sigma}_{1(AMML)}$, $\hat{\sigma}_{2(AMML)}$) are given by using the function

```
> RW(y1,y2)
```

The p -value and AMML estimates of the location and scale parameters are given for the RF test by using the function

```
> RF(y1,y2,iter=5000)
```

It should be noted that the p -value for the RF test is obtained using computational approach and it is based on the replication number in Algorithm 1, denoted as `iter` in the RF function. When the above mentioned functions in the **RobustBF** package are performed, following results are obtained

```
> RW(y1,y2)
```

Robust Welch's Two Sample t-Test

```
data: y1 and y2
RW = -3.1602, df = 36.892, p-value = 0.0031
alternative hypothesis: true difference between in means is not equal to 0
sample estimates:
  mean of y1  mean of y2   sd of y1   sd of y2
    0.0626    1.2391    1.0861    1.2876

> RF(y1,y2,iter=5000)
```

Robust Fiducial Based Test

```
data: y1 and y2
p-value = 0.0032
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of y1 mean of y2   sd of y1   sd of y2
  0.0626    1.2391    1.0861    1.2876
```


We also use `t.test` function in R to test the null hypothesis $H_0: \mu_1 = \mu_2$ and obtain its p -value as 0.0243. It can be seen from these results, RW , RF , and W tests reject the null hypothesis at $\alpha = 0.05$ significance level, since the p -values corresponding to these tests are all less than 0.05. However, p -values for RW and RF tests are much smaller than the one obtained for W . Results of the RW and RF tests are more reliable since the AMML estimates of the σ_1 and σ_2 ($\hat{\sigma}_{1(AMML)} = 1.0861$, $\hat{\sigma}_{2(AMML)} = 1.2876$) are less than the corresponding LS estimates ($\hat{\sigma}_{1(LS)} = 1.2819$, $\hat{\sigma}_{2(LS)} = 1.6542$). It should be noted that RW and RF tests reject the null hypothesis while W fails to reject it at the significance level $\alpha = 0.01$. These results are in agreement with the simulation results in the context of long-tailed symmetric distributions.

Conclusion

Reviewing the literature shows that comparing two means is a commonly encountered problem, especially in applied sciences when the usual normality and homogeneity of variances assumptions are violated. For this reason, in this study, we present **RobustBF** package and propose RW and RF tests to test the equality of two LTS means when the variances are unknown and arbitrary. The first test included in the package is a robust version of Welch's t -test, and the other one is a robust fiducial-based test. The proposed tests are based on AMML estimators. Also, we use `t.test` function available in R to compare the proposed tests with the Welch's t -test in terms of Type I error rates and powers. Examining the results of the simulation study reveals that Type I error rates of the RW test are closer to the nominal level in general. Therefore, the RW test verifies the obtained null distribution for long-tailed symmetric samples. This test is followed by RF . RF does not require the knowledge of sampling distribution of the test statistics. W test appears to be conservative except for the t_5 , Logistic and Laplace distributions. RW shows the best power performance among the others besides being robust for the contamination model for the scenarios considered in this study. Therefore, the proposed RW test can be recommended for testing the equality of two LTS means under heterogeneity of variances. W test performs poorly in almost all cases. According to our knowledge, the proposed tests presented in the **RobustBF** package are not available in any other R tool.

Bibliography

- S. Acıtaş, P. Filzmoser, and B. Şenoğlu. A new partial robust adaptive modified maximum likelihood estimator. *Chemometrics and Intelligent Laboratory Systems*, 204:104068, 2020. URL <https://doi.org/10.1016/j.chemolab.2020.104068>. [p3]
- S. Acıtaş, P. Filzmoser, and B. Şenoğlu. A robust adaptive modified maximum likelihood estimator for the linear regression model. *Journal of Statistical Computation and Simulation*, 91(7):1394–1414, 2021. URL <https://doi.org/10.1080/00949655.2020.1856847>. [p3]
- D. J. Best and J. C. W. Rayner. Welch's approximate solution for the behrens – fisher problem. *Technometrics*, 29(2):205–210, 1987. URL <https://doi.org/10.1080/00401706.1987.10488211>. [p1]
- C. H. Chang and N. Pal. A revisit to the behrens–fisher problem: comparison of five test methods. *Communications in Statistics—Simulation and Computation*, 37(6):1064–1085, 2008. URL <https://doi.org/10.1080/03610910802049599>. [p1]
- W. G. Cochran and G. M. Cox. *Experimental designs*. John Wiley and Sons, New York, 1957. [p1]
- J. M. Davenport and J. T. Webster. The behrens-fisher problem, an old solution revisited. *Metrika*, 22: 47–54, 1975. URL <https://doi.org/10.1007/BF01899713>. [p1]
- A. P. Dawid and M. Stone. The functional-model basis of fiducial inference. *The Annals of Statistics*, 10 (4):1054–1067, 1982. URL <https://www.jstor.org/stable/2240708>. [p1, 5]
- A. Dönmez. *Adaptive estimation and hypothesis testing methods [dissertation]*. Ankara:METU, 2010. [p2, 3]
- H. Fairfield Smith. The problem of comparing the result of two experiments with unequal errors. *Journal Council for Scientific and Industrial Research Australia*, 9:211–212, 1936. URL <https://publications.csiro.au/publications/publication/PIprocite:a35f0b77-57ff-496e-869f-10ba39834723>. [p1]
- M. P. Fay. *asht: Applied Statistical Hypothesis Tests*, 2020. URL <https://cran.r-project.org/web/packages/asht/index.html>. R package version 0.9.6. [p2]

