ABSTRACT

Title of dissertation: Robust Means Modeling: An Alternative to Hypothesis Testing Of Mean Equality in the Between-Subjects Design under Variance Heterogeneity and Nonnormality

Weihua Fan, Doctor of Philosophy, 2006

Dissertation directed by: Professor Gregory R. Hancock

Department of Educational Measurement, Statistics and Evaluation

The study describes the various alternatives to the between-subjects ANOVA *F* test that have been performing reasonably well in the literature under different experimental conditions of sample sizes, variance ratios or nonnormality. Drawing from structural equation modeling (SEM), the robust means modeling (RMM) approach is developed, in which the assumption of variance homogeneity is not part of the model or its estimation. Specifically, univariate structured means modeling (SMM) is applied to the independent groups design with robust estimation strategies such as the Browne's asymptotic distribution free (ADF) estimator (1982, 1984) and its alternatives for non-normal continuous variables in order to achieve robustness to the biasing effects of nonnormality. A Monte Carlo simulation investigation is conducted to compare the Type I error rate and the power of the ANOVA-based methods as well as the proposed RMM approaches. Various factors including variance inequality, sample-size pairings with group variances, and degree of nonnormality are manipulated in the simulation. The results show that the proposed RMM methods are indeed superior to the ANOVA-based methods across conditions, especially when the distribution is asymmetric nonnormal.

Robust Means Modeling: An Alternative to Hypothesis Testing Of Mean Equality in the Between-Subjects Design under Variance Heterogeneity and Nonnormality

By

Weihua Fan

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Gregory R. Hancock, Chair/Advisor
Professor Amy B. Hendrickson
Professor Robert W. Lissitz
Professor George Macready
Professor Kevin E. O'Grady

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

<u>Overview of the Literature</u>

<u>Overview of ANOVA Alternatives under Assumption Violations</u>

The omnibus analysis of variance (ANOVA) *F* test and Student's *t* test have to satisfy certain assumptions to produce valid results, which briefly are 1) independence of observations, 2) normality of the treatment populations, and 3) homogeneity of population variances. The first assumption requires that the data are independent within and between groups, which is primarily about the study design and can be satisfied through random sampling. The latter two are functions of the populations under investigation and therefore are generally beyond the control of researchers. Failure to satisfy these assumptions, especially under unequal sample sizes, heterogeneous population variances and population nonnormality, alters Type I error rates (the probability of erroneously rejecting a true null hypothesis) and power (the probability of correctly rejecting a false null hypothesis).

The problem of testing for equality of means under assumption violations arises frequently in educational and psychological research. There is a great volume of literature devoted to describing the behavior of the ANOVA *F* test under assumption violations, under various degrees of each violation and under different conditions such as unequal sample sizes, heterogeneous variances and nonnormality. The issue of testing for means equality under variance heterogeneity in the early literature dates back to the time of Fisher (1935). Research shows that when a larger variance is associated with a larger sample size, the probability of Type I error declines below the nominal level (known as a "positive condition"). In contrast, when a larger variance is associated with a smaller sample size, the probability of Type I errors increases, sometimes far above the nominal level (known as a "negative condition"). See examples by Hsu (1938), Glass, Peckman, and Sanders (1972), and Overall, J. E., Atlas, R.

S., Gibson, J. M. (1995). Even when sample sizes are equal, empirical evidence shows that Type I error changes when variances are heterogeneous (Brown & Forsythe, 1974; Hsiung & Olejnik, 1996; Rogan & Keselman, 1977; Tomarken & Serlin, 1986), especially when the data are nonnormal.

Thus many applied researchers choose to use an alternative test procedure that is robust to assumption violations. A robust test is supposed to maintain the actual Type I error rate close to the nominal level and maintain actual statistical power close to theoretical power, even when the data do not conform to the assumptions of normal treatment populations and homogeneous population variances. It has been well documented in the statistical literature that these alternatives are generally superior to the ANOVA *F* test and Student's *t* test in the majority of assumption violation situations (e.g., Levy, 1978; Tomarken & Serlin, 1986).

The first type of alternative test procedure is the Welch (1938, 1951) type statistics, which deal with the deleterious effects of variance heterogeneity on the usual ANOVA *F* test and Student's *t* test (see Keselman, Lix, & Kowalchuk, 1998; Lix & Keselman, 1998). In addition, researchers (see Lix & Keselman, 1998; Wilcox, 1997) proposed using trimmed means and Winsorized variances rather than the usual least squares statistics to reduce the biasing effects of nonnormality. A number of papers have studied the robustness to nonnormality and variance heterogeneity in unbalanced independent designs by using robust estimators with various test statistics (Keselman, Algina, Wilcox, & Kowalchuk, 2000; Keselman, Kowalchuk, & Lix, 1998; Wilcox, R. R., Keselman, H. J., Muska, J., & Cribbie, R., 2000). However, this method is usually criticized in that it alters the hypothesis of test for equality of means by removing the extreme observations.

Bootstrap methods may be incorporated to further improve Type I error control by obtaining critical values for test statistics. Such improvement has been demonstrated with statistics for independent group designs (Wilcox, Keselman, & Kowalchuk, 1998). Wasserman and Bockenholt (1989) indicated how various inferential problems (e.g., correlational and general linear model analyses) could be addressed via bootstrap methods. However, the overwhelming burden of conducting these methods for applied researchers, because of the lack of statistical software, has restricted the use of these methods.

The purpose of this dissertation is threefold: (a) to describe the various alternatives to the between-subjects ANOVA F test that have been performing reasonably well in the literature under different experimental conditions of sample sizes, variance ratios or nonnormality, (b) to develop the robust means modeling (RMM) approach drawing from structural equation modeling (SEM) in which the assumption of variance homogeneity is not part of the model or its estimation. Specifically, univariate Structured Means Modeling (SMM) is applied to the independent groups design with robust estimation strategies such as the Browne's asymptotic distribution free (ADF) estimator (1982, 1984) and its alternatives for non-normal continuous variables in order to achieve robustness to the biasing effects of nonnormality. The third propose of the study is (c) to detail the data- analytic conditions under which it may be advantageous to use the proposed RMM methods. A simulation investigation was conducted on the proposed RMM approaches as well as on the ANOVA procedures including the ANOVA *F* test and its alternatives to assess the Type I error and power rates under a wide variety of experimental conditions.

This dissertation has the following format. The remainder of this chapter looks very briefly at the literature on assumption violations and the robustness studies of the ANOVA *F* test as well as its alternatives. A detailed description of the various alternatives to the ANOVA F test follows. Next, the methodology of robust means modeling is developed, followed by the results of the simulation investigation. A discussion of the implications of this study for educational and psychological researchers concludes the dissertation.

Alternatives to the ANOVA *F* Test

Nonparametric Tests

Researchers faced with data that appear to violate the assumptions of the ANOVA *F* test usually consider two approaches. One approach is the nonparametric rank tests, such as the Wilcoxon-Mann-Whitney test and the Kruskal-Wallis test (Kruskal & Wallis, 1952). These tests combine the scores of two or more treatment groups together and convert them to a single set of ranks. Then, the tests replace scores in various treatment groups by their corresponding ranks. Finally, test statistics are calculated, such as the Wilcoxon *T*, the Mann-Whitney *U*, or the Kruskal-Wallis *H*, from the sums of ranks in the respective groups. The Wilcoxon-Mann-Whitney test is equivalent to the Student's *t* statistic calculated from ranks (Conover & Iman, 1981). Similarly, the Kruskal-Wallis test is equivalent to an ordinary *F* test performed on ranks instead of scores. Further discussion of the rank-transformation concept can be found in Sawilowsky, Blair, and Higgins (1989).