- R. A. Fisher. The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London. Series A*, 139(838):343–348, 1933. URL <https://doi.org/10.1098/rspa.1933.0021>. [p1, 5]
- R. A. Fisher. The fiducial argument in statistical inference. *Annals of eugenics*, 6(4):391–398, 1935. URL <https://doi.org/10.1111/j.1469-1809.1935.tb02120.x>. [p1, 5]
- R. A. Fisher. The comparison of samples with possibly unequal variances. *Annals of Eugenics*, 9(2): 174–180, 1939. URL <https://doi.org/10.1111/j.1469-1809.1939.tb02205.x>. [p1]
- G. Güven, O. Güreş, H. Şamkar, and B. Şenoğlu. A fiducial-based approach to the one-way anova in the presence of nonnormality and heterogeneous error variances. *Journal of Statistical Computation and Simulation*, 89(9):1715–1729, 2019. URL <https://doi.org/10.1080/00949655.2019.1593985>. [p3]
- S. H. Kim and A. S. Cohen. On the behrens-fisher problem: A review. *Journal of Educational and Behavioral Statistics*, 23(4):356–377, 1998. URL <https://doi.org/10.3102/10769986023004356>. [p1]
- X. Li, J. Wang, and H. Liang. Comparison of several means: A fiducial based approach. *Computational statistics & data analysis*, 55(5):1993–2002, 2011. URL <https://doi.org/10.1016/j.csda.2010.12.009>. [p6]
- P. Mair and R. Wilcox. *WRS2: A Collection of Robust Statistical Methods*, 2021. URL <https://cran.r-project.org/web/packages/WRS2/index.html>. R package version 1.1–1. [p2]
- J. S. Mehta and R. Srinivasan. On the behrens-fisher problem. *Biometrika*, 57(3):649–655, 1970. URL <https://doi.org/10.1093/biomet/57.3.649>. [p1]
- S. Puthenpura and N. K. Sinha. Modified maximum likelihood method for the robust estimation of system parameters from very noisy data. *Automatica*, 22(2):231–235, 1986. URL [https://doi.org/10.1016/0005-1098\(86\)90085-3](https://doi.org/10.1016/0005-1098(86)90085-3). [p2]
- R Core Team. *The R Stats Package*, 1970. URL <https://www.rdocumentation.org/packages/stats/versions/3.6.2>. R package version 3.6.2. [p2]
- F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114, 1946. URL <https://doi.org/10.2307/3002019>. [p1, 4]
- B. Şenoğlu and M. L. Tiku. Analysis of variance in experimental design with nonnormal error distributions. *Communications in Statistics-Theory and Methods*, 30(7):1335–1352, 2001. URL <https://doi.org/10.1081/STA-100104748>. [p3]
- P. Singh, K. K. Saxena, and O. P. Srivastava. Power comparisons of solutions to the behrens-fisher problem. *American Journal of Mathematical and Management Sciences*, 22(3):233–250, 2002. URL <https://doi.org/10.1080/01966324.2002.10737589>. [p1]
- M. L. Tiku. Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, 54:155–165, 1967. URL <https://doi.org/10.1093/biomet/54.1-2.155>. [p2]
- M. L. Tiku. Estimating the parameters of log-normal distribution from censored samples. *Journal of the American Statistical Association*, 63(321):134–140, 1968. URL <https://doi.org/10.2307/2283834>. [p2]
- M. L. Tiku and A. D. Akkaya. *Robust estimation and hypothesis testing*. New Age International Limited Publishers, New Delhi, 2004. [p8]
- M. L. Tiku and S. Kumra. Expected values and variances and covariances of order statistics for a family of symmetric distributions (student’s t). *Selected tables in mathematical statistics*, 8:141–270, 1985. [p2]
- M. L. Tiku and M. Singh. Robust test for means when population variances are unequal. *Communications in Statistics—Theory and Methods*, 10(20):2057–2071, 1981. URL <https://doi.org/10.1080/03610928108828173>. [p1]
- M. L. Tiku and R. P. Suresh. A new method of estimation for location and scale parameters. *Journal of Statistical Planning and Inference*, 30(2):281–292, 1992. URL [https://doi.org/10.1016/0378-3758\(92\)90088-A](https://doi.org/10.1016/0378-3758(92)90088-A). [p3]
- M. L. Tiku and B. Sürücü. Mmles are as good as m-estimators or better. *Statistics & probability letters*, 79(7):984–989, 2009. URL <https://doi.org/10.1016/j.spl.2008.12.001>. [p2, 3]

- D. C. Vaughan. On the tiku-suresh method of estimation. *Communications in Statistics-Theory and Methods*, 21(2):451–469, 1992. URL <https://doi.org/10.1080/03610929208830788>. [p2]
- Y. Y. Wang. Probabilities of the type i errors of the welch tests for the behrens-fisher problem. *Journal of the American Statistical Association*, 66(335):605–608, 1971. URL <https://doi.org/10.2307/2283538>. [p1]
- B. L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29:350–362, 1938. URL <https://doi.org/10.2307/2332010>. [p1, 4]
- K. K. Yuen. The two-sample trimmed t for unequal population variances. *Biometrika*, 61(1):165–170, 1974. URL <https://doi.org/10.1093/biomet/61.1.165>. [p1]

Gamze Güven
Eskisehir Osmangazi University
Department of Statistics
Eskisehir, Turkey
<https://orcid.org/0000-0002-8821-3179>
gamzeguven@ogu.edu.tr

Şükriü Acıtaş
Eskisehir Technical University
Department of Statistics
Eskisehir, Turkey
<https://orcid.org/0000-0002-4131-0086>
sacitas@eskisehir.edu.tr

Hatice Şamkar
Eskisehir Osmangazi University
Department of Statistics
Eskisehir, Turkey
<https://orcid.org/0000-0003-1400-2392>
hfidan@ogu.edu.tr

Birdal Şenoğlu
Ankara University
Department of Statistics
Ankara, Turkey
<https://orcid.org/0000-0003-3707-2393>
senoglu@science.ankara.edu.tr

Table 1: Simulated Type I error rates of the *RW*, *RF* and *W* tests for Models (a)-(h)

Model (a)					Model (b)		
<i>n</i> ₁	<i>n</i> ₂	<i>RW</i>	<i>RF</i>	<i>W</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
6	6	0.023	0.014	0.015	0.031	0.025	0.019
6	10	0.026	0.016	0.017	0.033	0.027	0.020
10	10	0.025	0.020	0.018	0.031	0.027	0.020
10	15	0.023	0.016	0.017	0.029	0.027	0.020
10	30	0.030	0.025	0.020	0.029	0.028	0.020
20	20	0.024	0.022	0.020	0.028	0.026	0.021
20	30	0.028	0.025	0.024	0.027	0.026	0.022
20	50	0.026	0.024	0.021	0.028	0.027	0.020
30	50	0.027	0.025	0.021	0.026	0.026	0.022
50	50	0.030	0.028	0.020	0.027	0.026	0.020

Model (c)					Model (d)		
<i>n</i> ₁	<i>n</i> ₂	<i>RW</i>	<i>RF</i>	<i>W</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
6	6	0.035	0.024	0.022	0.047	0.036	0.030
6	10	0.037	0.030	0.020	0.053	0.047	0.037
10	10	0.030	0.025	0.020	0.046	0.042	0.033
10	15	0.031	0.026	0.018	0.046	0.043	0.033
10	30	0.031	0.029	0.022	0.046	0.044	0.030
20	20	0.030	0.027	0.019	0.042	0.040	0.027
20	30	0.030	0.028	0.021	0.046	0.044	0.030
20	50	0.030	0.030	0.022	0.045	0.044	0.030
30	50	0.031	0.029	0.020	0.042	0.041	0.026
50	50	0.030	0.028	0.019	0.044	0.044	0.025

Model (e)					Model (f)		
<i>n</i> ₁	<i>n</i> ₂	<i>RW</i>	<i>RF</i>	<i>W</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
6	6	0.040	0.030	0.034	0.050	0.042	0.042
6	10	0.050	0.045	0.035	0.054	0.042	0.046
10	10	0.044	0.038	0.038	0.049	0.043	0.044
10	15	0.045	0.042	0.033	0.054	0.048	0.049
10	30	0.042	0.040	0.037	0.055	0.052	0.051
20	20	0.044	0.042	0.028	0.054	0.049	0.047
20	30	0.040	0.037	0.038	0.052	0.049	0.048
20	50	0.041	0.041	0.026	0.051	0.049	0.046
30	50	0.043	0.042	0.036	0.053	0.053	0.049
50	50	0.044	0.043	0.028	0.054	0.052	0.049

Model (g)					Model (h)		
<i>n</i> ₁	<i>n</i> ₂	<i>RW</i>	<i>RF</i>	<i>W</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
6	6	0.053	0.041	0.045	0.048	0.032	0.044
6	10	0.056	0.052	0.051	0.044	0.034	0.043
10	10	0.055	0.048	0.047	0.044	0.036	0.042
10	15	0.055	0.052	0.048	0.045	0.039	0.044
10	30	0.052	0.050	0.044	0.045	0.039	0.045
20	20	0.054	0.053	0.049	0.044	0.041	0.044
20	30	0.054	0.053	0.048	0.047	0.044	0.047
20	50	0.055	0.055	0.049	0.050	0.046	0.049
30	50	0.054	0.054	0.048	0.046	0.045	0.046
50	50	0.054	0.053	0.049	0.054	0.051	0.052

→ The table's caption should be placed below the table.