For many years, researchers and applied statisticians have assumed that nonparametric tests are not influenced by heterogeneous variances. Unfortunately, these nonparametric methods are sensitive to unequal variances due to the properties of variances of ranks. Moreover, it is easy to understand why methods based on test statistics that are functions of ranks, such as the normal scores test (van der Waerden, 1952), which replaces ranks by quantiles of a standard normal distribution, share the same property.

More recently, however, it has become apparent that these test statistics also are biased when sample sizes are unequal (see, for example, Oshima & Algina, 1992a, Tomarken & Serlin, 1986; Zimmerman, 1996; Zimmerman & Zumbo, 1993). Even when sample sizes are equal, the statistical nominal levels of the Wilcoxon-Mann-Whitney test and the Kruskal-Wallis test are substantially biased by heterogeneous variances of treatment groups (Zimmerman, 2000). Given the unsatisfactory performance of both the ANOVA *F* test, the

Kruskal-Wallis test and the Wilcoxon-Mann-Whitney test under variance heterogeneity, researchers have concentrated their efforts on investigating parametric alternatives to the ANOVA *F* test that do not assume variance homogeneity.

Parametric Tests

Many solutions have been developed for testing the hypothesis of mean equality in the one-way independent groups design when variance homogeneity is not a tenable assumption. Four of the most popular tests in this category are the Welch test (1938, 1947), the Brown-Forsythe test (Brown & Forsythe, 1974), the James second-order test (James, 1951), and an approximation proposed by Alexander and Govern in 1994. Although other procedures are available, these tests have been investigated in numerous Monte Carlo simulations; overall, they have proved to be the most effective tests in controlling Type I error as well as in providing competitive power under varying experimental conditions (Harwell, 1992; Lix, Keselman, & Keselman, 1996). The following are the descriptions for the ANOVA *F* test as well as the four mentioned parametric tests.

Let X_{ik} be the *i*th observation in the *k*th group, where $i = 1... n_k$ and $k = 1…k;$ let $\sum n_k = N$. The X_{ik} are assumed to be independent and normally distributed with expected values μ_k and variances σ_k^2. The estimates of μ_k and σ_k^2 are

$$\overline{X}_{.k} = \sum_i X_{ik} / n_k \,, \tag{1}$$

and

$$S_k{}^2 = \sum_i (X_{ik} - \overline{X}_{.k})/(n_k - 1) \,. \tag{2}$$

1. ANOVA *F* test is computed as

$$F = \frac{\sum_k n_k (\overline{X}_{.k} - \overline{X}..)^2 / (k-1)}{\sum_l \sum_k (X_{ik} - \overline{X}_{.k})^2 / (N-k)},$$

where $\overline{X}.. = \sum_k n_k \overline{X}_{.K} / N$. The degrees of freedom for the numerator and denominator are k-1 and N-k. The four parametric ANOVA alternative test statistics follow.

2. Welch v_w (1951):

$$v_w = \frac{\sum_k w_k (\overline{X}_{.k} - X'..)^2 / (k-1)}{[1 + \frac{2}{3}((k-2)\Lambda)]}, \tag{3}$$

where $w_k = n_k / S_k^2$, (4)

$$X'.. = \sum_k w_k \overline{X}_{.k} / W, \tag{5}$$

where $W = \sum_k w_k$ and

$$\Lambda = \frac{3\sum_k (1 - w_k / W)^2 / (n_k - 1)}{k^2 - 1}. \tag{6}$$

The Welch v_w statistic is approximately distributed as a central F variable with v_1=k- l and v_2 = $1/\Lambda$ degrees of freedom. It is undoubtedly one of the best-known approximate procedures, which is available in the BMDP (Dixon, 1992) and SPSS statistical packages. As well, a generalized version of this procedure is available in a SAS/IML program (SAS Institute, 1990) developed by Lix and Keselman (1995).

3. Brown and Forsythe F^* (1974):

The procedure presented by Brown and Forsythe (1974; BF) is based on a test statistic, in which the numerator and denominator have the same expected value under the null hypothesis,

$$F^* = \frac{\sum_k n_{.k}(\overline{X}_{.k} - \overline{X}_{..})^2}{\sum_k (1 - n_k / N) S_k^{\ 2}}. \tag{7}$$

The F^* statistic is approximately distributed as an F variable with $v_1 = k$-1 and $v_2 = f$ degrees of freedom, where f is obtained with the Satterthwaite (1941) approximation as

$$\frac{1}{f} = \sum_k c_k^{\ 2} / (n_k - 1) \tag{8}$$

with

$$c_k = (1 - n_k / N) S_k^{\ 2} / [\sum_k (1 - n_k / N) S_k^{\ 2}]. \tag{9}$$

The BF procedure is available in the BMDP (Dixon, 1992) and SPSS statistical software packages.

4. Alexander and Govern A (1994):

In 1994, Alexander and Govern proposed a new test based on an approximation of the probabilities associated with a t distribution formulated by Hill (1970). This approximation is used in the development of Alexander and Govern's statistic,

$$A = \sum_{i=1}^{k} z_k^2 \tag{10}$$

where

$$z_k = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855c)}{(10b^2 + 8bc^4 + 1000b)} \tag{11}$$

when $a = v_k$-.5, b=48a^2, $c = [a \ln(1 + t_k^2 / v_k)]^{1/2}$, $v_k = n_k - 1$ and t_k is the sample t statistic for each group. The following formula is used to calculate t_k:

$$t_k = \frac{\overline{X}_{.k} - X^+}{S''_k} \tag{12}$$

where $X^{+} = \sum_{i=1}^{k} w_k \overline{X}_{.k}$. For each of the k groups, a weight (w_k) is defined such that $\sum w_k = 1$:

$$w_k = \frac{1/S''^{2}_{k}}{\sum_{i=1}^{k} 1/S''^{2}_{k}}. \tag{13}$$

On relaxing the assumptions of equal variances and equal *ns,* each of the k groups of size n_k will have sample mean ($\overline{X}_{.k}$), and each mean will have a standard error (S''_{k}):

$$S''_{k} = S_{\overline{X}_{.k}} = \left\{ \frac{\sum_{i=1}^{n_k} (X_i - \overline{X}_{.k})}{n_k(n_k - 1)} \right\}^{1/2}. \tag{14}$$

The A statistic is approximately distributed as χ^2 with k-1 degrees of freedom.

5. James second-order U (1951):

Two statistical procedures are proposed by James (1951), referred to as his first- and second-order procedures. The two procedures are equivalent in terms of the defined test statistic, but differ in the critical value used to assess statistical significance. The former is not widely recommended, as it cannot effectively control the rate of Type I errors under variance heterogeneity (Brown & Forsythe, 1974). The James second-order procedure (James) is based on the statistic

$$U = \sum_{k} w_k (\overline{X}_{.k} - X'..)^2 , \tag{15}$$

where w_k and $X'..$ are as previously defined. For any integers t and s, let

$$R_{st} = \sum_{k} \frac{1}{(n_k - 1)^s} \left(\frac{w_k}{W} \right)^t . \tag{16}$$

Furthermore, let $x_{2s} = r^s / [(k-1)(k+1)\ldots(k+2s-3)]$, where r is the (1-α) centile of a chi-square distribution with (k-1) *df*. The statistic U is compared to the critical value, $h(\alpha)$, where

$$h(\alpha) = [(r+(1/2)(3x_4+x_2)\sum_k (1-w_k/W)^2/(n_k-1)$$

$$+\{(1/16)(3x_4+x_2)^2(1-(k-3)/r)\left[\sum_k (1-w_k/W)^2/(n_k-1)\right]^2$$

$$+(1/2)(3x_4+x_2)[(8R_{23}-10R_{22}+4R_{21}-6R^2_{12}+8R_{12}R_{11}-4R^2_{11})$$

$$+(2R_{23}-4R_{22}+2R_{21}-2R^2_{12}+4R_{12}R_{11}-2R^2_{11})(x_2-1)$$

$$+(1/4)(-R^2_{12}+4R_{12}R_{11}-2R_{12}R_{10}-4R^2_{11}+4R_{11}R_{10}-R^2_{10})(3x_4-2x_2-1)]$$

$$+(R_{23}-3R_{22}+3R_{21}-R_{20})(5x_6+2x_4+x_2)$$

$$+(3/16)(R^2_{12}-4R_{23}+6R_{22}-4R_{21}+R_{20})(35x_8+15x_6+9x_4+5x_2)$$

$$+(1/16)(-2R_{22}+4R_{21}-R_{20}+2R_{12}R_{10}-4R_{11}R_{10}+R^2_{10})(9x_8-3x_6-5x_4-x_2)$$

$$+(1/4)(-2R_{22}+R^2_{11})(27x_8+3x_6+x_4+x_2)$$

$$+(1/4)(R_{23}+R_{12}R_{11})(45x_8+9x_6+7x_4+3x_2) \tag{17}$$

and the null hypothesis is rejected if $U > h(\alpha)$.