Table 2: Simulated powers of the *RW*, *RF* and *W* tests for Model (a)

	<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>		<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
<i>n</i> = (6, 6)	0.00	0.023	0.014	0.015	<i>n</i> = (20, 20)	0.00	0.024	0.022	0.020
	1.60	0.19	0.15	0.11		0.60	0.10	0.09	0.04
	3.20	0.51	0.46	0.30		1.20	0.35	0.33	0.09
	4.80	0.74	0.70	0.46		1.80	0.65	0.64	0.17
	6.40	0.84	0.82	0.56		2.40	0.84	0.83	0.25
	8.00	0.91	0.89	0.64		3.00	0.94	0.94	0.33
<i>n</i> = (6, 10)	0.00	0.026	0.016	0.017	<i>n</i> = (20, 30)	0.00	0.028	0.025	0.024
	1.50	0.21	0.17	0.10		0.50	0.09	0.08	0.03
	3.00	0.57	0.53	0.29		1.00	0.31	0.29	0.07
	4.50	0.78	0.76	0.44		1.50	0.59	0.58	0.13
	6.00	0.89	0.88	0.57		2.00	0.80	0.79	0.20
	7.50	0.94	0.93	0.62		2.50	0.92	0.91	0.27
<i>n</i> = (10, 10)	0.00	0.025	0.020	0.018	<i>n</i> = (20, 50)	0.00	0.026	0.024	0.021
	1.00	0.13	0.11	0.06		0.46	0.10	0.09	0.04
	2.00	0.43	0.39	0.18		0.92	0.32	0.30	0.06
	3.00	0.70	0.67	0.32		1.38	0.60	0.59	0.12
	4.00	0.85	0.83	0.42		1.84	0.81	0.81	0.18
	5.00	0.92	0.91	0.51		2.30	0.92	0.92	0.25
<i>n</i> = (10, 15)	0.00	0.023	0.016	0.017	<i>n</i> = (30, 50)	0.00	0.027	0.025	0.021
	0.80	0.11	0.09	0.05		0.40	0.09	0.09	0.03
	1.60	0.36	0.33	0.14		0.80	0.30	0.29	0.06
	2.40	0.64	0.61	0.24		1.20	0.60	0.59	0.10
	3.20	0.80	0.79	0.35		1.60	0.83	0.83	0.16
	4.00	0.90	0.89	0.43		2.00	0.94	0.94	0.22
<i>n</i> = (10, 30)	0.00	0.030	0.025	0.020	<i>n</i> = (50, 50)	0.00	0.030	0.028	0.020
	0.70	0.12	0.10	0.05		0.32	0.08	0.08	0.030
	1.40	0.37	0.35	0.11		0.64	0.26	0.26	0.05
	2.10	0.65	0.63	0.21		0.96	0.55	0.54	0.08
	2.80	0.80	0.79	0.30		1.28	0.79	0.78	0.11
	3.50	0.90	0.89	0.39		1.60	0.93	0.92	0.16

Place the caption here.

Table 3: Simulated powers of the *RW*, *RF* and *W* tests for Model (b)

	<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>		<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
<i>n</i> = (6, 6)	0.00	0.031	0.025	0.019	<i>n</i> = (20, 20)	0.00	0.028	0.026	0.021
	5.40	0.22	0.19	0.13		2.00	0.11	0.11	0.04
	10.80	0.52	0.50	0.34		4.00	0.34	0.34	0.11
	16.20	0.72	0.71	0.49		6.00	0.61	0.60	0.19
	21.60	0.83	0.83	0.60		8.00	0.80	0.79	0.30
	27.00	0.90	0.90	0.68		10.00	0.90	0.90	0.38
<i>n</i> = (6, 10)	0.00	0.033	0.027	0.020	<i>n</i> = (20, 30)	0.00	0.027	0.026	0.022
	5.30	0.22	0.21	0.13		2.00	0.10	0.10	0.05
	10.60	0.53	0.52	0.35		4.00	0.33	0.33	0.10
	15.90	0.72	0.71	0.50		6.00	0.62	0.62	0.20
	21.20	0.82	0.82	0.59		8.00	0.81	0.80	0.29
	26.50	0.90	0.90	0.68		10.00	0.90	0.90	0.37
<i>n</i> = (10, 10)	0.00	0.031	0.027	0.020	<i>n</i> = (20, 50)	0.00	0.028	0.027	0.20
	3.60	0.17	0.16	0.09		2.00	0.11	0.11	0.05
	7.20	0.47	0.46	0.23		4.00	0.35	0.34	0.11
	10.80	0.71	0.71	0.38		6.00	0.62	0.62	0.20
	14.40	0.84	0.84	0.49		8.00	0.80	0.80	0.29
	18.00	0.92	0.92	0.58		10.00	0.91	0.91	0.37
<i>n</i> = (10, 15)	0.00	0.029	0.027	0.020	<i>n</i> = (30, 50)	0.00	0.026	0.026	0.022
	3.60	0.17	0.16	0.09		1.60	0.10	0.10	0.04
	7.20	0.49	0.48	0.23		3.20	0.32	0.32	0.08
	10.80	0.72	0.72	0.39		4.80	0.62	0.62	0.15
	14.40	0.85	0.85	0.50		6.40	0.82	0.82	0.23
	18.00	0.92	0.91	0.57		8.00	0.92	0.92	0.30
<i>n</i> = (10, 30)	0.00	0.029	0.028	0.020	<i>n</i> = (50, 50)	0.00	0.027	0.026	0.020
	3.40	0.16	0.16	0.09		1.12	0.08	0.08	0.03
	6.80	0.45	0.45	0.22		2.24	0.26	0.26	0.05
	10.20	0.69	0.69	0.36		3.36	0.53	0.53	0.09
	13.60	0.83	0.83	0.47		4.48	0.76	0.76	0.13
	17.00	0.90	0.90	0.55		5.60	0.90	0.90	0.18

Place the caption here.

Table 4: Simulated powers of the RW , RF and W tests for Model (c)

	d	RW	RF	W		d	RW	RF	W
$n = (6, 6)$	0.00	0.035	0.024	0.022	$n = (20, 20)$	0.00	0.030	0.027	0.019
	5.00	0.23	0.20	0.14		1.70	0.10	0.10	0.04
	10.00	0.55	0.52	0.37		3.40	0.33	0.32	0.11
	15.00	0.76	0.75	0.53		5.10	0.60	0.59	0.19
	20.00	0.87	0.86	0.62		6.80	0.80	0.80	0.29
	25.00	0.92	0.92	0.70		8.50	0.92	0.91	0.38
$n = (6, 10)$	d	RW	RF	W	$n = (20, 30)$	d	RW	RF	W
	0.00	0.037	0.030	0.020		0.00	0.030	0.028	0.021
	4.80	0.23	0.21	0.14		1.70	0.10	0.10	0.04
	9.60	0.56	0.54	0.36		3.40	0.33	0.33	0.11
	14.40	0.75	0.74	0.52		5.10	0.62	0.61	0.20
	19.20	0.86	0.86	0.62		6.80	0.81	0.81	0.30
$n = (10, 10)$	24.00	0.92	0.92	0.69	$n = (20, 50)$	8.50	0.92	0.92	0.37
	d	RW	RF	W		d	RW	RF	W
	0.00	0.030	0.025	0.020		0.00	0.030	0.030	0.022
	3.00	0.15	0.14	0.08		1.70	0.11	0.11	0.05
	6.00	0.44	0.42	0.22		3.40	0.35	0.35	0.11
	9.00	0.71	0.69	0.37		5.10	0.63	0.63	0.20
$n = (10, 15)$	12.00	0.84	0.84	0.48	$n = (30, 50)$	6.80	0.82	0.82	0.30
	15.00	0.92	0.92	0.56		8.50	0.92	0.92	0.38
	d	RW	RF	W		d	RW	RF	W
	0.00	0.031	0.026	0.018		0.00	0.031	0.029	0.20
	2.60	0.13	0.12	0.06		1.30	0.09	0.09	0.03
	5.20	0.38	0.37	0.18		2.60	0.31	0.30	0.07
$n = (10, 30)$	7.80	0.64	0.63	0.32	$n = (50, 50)$	3.90	0.58	0.58	0.13
	10.40	0.79	0.79	0.44		5.20	0.79	0.79	0.20
	13.00	0.90	0.90	0.54		6.50	0.92	0.92	0.29
	d	RW	RF	W		d	RW	RF	W
	0.00	0.031	0.029	0.022		0.00	0.030	0.028	0.019
	2.60	0.14	0.13	0.07		0.96	0.08	0.08	0.03
	5.20	0.39	0.39	0.18		1.92	0.25	0.25	0.05
	7.80	0.64	0.64	0.32		2.88	0.54	0.53	0.09
	10.40	0.80	0.80	0.44		3.84	0.78	0.77	0.13
	13.00	0.90	0.90	0.54		4.80	0.91	0.91	0.20

Place the caption here.