Applied researchers may feel that one of the main drawbacks of using the James procedure is its computational complexity. Oshima and Algina (1992a) developed a program written in the SAS language that tests H_0 using this approach. The required input to the program is the sample size, mean, and variance for each group. The program can handle data for one-way designs containing as many as 10 groups.

Choosing among parametric alternatives to the ANOVA F test

Studies of statistical robustness of the various parametric alternatives to the ANOVA *F* test strongly suggest that no one approach is best in all situations. The Welch test, developed in 1951, has been included in almost all studies investigating an acceptable alternative to the ANOVA *F* test (Algina, Oshima, & Lin, 1994; Clinch & Keselman, 1982; Dijkstra & Werter, 1981; Lix, Keselman, & Keselman, 1996; Oshima & Algina, 1992; Wilcox, Charlin, &

Thompson, 1986). The literature indicates that the Welch test is relatively robust to departures from variance heterogeneity. However, if sample size is small, the ability of Welch's test to limit the Type I error rate to the nominal level may decrease as variance heterogeneity increases or as the number of groups increases (Wilcox, 1988).

Brown and Forsythe’s test, along with the Welch test, has been shown to provide good control of the Type I error rate. However, a number of researchers (e.g. Tomarken & Serlin, 1986; Wilcox et al., 1986) have shown that the BF procedure is not comparable to the Welch test or the James second-order test in some situations. The ability of the BF method to limit the Type I error rate to the nominal level may be compromised depending on the pattern of the population variances (Tomarken & Serlin, 1986). While several studies recommend the use of the James second-order test (e.g., Lix, Keselman, & Keselman, 1996; Oshima & Algina, 1992b; Wilcox, 1988), these studies have also shown that this approach may not always control the Type I error rate when sample size is small and when the data are obtained from populations with asymmetric distributions. The Alexander and Govern approximation is recommended by Schneider and Penfield (1997) as the best alternative to the ANOVA *F* test when variances are heterogeneous for the following reasons: the simplicity of computation and its overall superiority when considering both Type I error and power rate under most experimental conditions.

In sum, the evidence of all of these parametric tests suggests that these methods can generally control the rate of Type I error when group variances are heterogeneous or the data are normally distributed. However, the literature also indicates that these tests can become liberal when the data are both heterogeneous and nonnormal, particularly when the design is unbalanced. Thus, these statistics have limitations, namely, their sensitivity to the nature of the population distributions.

Robust Estimators

When the score distribution has heavy tails, that is, there are extreme observations, the usual group means and variances as well as the standard error of the usual mean will be greatly influenced by these extreme observations. The standard error of the usual mean can become seriously inflated when the underlying distribution has heavy tails. Accordingly, researchers seek to substitute robust measures of location and scale for the usual mean and variance in order to achieve test statistics that are insensitive to the combined effects of variance heterogeneity and nonnormality.

Although a wide range of robust estimators have been proposed in the literature, the trimmed mean and Winsorized variance are most appealing because of their computational simplicity and good theoretical properties (Wilcox, 1995). By censoring or removing extreme observations, the standard error of the trimmed mean is less affected by departures from normality than the usual mean. Similarly, the most extreme observations are replaced with less extreme values in the distribution of scores in computing the Winsorized variance.

However, the trimmed means and Winsorized variance should only be adopted if the researcher is interested in testing for treatment effects across groups using a measure of location that more accurately reflects the typical score within a group when working with heavy-tailed distributions (Lix and Keselman, 1998). These robust estimates are sometimes criticized that their null hypothesis has been altered to test the equality of population trimmed means rather than what ANOVA *F* is testing, namely the equality of population means. Thus, researchers need to be clear on the goals of data analysis prior to choosing a particular method of statistical inference.

Trimmed means and Winsorized variances can be used in conjunction with Welch v_w (1951) statistic for heavy-tailed symmetric distributions, initially suggested by Yuen (1974). The Welch test with trimming adequately controls the rate of Type I errors and results in greater power than statistics based on the usual mean and variance for two group designs. Lix and Keselman (1998) expanded the study to more groups and more test statistics. As a result, they recommend the trimmed means and Winsorized variances with the Welch, Alexander and Govern, or James statistics to test for mean equality in between-subjects designs.

When trimmed means are being compared, the null hypothesis pertains to the equality of population-trimmed means, i.e. the u_t. That is, Ho: $\mu_{t1}=\mu_{t2}=\ldots=\mu_{tk}$. Let $X_{(1)k}\leq X_{(2)k}\leq\cdots\leq X_{(n_j)k}$ represent the ordered observations associated with the kth group. Let $g_k=[\gamma n_k]$, where γ represents the proportion of observations that are to be trimmed in each tail of the distribution. Wilcox (1995) suggested that 20% trimming should be used, that is, $\gamma=.2$.

The effective sample size for the kth group becomes $h_k=n_k-2g_k$. The kth sample trimmed mean is

$$\overline{X}_{tk}=\frac{1}{h_k}\sum_{i=g_k+1}^{n_k-g_k}X_{(i)k}\ . \tag{18}$$

To compute the sample Winsorized variance, the sample Winsorized mean is necessary and is computed as

$$\overline{X}_{wk}=\frac{1}{n_k}\sum_{i=1}^{n_k}Y_{ik}\ , \tag{19}$$

where

$$Y_{ik}=X_{(g_k+1)k} \quad \text{if } X_{ik}\leq X_{(g_k+1)k} \tag{20}$$

$$=X_{ik} \quad \text{if } X_{(g_k+1)k}<X_{ik}<X_{(n_k-g_k)k} \tag{21}$$

$$=X_{(n_k-g_k)k} \quad \text{if } X_{ik}\geq X_{(n_k-g_k)k}\ . \tag{22}$$

The sample Winsorized variance according to Lix and Keselman (1998) is

$$s_{wk}^2=\frac{1}{h_k-1}\sum_{i=1}^{n_k}(Y_{ik}-\overline{X}_{wk})^2\ , \tag{23}$$

and

$$\tilde{s}_{wk}^2 = \frac{s_{wk}^2}{h_k} \tag{24}$$

estimates the squared standard error of the sample-trimmed mean (see Wilcox, 1996).

Thus, with robust estimation, the trimmed group means ($\overline{X}_{tk}$) replace the least squares group means ($\overline{X}_k$), the Winsorized group variances estimators (s_{wk}^2) replace the least squares variances (s_k^2), and $\Sigma_k h_k$ replaces N, in the test statistics and their degrees of freedom (*df*). For example the Welch test with the robust estimation becomes

$$v_{wt} = \frac{\sum_k w_{tk}(\overline{X}_{tk} - \overline{X}_t)^2/(K-1)}{[1+\frac{2}{3}((K-2)\Lambda)]}, \tag{25}$$

where $w_{tk} = 1/\tilde{s}_{wk}{}^2$, (26)

$$\overline{X}_t = \sum_k w_{tk}\overline{X}_{tk}/W_t, \tag{27}$$

where $W_t = \sum_k w_{tk}$ and

$$\Lambda = \frac{3\sum_k (1-w_{tk}/W_t)^2/(h_k-1)}{K^2-1}. \tag{28}$$

This Welch v_w statistic is approximately distributed as a central F variable with $v_1 = k- 1$ and $v_2 = 1/\Lambda$ degrees of freedom with the trimmed and Winsorized statistics.