Table 5: Simulated powers of the *RW*, *RF* and *W* tests for Model (d)

	<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>		<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
<i>n</i> = (6,6)	0.00	0.047	0.036	0.030	<i>n</i> = (20,20)	0.00	0.42	0.040	0.027
	2.00	0.15	0.13	0.11		0.80	0.11	0.10	0.07
	4.00	0.42	0.40	0.34		1.60	0.30	0.29	0.17
	6.00	0.69	0.67	0.57		2.40	0.57	0.56	0.32
	8.00	0.84	0.83	0.71		3.20	0.78	0.78	0.46
	10.00	0.91	0.91	0.78		4.00	0.91	0.91	0.57
<i>n</i> = (6,10)	0.00	0.053	0.047	0.037	<i>n</i> = (20,30)	0.00	0.046	0.044	0.030
	2.00	0.16	0.15	0.12		0.80	0.13	0.13	0.09
	4.00	0.43	0.42	0.34		1.60	0.37	0.36	0.25
	6.00	0.70	0.69	0.59		2.40	0.65	0.65	0.47
	8.00	0.84	0.84	0.72		3.20	0.82	0.82	0.62
	10.00	0.91	0.91	0.79		4.00	0.91	0.91	0.71
<i>n</i> = (10,10)	0.00	0.046	0.042	0.033	<i>n</i> = (20,50)	0.00	0.045	0.044	0.030
	1.30	0.12	0.12	0.09		0.80	0.11	0.11	0.06
	2.60	0.36	0.34	0.26		1.60	0.31	0.31	0.17
	3.90	0.63	0.62	0.46		2.40	0.57	0.56	0.32
	5.20	0.81	0.81	0.60		3.20	0.79	0.79	0.46
	6.50	0.92	0.92	0.71		4.00	0.92	0.92	0.58
<i>n</i> = (10,15)	0.00	0.046	0.043	0.033	<i>n</i> = (30,50)	0.00	0.042	0.041	0.026
	1.30	0.13	0.12	0.09		0.64	0.11	0.11	0.05
	2.60	0.37	0.36	0.26		1.28	0.30	0.30	0.14
	3.90	0.65	0.64	0.47		1.92	0.56	0.56	0.27
	5.20	0.83	0.82	0.62		2.56	0.79	0.79	0.39
	6.50	0.92	0.91	0.71		3.20	0.92	0.92	0.51
<i>n</i> = (10,30)	0.00	0.046	0.044	0.030	<i>n</i> = (50,50)	0.00	0.044	0.044	0.025
	1.30	0.13	0.13	0.09		0.48	0.10	0.10	0.05
	2.60	0.37	0.36	0.25		0.96	0.27	0.27	0.10
	3.90	0.65	0.65	0.47		1.44	0.53	0.53	0.19
	5.20	0.82	0.82	0.62		1.92	0.78	0.77	0.30
	6.50	0.91	0.91	0.71		2.40	0.92	0.92	0.39

→ Place the caption here.

Table 6: Simulated powers of the *RW*, *RF* and *W* tests for Model (e)

	<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>		<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
<i>n</i> = (6,6)	0.00	0.040	0.030	0.034	<i>n</i> = (20,20)	0.00	0.044	0.042	0.028
	2.00	0.16	0.13	0.13		0.80	0.11	0.11	0.07
	4.00	0.44	0.40	0.38		1.60	0.30	0.29	0.17
	6.00	0.70	0.67	0.62		2.40	0.55	0.55	0.31
	8.00	0.84	0.83	0.76		3.20	0.78	0.78	0.46
	10.00	0.92	0.91	0.85		4.00	0.91	0.91	0.57
<i>n</i> = (6,10)	0.00	0.050	0.045	0.035	<i>n</i> = (20,30)	0.00	0.040	0.037	0.038
	1.90	0.15	0.14	0.11		0.80	0.10	0.10	0.08
	3.80	0.41	0.40	0.33		1.60	0.31	0.30	0.23
	5.70	0.67	0.66	0.55		2.40	0.59	0.58	0.42
	7.60	0.83	0.83	0.71		3.20	0.81	0.81	0.60
	9.50	0.90	0.90	0.77		4.00	0.92	0.92	0.74
<i>n</i> = (10,10)	0.00	0.044	0.038	0.038	<i>n</i> = (20,50)	0.00	0.041	0.041	0.026
	1.26	0.12	0.10	0.10		0.80	0.11	0.11	0.06
	2.52	0.35	0.33	0.29		1.60	0.30	0.30	0.17
	3.78	0.62	0.60	0.50		2.40	0.58	0.58	0.32
	5.04	0.80	0.80	0.67		3.20	0.79	0.79	0.46
	6.30	0.91	0.91	0.78		4.00	0.91	0.91	0.57
<i>n</i> = (10,15)	0.00	0.045	0.042	0.033	<i>n</i> = (30,50)	0.00	0.043	0.042	0.041
	1.24	0.12	0.12	0.09		0.64	0.11	0.11	0.08
	2.48	0.33	0.33	0.23		1.28	0.31	0.31	0.21
	3.72	0.60	0.60	0.44		1.92	0.59	0.58	0.38
	4.96	0.80	0.80	0.59		2.56	0.81	0.81	0.56
	6.20	0.90	0.90	0.68		3.20	0.93	0.93	0.70
<i>n</i> = (10,30)	0.00	0.042	0.040	0.037	<i>n</i> = (50,50)	0.00	0.044	0.043	0.028
	1.24	0.13	0.13	0.11		0.48	0.10	0.10	0.05
	2.48	0.36	0.36	0.29		0.96	0.27	0.27	0.10
	3.72	0.63	0.63	0.52		1.44	0.55	0.55	0.20
	4.96	0.81	0.81	0.69		1.92	0.78	0.78	0.30
	6.20	0.91	0.91	0.80		2.40	0.92	0.92	0.41

Place the caption here.