Chapter II

Robust Means Modeling

Introduction to Structured Means Modeling

Structural equation modeling (SEM) has become a popular data analytic method in educational and psychological research. Growing out of but more powerful than multiple regression, SEM can deal with interactions, nonlinearities, measurement error, correlated independents and error terms, multiple latent independents/dependents with multiple indicators. It includes specialized versions of a number of other analysis methods as special cases, such as multiple regression, path analysis, factor analysis, time series analysis, and analysis of covariance. Developed from SEM, Sörbom's (1974) method of structured means modeling (SMM) seeks to combine ANOVA principles with confirmatory factor analysis, leading to tests of between- or within-population differences on an underlying latent construct. The hypothesis testing regarding means may then be conducted at the latent variable level (see, e.g., Hancock, 2003, 2004).

In order to estimate the differences between groups on latent variable means, it is necessary to expand the factor model to incorporate intercepts. For a set of p observed variable x for construct ξ, $\mathbf{x}$ values in a single group may be expressed in a $p\times1$ vector as follows:

$$\mathbf{x} = \mathbf{v} + \boldsymbol{\Lambda}\xi + \delta, \tag{29}$$

where $\mathbf{v}$ is a $p\times1$ vector of intercept values, $\boldsymbol{\Lambda}$ is a $p\times1$vector of λ loadings, and δ is a $p\times1$vector of normal errors. Thus, the expected values or means for the observed indicators can be computed as

$$\mathrm{E}[\mathbf{x}] = \boldsymbol{\mu} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\kappa, \tag{30}$$

where $\boldsymbol{\mu}$ is the $p \times 1$ population mean vector for the observed variables, and κ is the mean of factor ξ. Assuming that the errors are independent, the model implied variance-covariance matrix becomes:

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'] = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Theta}, \quad (31)$$

where Φ is the factor variance for ξ and $\boldsymbol{\Theta}$ is the $p \times p$ error covariance matrix for δ. Hancock (1997) provides a more detailed introduction to SMM.

Robust Means Modeling

Application of SMM to the independent groups design

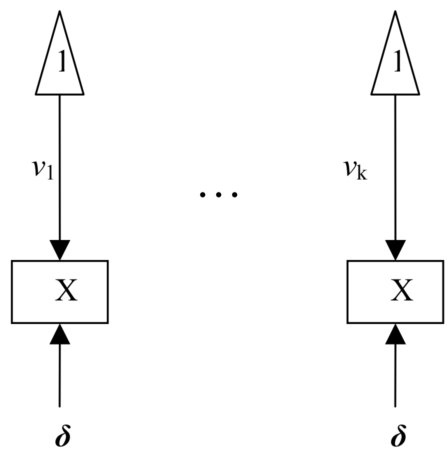
Two aspects of SMM are directly relevant for between-subject ANOVA-type designs, the focus of the currently proposed investigation. First, the above model makes no assumptions about homogeneity of variance; that is, the SMM framework allows each of the k populations to have its own latent variable variance, σ_k^2. Second, a special case of SMM is that a simple model with only one measured indicator variable without any external information is necessarily set equal to the construct ξ. Thus, when there are no effects from a latent factor, but only a single observed variable, the latent σ_k^2 values can be set to zero and the latent model can be simplified to a measured variable mean structure model

$$x_k = \nu_k + \varepsilon_k, \quad (32)$$

where ν_k is the kth population mean for the observed variable, ε_k is the kth population residual (within-group) error term. Graphically, we can imagine that a constant number of one, as well as the error variable, have direct arrows to the observed variable (See Figure 1). The null hypothesis for this univariate SMM thus becomes H_0: $\nu_1 = \nu_2 = \ldots = \nu_k$.

Figure 1.

Application of SMM to Independent Groups Design



To test the equivalence of the means/intercepts, a constraint can be imposed forcing intercepts equivalent across populations while still allowing for heterogeneous variances (i.e., var(ε_k)=$\theta_k \neq \theta$). The model implied variance for each population is just the implied error variance for the observed variable. That is, $\hat{\boldsymbol{\Sigma}}_k = Var(\delta_k)$. To understand this equivalency, consider the following computation:

$$\hat{\boldsymbol{\Sigma}}_k = Var(X_k) = E[(X_k - E(X_k))^2] = E[((v_k + \delta_k) - v_k)^2] = E(\delta_k^{\ 2}) = Var(\delta_k). \tag{33}$$

Since the intercepts are constrained to be equal across groups, we then have the expected mean for both groups as $E(X_1)=\ldots= E(X_k)=v=u$ with p=1.

Estimation Methods

In the application of a SMM, there is a distributional assumption that underlies the statistical methods. Traditional parameter estimation methods such as Maximum Likelihood (ML, T_{ML}) make the distributional assumption that the measured variables have a multivariate normal distribution in the population (Satorra, 1990). Any violation of the assumption can produce inaccurate parameter estimates as well as standard error estimates, which may lead to unreasonable interpretation of the results. However, the majority of data collected in behavioral research do not follow univariate or multivariate normal distributions. Violations of the assumption of multivariate normality are common (and often unavoidable) in practice and can potentially lead to seriously misleading results (Micceri, 1989).

Many alternative estimation methods to ML have been proposed to address the biasing effects caused by nonnormal distributions. The first alternative estimation method is the Browne's (1982, 1984) asymptotically distribution free (ADF) estimation method, which relaxes distributional assumptions and yields a model test statistic, T_{ADF}. Although the ADF approach has the disadvantage of requiring large sample sizes for complex models, the incorporation of the ADF estimation to a very simple model involving only a single observed variable is expected to perform reasonably well. The second alternative estimation method is Satorra and Bentler's (1988) rescaled test statistic (SB), which adjusts T_{ML} and ML standard errors to yield a test statistic approximating the referenced chi-square distribution (Browne, 1982, 1984). The third and fourth alternative estimation methods are Yuan and Bentler's (1997, 1999) test statistics T_{YB1} and T_{YB2}, which make corrections to T_{ADF} for small sample sizes. The last alternative estimation method is Bartlett's correction (Fouladi, 1998, 1999, 2000, Bartlett, 1950) to the ML test statistic. Various studies have been conducted to understand the effects of the multivariate nonnormality on maximum likelihood estimation and the five alternative estimators used in SEM. For more details for the above estimation methods, see studies by Anderson & Gerbing, 1984; Browne, 1982; Chou, Bentler, & Satorra, 1991; Curran, West, & Finch, 1996; Finch, West, & MacKinnon, 1997; Hu, Bentler, & Kano, 1992.

These estimation strategies can be applied to univariate SMM to create a framework for between-subjects designs that has neither variance nor distributional restrictions, which are termed robust means modeling (RMM) approaches in this study. To be more specific, the approach is to apply alternate estimation to SMM to the hypothesis test of mean equality. If the utilization of the robust estimates handles nonnormality, the RMM approach should be a convenient and efficient way of testing mean equality under heterogeneity and nonnormality conditions simultaneously. What makes the RMM approaches even more attractive is that the techniques are available now with popular software such EQS 6.1. Programming code of EQS will be provided for the estimation strategies of T_{ML}, T_{ADF}, T_{SB}, based on which the test statistics of T_{YB1}, T_{YB2}, and T_{BC} can be easily computed. Alternate estimation options are described briefly below. Discussed are some widely used methods of estimation in SEM, the ML estimation and its alternatives -- T_{ML}, T_{ADF}, T_{SB}, T_{YB1}, T_{YB2}, and T_{BC}.