Table 7: Simulated powers of the RW , RF and W tests for Model (f)

	d	RW	RF	W		d	RW	RF	W
$n = (6, 6)$	0.00	0.050	0.042	0.042	$n = (20, 20)$	0.00	0.054	0.049	0.047
	0.90	0.13	0.10	0.11		0.40	0.11	0.11	0.10
	1.80	0.33	0.29	0.30		0.80	0.27	0.26	0.24
	2.70	0.60	0.54	0.56		1.20	0.53	0.51	0.47
	3.60	0.80	0.77	0.77		1.60	0.76	0.75	0.69
	4.50	0.92	0.90	0.89		2.00	0.90	0.90	0.85
$n = (6, 10)$	d	RW	RF	W	$n = (20, 30)$	d	RW	RF	W
	0.00	0.054	0.042	0.046		0.00	0.052	0.049	0.048
	0.90	0.13	0.11	0.12		0.40	0.11	0.11	0.09
	1.80	0.37	0.34	0.34		0.80	0.29	0.29	0.25
	2.70	0.63	0.59	0.59		1.20	0.55	0.54	0.48
	3.60	0.83	0.81	0.80		1.60	0.77	0.77	0.71
$n = (10, 10)$	4.50	0.92	0.92	0.90	$n = (20, 50)$	2.00	0.92	0.92	0.87
	d	RW	RF	W		d	RW	RF	W
	0.00	0.049	0.043	0.044		0.00	0.051	0.049	0.046
	0.64	0.11	0.10	0.10		0.40	0.12	0.11	0.10
	1.28	0.32	0.29	0.29		0.80	0.30	0.30	0.27
	1.92	0.59	0.55	0.54		1.20	0.57	0.57	0.51
$n = (10, 15)$	2.56	0.80	0.78	0.76	$n = (30, 50)$	1.60	0.79	0.78	0.72
	3.20	0.93	0.92	0.90		2.00	0.92	0.92	0.88
	d	RW	RF	W		d	RW	RF	W
	0.00	0.054	0.048	0.049		0.00	0.053	0.053	0.049
	0.60	0.12	0.11	0.10		0.32	0.11	0.11	0.10
	1.20	0.31	0.29	0.28		0.64	0.29	0.28	0.25
$n = (10, 30)$	1.80	0.56	0.54	0.52	$n = (50, 50)$	0.96	0.55	0.55	0.49
	2.40	0.78	0.77	0.73		1.28	0.78	0.78	0.71
	3.00	0.92	0.91	0.88		1.60	0.92	0.92	0.87
	d	RW	RF	W		d	RW	RF	W
	0.00	0.055	0.052	0.051		0.00	0.054	0.052	0.049
	0.60	0.12	0.11	0.11		0.26	0.11	0.11	0.10
$n = (10, 30)$	1.20	0.32	0.31	0.29	$n = (50, 50)$	0.52	0.30	0.29	0.25
	1.80	0.58	0.57	0.54		0.78	0.55	0.54	0.47
	2.40	0.80	0.79	0.75		1.04	0.79	0.79	0.71
	3.00	0.93	0.92	0.89		1.30	0.93	0.93	0.88
	d	RW	RF	W		d	RW	RF	W
	0.00	0.055	0.052	0.051		0.00	0.054	0.052	0.049

Place the caption here

Table 8: Simulated powers of the *RW*, *RF* and *W* tests for Model (g)