1. *Maximum Likelihood*

Maximum Likelihood (ML) is the most widely used fitting function for structural equation models under normality. Nearly all of the major software programs use ML as the default estimator. According to Bollen (1989), the general fit function becomes a weighted combination of the fit for the multi-group situation:

$$F_{\mathrm{ML}}=\sum_{K=1}^{K}\left(\frac{n_k}{N}\right)F_{\mathrm{k}}(\mathbf{S}_k,\ \hat{\mathbf{\Sigma}}_k), \qquad (34)$$

where $\mathbf{S}_k$ is the group's covariance matrix, n_{k} is the sample size in the kth group, N= n_1 + n_2+…+ n_k, $\hat{\mathbf{\Sigma}}_k$ is the hypothesized structure implied covariance matrix for each group, and $F_{\mathrm{k}}(\mathbf{S}_k, \Sigma_{\mathrm{k}})$ is the fit function for the kth group. The maximum likelihood fitting functions are computed in the same way for each group with the inclusion of parameter restrictions across groups and the simultaneous minimization of a composite fit function of two or more groups. In addition, according to Bentler (1995), the maximum likelihood function for means structures adds weighted sum of squares resulting from the discrepancy from $\mathbf{m}_k$ and $\hat{\boldsymbol{\mu}}_k$

$$F_{ML(k)} = [\ln|\hat{\boldsymbol{\Sigma}}_k| - tr(\mathbf{S}_k\hat{\boldsymbol{\Sigma}}_k^{-1}) - \ln|\mathbf{S}_k| - p] + (\mathbf{m}_k - \hat{\boldsymbol{\mu}}_k)'\hat{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{m}_k - \hat{\boldsymbol{\mu}}_k), \quad (35)$$

where $\mathbf{m}_k$ is the observed mean vector for the kth group and $\hat{\boldsymbol{\mu}}_k$ is the model implied mean for each population. The general fit value, in turn, can be converted to a model test statistic that presents fit information over groups. The overall chi-square estimate for all the groups can be approximated by

$$T_{ML} = (n_1-1)F_{ML(1)} + (n_2-1)F_{ML(2)} + \ldots + (n_k-1)F_{ML(k)}. \quad (36)$$

The null hypothesis is that the constraints of the model in all groups are correct. The degrees of freedom equal $kp(p+3)/2-q$, where k is the number of groups, and q is the number of parameters estimated in all groups.

Inputting all of the information from equation (33) to equation (35) yields a simplified fit function of each group for the RMM approach, which applies SMM to a single observed variable:

$$F_{ML(k)} = [\ln[\mathrm{var}(\delta_k)] - \ln(s_k^2) + s_k^2/\mathrm{var}(\delta_k) - 1] + (\overline{X}_k - \hat{\mu})^2/\mathrm{var}(\delta_k)$$

$$= \ln[\frac{\mathrm{var}(\delta_k)}{s_k^2}] + \frac{s_k^2}{\mathrm{var}(\delta_k)} + \frac{(\overline{X}_k - \hat{\mu})^2}{\mathrm{var}(\delta_k)} - 1, \quad (37)$$

where s_k^2 is the sample variance for each group, $\mathrm{var}(\delta_k)$ is the model implied error variance for each group, and $\overline{X}_k$ is the observed mean for each group. Thus, the overall chi-square statistic for the test will be calculated as formula (36). The test statistic is approximately distributed as a chi-square distribution with degrees-of-freedom of (k-1). That is, a total of k+1 parameters (k variances and one common mean) are estimated from the 2k means and variances in the data, thereby yielding a test with 2k-(k+1)=k-1 *df*. SEM software such as EQS, LISREL, or Mplus can easily conduct the test of the above equality constraint and provide an associated test statistic, T_{ML}. Under assumptions of normality, this test statistic is approximately distributed as a χ^2 with *df=k-1*. If T_{ML} exceeds the desired α-level critical value then the null hypothesis of population mean equality is rejected, indicating group mean differences. If T_{ML} does not exceed the critical value then population mean equality remains tenable, showing that there is no significant mean difference between groups.

Numerical Example. Here is a numerical example of the application of the RMM approach to a two-independent samples design. The two samples have n_1=80 and n_2=20 with the observed sample means and variances of $\overline{X}_1$=.07, $\overline{X}_2$=-.2, s_1^2=(1.5)²=2.25 and s_2^2=(1.2)²=1.44. EQS 6.1 was used to run the study (See Appendix A.1.). Key estimates from the standard portion of the output included $\chi^2{}_{(1)}$=.692, $\text{var}(\delta_1) = 2.255$, $\text{var}(\delta_2) = 1.479$ and $\hat{\mu}$=-.002. Using the derived formula (37) for the chi-square statistic

$$F_{ML(1)} = \ln[\frac{\text{var}(\delta_1)}{s^2{}_1}] + \frac{s^2{}_1}{\text{var}(\delta_1)} + \frac{(\overline{X}_1 - \hat{\mu})^2}{\text{var}(\delta_1)} - 1$$

$$= \ln\left(\frac{2.255}{2.25}\right) + \frac{2.25}{2.255} + \frac{(.07-(-.002))^2}{2.255} - 1$$

$$= .0023\ . \tag{38}$$

$$F_{ML(2)} = \ln[\frac{\text{var}(\delta_2)}{s^2{}_2}] + \frac{s^2{}_2}{\text{var}(\delta_2)} + \frac{(\overline{X}_2 - \hat{\mu})^2}{\text{var}(\delta_2)} - 1$$

$$= \ln\left(\frac{1.479}{1.44}\right) + \frac{1.44}{1.479} + \frac{(-.2-(-.002))^2}{1.479} - 1$$

$$= .02686. \tag{39}$$

Thus the chi-square statistic equals

$$T_{ML} = (n_1\text{-}1)F_{ML(1)} + (n_2\text{-}1)F_{ML(2)} = 79{*}.0023+19{*}.026861=.692 \tag{40}$$

The chi-square statistic computed from the formula is identical to the value provided by EQS output. Since .692 is less than 3.84, the chi-square critical value with *df*=1, the null hypothesis of population mean equality is retained, indicating that there is no significant mean difference between groups.

A limitation of traditional ML estimation is the strong assumption of multivariate normality. For ML estimation with small samples, T_{ML} was not robust to departures from multivariate normality, yielding inflated Type I error rates. That is, in practice a researcher may mistakenly reject or incidentally modify a model just because the distribution of the observed variables is not multivariate normal rather than because the model itself is not correct. With increasing nonnormality, ML has been found to be increasingly biased (Curran, West, & Finch, 1996; West, Finch, & Curran, 1995). With normally distributed data, T_{ML} performed reasonably well for properly specified models (Curran, West, & Finch, 1996; West, Finch, & Curran, 1995), but was inflated most notably at sample size ratios that meet or exceeded the 5:1 guideline for using ML estimation (Nevitt & Hancock, 2004).

2. *Browne's asymptotic distribution free test statistic*

If the data are continuous but nonnormal, the estimation method most often recommended is the asymptotically distribution free (ADF) method (Browne, 1984). The method may also be applied if the distributions of the continuous variables deviate considerably from normality, even if some of the observed variables are ordinal and others continuous, or the models include dichotomous variables. The ADF method is available in LISREL under the name "weighted least squares (WLS)" and in EQS under "arbitrary distribution generalized least squares (AGLS)". In contrast to ML, raw data are needed for data analysis with this method. EQS code to obtain the test statistic is provided by Appendix A.2.

The extension of the ADF estimator to SMM results in minimizing the fitting function (Muthén, 1989)

$$\hat{F}_{ADF} = \sum_{k=1}^{K} (\mathbf{s}^{(k)}\text{-}\boldsymbol{\sigma}^{(k)})' \,\Gamma^{(k)^{-1}}(\mathbf{s}^{(k)}\text{-}\boldsymbol{\sigma}^{(k)}), \tag{41}$$

with respect to the model parameters. Here, $\boldsymbol{\sigma}^{(k)'}=(\boldsymbol{\sigma}_1^{(k)'}\ \boldsymbol{\sigma}_2^{(k)'})$. $\boldsymbol{\sigma}_1^{(k)'}$ is a $(p\times1)$ vector-valued function of the mean vector parameters, which is the model-implied population mean in the RMM approach. $\boldsymbol{\sigma}_2^{(k)'}$ is a $(p(p+1)/2\times1)$ vector of valued function corresponding to the distinct covariance matrix elements for each population. To be more specific, $\boldsymbol{\sigma}^{(k)'}$ becomes a (2×1) vector in the RMM approach, which has the first element as the model-implied population mean and the second element as the model-implied variance for each population. Containing both the mean and variance information, $\mathbf{s}^{(k)'}=(\mathbf{s}_1^{(k)'}\ \mathbf{s}_2^{(k)'})$. Similarly, $\mathbf{s}_1^{(k)}$ $(p\times1)$ and $\mathbf{s}_2^{(k)}$ $(p(p+1)/2\times1)$ denote sample mean and sample covariance matrix respectively. In the RMM approach, $\mathbf{s}_1^{(k)'}$ is a (2×1) vector with the first element as the sample mean for each group and the second element as the sample variance for each group. Γ represents the weight matrix utilized with ADF, which is the asymptotic covariance matrix, a matrix of the observed sample variances and covariances (Bollen, 1989).

Under multivariate normality, $\Gamma_{11}^{(k)}=\mathrm{N}^{(k)\text{-}1}\mathbf{S}^{(k)}$, $\Gamma_{21}^{(k)}=0$ and $\Gamma_{22}^{(k)}=K^{*}(\mathbf{S}^{(k)}\times\mathbf{S}^{(k)})\,K^{*\prime}$, where $\mathrm{N}^{(k)}$ is the sample size for group k, $\mathbf{S}^{(k)}$ is the sample covariance matrix, and K^{*} is a constant matrix that selects elements (Browne, 1974; Kendall & Stuart, 1977). Since $\Gamma_{11}^{(k)}$ does not change when the normality assumption is relaxed, it only remains to find $\Gamma_{21}^{(k)}$. Consider the p-dimensional vector $\mathbf{y}^{*}$ for observation i,

$$\mathbf{y_i}^{*\prime} = (\mathrm{y}_{11} - \bar{\mathrm{y}}_1 \ldots \mathrm{y}_{pi} - \bar{\mathrm{y}}_p), \tag{42}$$

where $\overline{\mathrm{y}}_{\mathrm{i}}$ is the sample mean of $\mathbf{y}_{\mathrm{i}}$ and creates the p(p+1)/2-dimensional vector $\mathbf{a}_{\mathrm{i}}$,

$$\mathbf{a}_{\mathrm{i}}' = *(y*_{li}\; y_{li} * y*_{2i}\; y_{li} * y*_{2i}\; y_2 * y*_{3i}\; y_{1i} \ldots y*_{pi}\; y_{p-1,i} y*_{pi}\; y*_{pi}) . \quad (43)$$

For each group *k*, deleting the group index, we may then create Γ_{21} as

$$\Gamma_{21} = n_k^{-2} \sum_{i=l}^{N} a_i y_i^{*}{}', \quad (44)$$

derived from Muthén (1989).

The multisample fit function, *G*, is a weighted function of *k* single sample fit functions F_k, where $G = \sum_{k=1}^{K} \left(\frac{n_k}{N} \right) F_k$, and where each *k*th sample has *p* measured variables yielding $p^*=p(p+3)/2$ unique variances, covariances, and sample means. Under the null hypothesis, the associated test statistic

$$T_{\mathrm{ADF}} = (N-1)\hat{G}_{\mathrm{ADF}} = (N-1)\sum_{k=1}^{K} \left(\frac{n_k}{N} \right) \hat{F}_{\mathrm{ADF}(k)} \quad (45)$$

asymptotically follows a central χ^2 distribution with the corresponding number of *k*-1 *df* in the RMM approach.

The ADF method has several advantages, yet also some disadvantages (Bollen, 1989). One main advantage is that it requires only minimal assumptions about the distributions of the observed variables. Simulation research with nonnormal data shows that the ADF test statistic is relatively unaffected by distributional characteristics (West, Finch, & Curran, 1995). Monte Carlo experiments have demonstrated that with large sample sizes (e.g., $N \geq 5{,}000$) T_{ADF} yields Type I error rates at the nominal level (Chou, Bentler, & Satorra, 1991; Curran et al., 1996; Hu et al, 1992; Muthén & Kaplan, 1992).

However, with large models and small to moderate sample sizes, ADF leads to high rates of non-convergence and improper solutions and to inflated Type I error rates associated with inflated T_{ADF} values when models do not converge. In addition, there is research suggesting that this estimator performs less well at the small to moderate sample sizes that typify much of psychological research. Therefore, some modifications of ADF are discussed in the following sections.

3. *Satorra and Bentler Scaled Chi-square statistic*

Another strategy to control for nonnormality and potentially small samples is to estimate model parameters using ML and then assess data-model fit using a test statistic that has been corrected. Satorra and Bentler (1988) developed a major modification of standard normal theory goodness-of-fit tests such as T_{ML} to yield distributional behavior that should more closely approximate a chi-square variate. The modification to T_{ML} is generically referred to as the SB statistic, T_{SB}, which has been incorporated into the EQS program (Bentler, 1996). EQS code to obtain the test statistic is provided in Appendix A.3.

The SB statistic corrects the normal theory chi-square by a constant, a scalar value that is a function of the model implied residual weight matrix, the observed multivariate kurtosis, and the model degrees of freedom. Define $\dot{\hat{\boldsymbol{\sigma}}}$ as the $p^* \times q$ matrix of partial derivatives of the p^* elements in $\hat{\boldsymbol{\sigma}}$ with respect to the q model parameters (i.e., the Jacobian matrix), evaluated at the final model parameter estimates. Let $\mathbf{W}$ be the symmetric $p^* \times p^*$ matrix of unique fourth-order moments obtained by $\hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}$, and let

$$\hat{\mathbf{U}} = \mathbf{W} - [\mathbf{W}\dot{\hat{\boldsymbol{\sigma}}}(\dot{\hat{\boldsymbol{\sigma}}}'\mathbf{W}\dot{\hat{\boldsymbol{\sigma}}})^{-1}\dot{\hat{\boldsymbol{\sigma}}}'\mathbf{W}] \tag{46}$$

be the residual weight matrix of those inverted fourth-order moments. Then,

$$T_{\text{SB}} = [(Kp^* - q)/\text{tr}(\hat{\mathbf{U}}\,\mathbf{S}_{\mathbf{Y}})]T_{\text{ML}}, \tag{47}$$

where $\mathbf{S}_{\mathbf{Y}} = \hat{\Gamma}$. The asymptotic distribution of T_{SB} is generally unknown; however, when H_0 is true, its first moment matches a central $\chi 2$ distribution with $kp^* - q$ *df*, which is k–1 *df* in the

RMM approach. It has been implemented to deal with various types of initial test statistics T, whether based on normal, elliptical or heterogeneous kurtosis theory, and has also been extended to various correlation structure models.