	<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>		<i>d</i>	<i>RW</i>	<i>RF</i>	<i>W</i>
<i>n</i> = (6, 6)	0.00	0.053	0.041	0.045	<i>n</i> = (20, 20)	0.00	0.054	0.053	0.049
	1.90	0.13	0.11	0.12		0.86	0.11	0.11	0.10
	3.80	0.33	0.29	0.30		1.72	0.30	0.29	0.26
	5.70	0.60	0.56	0.57		2.58	0.55	0.54	0.50
	7.60	0.81	0.78	0.78		3.44	0.78	0.78	0.74
	9.50	0.92	0.91	0.90		4.30	0.92	0.91	0.89
<i>n</i> = (6, 10)	0.00	0.056	0.052	0.051	<i>n</i> = (20, 30)	0.00	0.054	0.053	0.048
	1.90	0.13	0.12	0.11		0.86	0.12	0.11	0.10
	3.80	0.35	0.33	0.32		1.72	0.30	0.30	0.27
	5.70	0.61	0.59	0.58		2.58	0.56	0.56	0.51
	7.60	0.82	0.81	0.80		3.44	0.80	0.79	0.76
	9.50	0.93	0.92	0.91		4.30	0.92	0.92	0.90
<i>n</i> = (10, 10)	0.00	0.055	0.048	0.047	<i>n</i> = (20, 50)	0.00	0.055	0.055	0.049
	1.24	0.11	0.10	0.10		0.82	0.11	0.11	0.10
	2.48	0.28	0.27	0.25		1.64	0.29	0.29	0.26
	3.72	0.53	0.51	0.49		2.46	0.53	0.53	0.49
	4.96	0.75	0.74	0.72		3.28	0.76	0.75	0.71
	6.20	0.90	0.89	0.87		4.10	0.90	0.90	0.87
<i>n</i> = (10, 15)	0.00	0.055	0.052	0.048	<i>n</i> = (30, 50)	0.00	0.054	0.054	0.048
	1.24	0.11	0.11	0.10		0.64	0.11	0.11	0.10
	2.48	0.29	0.28	0.27		1.28	0.27	0.27	0.24
	3.72	0.54	0.53	0.50		1.92	0.50	0.50	0.46
	4.96	0.76	0.75	0.72		2.56	0.74	0.74	0.70
	6.20	0.90	0.90	0.88		3.20	0.89	0.89	0.86
<i>n</i> = (10, 30)	0.00	0.052	0.050	0.044	<i>n</i> = (50, 50)	0.00	0.054	0.053	0.049
	1.24	0.12	0.11	0.10		0.50	0.11	0.11	0.09
	2.48	0.30	0.29	0.27		1.00	0.26	0.26	0.23
	3.72	0.55	0.54	0.51		1.50	0.50	0.49	0.45
	4.96	0.76	0.76	0.73		2.00	0.73	0.73	0.68
	6.20	0.90	0.90	0.88		2.50	0.90	0.89	0.86

Place the caption here.

Table 9: Simulated powers of the RW , RF and W tests for Model (h)

	d	RW	RF	W		d	RW	RF	W
$n = (6, 6)$	0.00	0.048	0.032	0.044	$n = (20, 20)$	0.00	0.044	0.041	0.044
	1.16	0.13	0.10	0.12		0.54	0.11	0.10	0.10
	2.32	0.35	0.30	0.31		1.08	0.30	0.29	0.26
	3.48	0.61	0.56	0.56		1.62	0.57	0.56	0.49
	4.64	0.80	0.77	0.76		2.16	0.79	0.79	0.71
	5.80	0.91	0.90	0.88		2.70	0.93	0.92	0.86
$n = (6, 10)$	0.00	0.044	0.034	0.043	$n = (20, 30)$	0.00	0.047	0.044	0.047
	0.84	0.11	0.09	0.10		0.44	0.11	0.10	0.09
	1.68	0.30	0.26	0.27		0.88	0.28	0.27	0.24
	2.52	0.57	0.52	0.51		1.32	0.56	0.54	0.47
	3.36	0.78	0.74	0.72		1.76	0.79	0.78	0.69
	4.20	0.91	0.89	0.69		2.20	0.92	0.92	0.86
$n = (10, 10)$	0.00	0.044	0.036	0.042	$n = (20, 50)$	0.00	0.050	0.046	0.049
	0.80	0.12	0.10	0.11		0.34	0.10	0.09	0.09
	1.60	0.32	0.30	0.28		0.68	0.27	0.26	0.23
	2.40	0.57	0.55	0.50		1.02	0.50	0.49	0.42
	3.20	0.79	0.77	0.73		1.36	0.74	0.73	0.64
	4.00	0.91	0.90	0.86		1.70	0.90	0.89	0.82
$n = (10, 15)$	0.00	0.045	0.039	0.044	$n = (30, 50)$	0.00	0.046	0.045	0.046
	0.64	0.11	0.10	0.10		0.34	0.11	0.10	0.09
	1.28	0.29	0.27	0.25		0.68	0.29	0.28	0.24
	1.92	0.55	0.52	0.48		1.02	0.55	0.54	0.45
	2.56	0.78	0.76	0.71		1.36	0.79	0.78	0.69
	3.20	0.91	0.90	0.85		1.70	0.92	0.92	0.85
$n = (10, 30)$	0.00	0.045	0.039	0.045	$n = (50, 50)$	0.00	0.054	0.051	0.052
	0.48	0.11	0.09	0.10		0.30	0.10	0.09	0.08
	0.96	0.28	0.26	0.24		0.60	0.25	0.25	0.21
	1.44	0.54	0.51	0.46		0.90	0.48	0.48	0.40
	1.92	0.77	0.75	0.68		1.20	0.73	0.72	0.62
	2.40	0.91	0.90	0.85		1.50	0.90	0.89	0.80

Place the caption here.