4. *Yuan and Bentler's two corrected test statistics*

ADF's ability to yield a correct test statistic under nonnormal conditions at large sample sizes inspired Yuan and Bentler (1997, 1999) to develop corrections to T_{ADF} for small sample sizes. Yuan and Bentler (1999) proposed some modified ADF test statistics whose distributions are approximated by an F distribution. They proposed

$$T_{\text{YB1}} = T_{\text{ADF}} / (1 + T_{\text{ADF}}/\text{N}), \tag{48}$$

which follows a central χ^2 distribution with the same model *df* as T_{ADF} (when H_0 is true). In EQS (Bentler, 1995), it is known as "Yuan-Bentler corrected AGLS test statistics". As sample size gets large, T_{YB1} becomes similar to T_{ADF}.

Further motivated to improve small sample performance associated with T_{ADF}, Yuan and Bentler (1999) proposed another modification to the ADF test statistic, appealing to Fisher's F distribution. They offered a transformation of T_{ADF} based upon the logic of the transformation applied to Hotelling's T^2 statistic in MANOVA. Observing that T^2 is a quadratic form, similar in structure to the ADF fit function, they proposed rescaling T_{ADF} to an F-distributed statistic,

$$T_{\text{YB2}} = [\text{N} - (kp^* - q)]/[(\text{N} - 1)(kp^* - q)]T_{\text{ADF}}, \tag{49}$$

with numerator and denominator *df* of $kp^* - q$ and $\text{N} - (kp^* - q)$ respectively, where $p^*=p(p+3)/2$. In the RMM approach, the numerator and denominator *df* for T_{YB2} become k-1 and N- (k-1).

The empirical study of Yuan and Bentler (1999) shows that the distributions of the modified ADF statistics are more closely approximated by F distributions than the original

ADF statistic when referred to a chi-square distribution. The Monte Carlo simulation to investigate the small sample performance of T_{ADF}, T_{YB1}, and T_{YB2} finds that T_{YB1} and T_{YB2} maintain adequate control of Type I error rates as compared to T_{ADF}, and yield adequate power with both normal and nonnormal data.

5. *Bartlett-corrected statistic*

Within the context of exploratory factor analysis with m latent constructs, Bartlett (1950) suggested that at small sample sizes the correction to the ML test statistic

$$T_{BC} = (N - p/3 - 2m/3 - 11/6)F_{ML} \quad (50)$$

more closely follows a central $\chi 2$ distribution (with $kp^* - q$ df) than the usual T_{ML} statistic. This adjusted statistic is equivalent to applying a multiplicative correction to T_{ML} (or to any test statistic) of the form

$$c = 1 - [(2p + 4m + 5)/6(N - 1)]. \quad (51)$$

Fouladi (1998, 2000) applied this correction factor to T_{ML} to improve its small sample performance with normal data. In addition, she investigated the m-factor correction applied to T_{SB} with nonnormal data and reported that this scaling correction can be effective in controlling Type I error rate under some experimental conditions (Fouladi, 1999). Nevitt (2000) also found that a Bartlett-corrected statistic performed best to evaluate model fit in small samples.

Chapter III

Study Methods

Test Statistics Examined

ANOVA alternatives that have evidence showing better performance under variance heterogeneity or nonnormality than traditional ANOVA *F* test based on previous investigation were included in the study. Thus, for ANOVA-based methods, this included the incorporation of trimmed means and Winsorized variances into the ANOVA *F* test (for reference), as well as into procedures by Welch, Brown and Forsyth, Alexander and Govern, and James. For the RMM methods, this included SMM with ML, SMM with ADF, SMM using the Satorra and Bentler Scaled Chi-square statistic, SMM using Yuan and Bentler's two corrected test statistic, and SMM using the Bartlett-corrected statistic. In sum, the study examined eleven test statistics, including the following: five ANOVA methods (F, W, BF, A, and U) and six SMM methods (T_{ML}, T_{ADF}, T_{SB}, T_{YB1}, T_{YB2}, and T_{BC}).

Data Generation and Modeling

All simulated data were generated by SAS (1990). The test statistics for Welch, Brown and Forsyth, James, and Alexander and Govern incorporated trimmed means and Winsorized variances into procedures and were all calculated and analyzed using SAS. The SAS program of James's second-order test for testing the hypothesis of equal means was developed by Oshima and Algina(1992a). The adjustment of trimmed means and Winsorized variances was applied to the program. The T_{ML}, T_{ADF} and T_{SB} test statistics were obtained directly from EQS 6.1, based on which the test statistics of T_{YB1}, T_{YB2}, and T_{BC} were computed.

Each replication generated *k* independent samples with sample sizes as described below. For each cell of the design 1000 replications were generated, tallying the rate of false rejections (i.e., Type I errors). All population means were set equal to zero when Type I error probabilities were assessed. All tests were omnibus tests of the overall null hypothesis conducted at the α= .01, .05 or .10 level. Type I error robustness was evaluated using the Bradley's liberal criterion (Bradley, 1978). Accordingly, the corresponding robustness intervals were (0.5%, 1.5%), (2.5%, 7.5%) and (5%, 15%) at the nominal level of .01, .05 and .10 respectively. Note that for all tests the nominal tabled critical value was used rather than one empirically derived from the sampling distribution of the replications; this was done in order to approximate the hypothesis testing procedures that researchers actually do in practice.

<u>Conditions Manipulated</u>

The simulation design systematically manipulated four conditions: number of groups (*k*), degree of nonnormality, degree of variance heterogeneity, and sample size. For all cells of this design, all of the aforementioned ANOVA-based and SMM-based test statistics were computed for each replication. Table 1 summarized the conditions manipulated for the simulation study.

K, number of groups. The conditions investigated will be *k*=2, 3, and 4 independent samples.

Degree of nonnormality. Data for each sample will have the same distributional shape, coming from one of three populations. The skew and kurtosis manipulated were (0, 0), (0, 3), and (3, 21), two of which were used in other simulation studies (e.g., Nevitt & Hancock, 2004). That is, distribution 1 is multivariate normal with both univariate skew and kurtosis set equal to 0. Distribution 2 represents an elliptical distribution — data are nonnormal but symmetric with univariate skew of 0 and kurtosis of 3. Distribution 3 is nonnormal and asymmetric with univariate skew of 3 and kurtosis of 21. Micceri (1989) studied over 440 distributions from journal articles in various applied fields as well as national, statewide and districtwide tests, indicating two patterns of the distributions: 1) gain scores tended to be fairly symmetric and 2) criterion/ mastery tests tended to be extremely asymmetric. The study showed that both general ability and achievement measures tended to distribute symmetrically while psychometric measures exhibited greater asymmetry. Our second and third distribution shapes approximate the two distribution patterns in some way. In order to restrict the scope of the dissertation and make the dissertation manageable, the distribution shape with certain amount of skewness but no kurtosis is not included in the study, such as the distribution with skewness and kurtosis of (3, 0).

Simulated data for the second distribution were generated in SAS (1990) by following Ramberg and Schmeiser's (1979) power transformation of uniform variables to obtain a generalized lambda distribution. Data for the extreme distribution 3 were also generated in SAS (1990) using the programming described by Nevitt and Hancock (1999) that follows the Fleishman's (1978) polynomial transformation.

Sample size and variance heterogeneity. The specific sample size and variance conditions are presented below for each of the k conditions. (Note, since sample sizes vary across different k, comparisons should not be made across groups.)

K=2. The scores in the first sample is multiplied by a constant, so that the standard deviation ratio σ_1/σ_2 has a value of 1, 2.5 or 4. The total sample size, $N=n_1+n_2$, is fixed at N=20, 100, and 500, and the ratio of sample sizes, $n_1:n_2$, is 4:1, 1:1, and 1:4. For example, when N=100, the three pairs of sample sizes are 80:20, 50:50 and 20:80; each of these conditions in turn is crossed with the standard deviation ratio conditions to create 24 non-redundant conditions.

K=3. The ratio of σ_1, σ_2 and σ_3 is made to follow 1:1:1, 1:2.5:4, or 4:2.5:1. The total sample size, $N=n_1+ n_2+ n_3$, is fixed at 90, 180, and 900, and the ratio of sample sizes, $n_1:n_2:n_3$, is 4:2.5:1, 1:1:1, and 1:2.5:4. For example, when N=90, the three pairs of sample sizes are 48:30:12, 30:30:30, and 12:30:48; each of these conditions in turn is crossed with the standard deviation ratio conditions to create 15 non-redundant conditions.

K=4. The ratio of σ_1, σ_2, σ_3, and σ_4 will be made to follow 4:3:2:1, 1:1:1:1, or 1:2:3:4. The total sample size, $N=n_1+n_2+n_3+n_4$, is fixed at 120, 240, and 1200, and the ratio of sample sizes, n_1:n_2:n_3:n_4, is 5:4:2:1, 1:1:1:1, and 1:2:4:5. For example, when N=120, the three pairs of sample sizes are 50:40:20:10, 30:30:30:30, and 10:20:40:50; each of these conditions in turn is crossed with the standard deviation ratio conditions to create 15 non-redundant conditions.

Table 1.
Conditions Manipulated for the Simulation Study

<table>
<tr><td colspan="2">D</td><td>(0,0)
(0,3)
(3,21)</td><td>α</td><td>.01
.05
.10</td><td>K</td><td>2
3
4</td></tr>
<tr><td colspan="2">k=2</td><td colspan="2">k=3</td><td colspan="3">k=4</td></tr>
<tr><td>σ_1/σ_2</td><td>1
2.5
4</td><td>$\sigma_1/\sigma_2/\sigma_3$</td><td>4:2.5:1
1:2.5:4
1:1:1</td><td>$\sigma_1/\sigma_2/\sigma_3/\sigma_4$</td><td colspan="2">4:3:2:1
1:2:3:4
1:1:1:1</td></tr>
<tr><td rowspan="3">n_1: n_2</td><td>4:16
16:4
10:10</td><td rowspan="3">n_1: n_2: n_3</td><td>48:30:12
30:30:30
12:30:48</td><td rowspan="3">n_1: n_2: n_3: n_4</td><td colspan="2">50:40:20:10
30:30:30:30
10:20:40:50</td></tr>
<tr><td>20:80
80:20
50:50</td><td>96:60:24
60:60:60
24:60:96</td><td colspan="2">100:80:40:20
60:60:60:60
20:40:80:100</td></tr>
<tr><td>100:400
400:100
250:250</td><td>480:300:120
300:300:300
120:300:480</td><td colspan="2">500:400:200:100
300:300:300:300
100:200:400:500</td></tr>
</table>

Note. D is the distributional shapes. (0, 0)= normal distribution with skewness and kurtosis of (0, 0); (0, 3)= elliptically symmetric nonnormal distribution with skewness and kurtosis of (0, 3); (3, 21)= asymmetric nonnormal distribution with skewness and kurtosis of (3, 21).
K indicates the number of groups and α is the nominal level.

Design and Execution

To examine the Type I error rates, 1000 replications were conducted. The study fully crossed the eleven test statistics with different ratios of sample sizes, three pairs of variance ratios, three distributional forms, three groups as well as three Type I error rate levels, yielding (24+15+15)×3×3=486 "between" cells for each of the eleven test statistics. For each cell, independent data sets were generated from the associated distribution form, group variances and sample sizes, and were tested to examine the mean equality. The same simulated data sets drawn from the same distributional form, sample sizes and group variances were used to calculate the eleven test statistics. The approach was used to reduce the biasing effect of sampling random error in making comparisons of the results for the test statistics. In addition, parameter estimates provided by three RMM test statistics of T_{ML}, T_{ADF} and T_{SB} were tracked for all cells in the Type I error portion of the study.

Power Analysis

For methods shown to have the best control over Type I error, the power of ANOVA-based methods and the proposed RMM methods will be examined through extensive Monte Carlo simulation. For the superior test statistic(s) that yielded an observed Type I error rate that was within or below the robustness interval, a new series of 1000 replications with unequal population means was generated using the same distribution form, sample sizes, group variances and number of groups. Since the liberal Type I error rates indicate inflated test statistics, it is anticipated that the power estimates are inflated under nonnull conditions and are not comparable to the power estimates from the test statistics providing reasonable Type I error rates. Thus, the study conditions under which the test statistics yielded empirical Type I error rates above the upper bound of Bradley's (1978) liberal criterion were eliminated from the power analysis. However, the study conditions under which the test statistics yielded empirical Type I error rates below the lower bound of the Bradley's (1978) liberal criterion were retained for the power analysis, because the test statistics could potentially provide acceptable power in this situation.

Bradley (1978) argued that a quantitative definition of robustness can be achieved by stating the range of empirical Type I error rates for which the test would be considered robust. He identifies three different Type I error rates of robustness as fairly stringent, moderate, and liberal. Bradley's liberal criterion is defined as the situation when the absolute value of empirical Type I error rate minus Type I error rate is less than or equal to Type I error rate divided by 2. A test fulfills his liberal criterion at a Type I Error rate of .10 if the empirical Type I error rate is between .05 and .15. For a Type I error rate of .05, the liberal criterion would require the empirical Type I error rate to lie between .025 and .075. Similarly, to meet the Bradley's liberal criterion for a Type I error rate of .01, the empirical Type I error rate should lie between .005 and .015. It's worth noting, though, that the robustness interval for the nominal level of .01 is smaller than that for the nominal level of .10, leading to more difficulty in rejecting H_0.

For each of the same conditions in the Type I error study there will be additional conditions created in which the null hypothesis is false. Two such conditions will be examined for each *K* condition. For k=2, mean differences will be induced that yield standardized effect sizes of d=.2 and .8 in both the homogeneous and heterogeneous variance cases. For k=3 and 4, mean differences will be induced that yield standardized effect sizes of f=.1 and .4 in both the homogeneous and heterogeneous variance cases as well. For each replication the number of rejections will be tallied, and relative power will be examined within each condition. Note again that not all methods may participate in this portion of the study.

Formulas were algebraically derived for generating the mean differences at specified level of Cohen's (1977) t test effect size index *d*,

$$d = | u_1 - u_2 | / \sigma \tag{52}$$

where u_1 and u_2 are population means and σ is the standard deviation of scores typically assumed to be homogeneous across both populations (Cohen, 1977). This formula was applied to obtain the population mean difference given a specified d for the two-group test of mean equality. When k is greater than two, the formula of Cohen's (1977) ANOVA effect size index f was utilized,

$$f = \frac{\sigma_m}{\sigma}, \tag{53}$$

$$\sigma_m = \sqrt{\frac{\sum_k n_k (\mu_k - \mu)^2}{N}}, \tag{54}$$

and

$$\mu = \frac{\sum_k n_k \mu_k}{N}. \tag{55}$$

It is obvious that the standard deviation of each group (σ_k) is equivalent to the population standard deviation (σ) in the formula for d and f shown above when population variances are homogenous. However, when population variances are heterogeneous, the theoretical power cannot be estimated because the population variance (σ^2) is unknown (Glass et al., 1972). Thus, a sample size weighted approximation was used instead,

$$\bar{\sigma}^2 = \frac{\sum_k n_k \sigma_k^2}{N}. \tag{56}$$

Note that the any-pairs power rate was assessed for these omnibus tests, that is , the probability of detecting at least one true pairwise difference. There is evidence showing that this approximation generates estimates generally corresponding well with the empirical power of the ANOVA (Budescu, 1982). Tomarken and Serlin (1986) have also used this approximation to compare ANOVA alternatives, which performs reasonably well. Our study is similar in the way that its purpose is to compare different approaches rather than to evaluate the effects of variance heterogeneity on the relation between the theoretical and empirical power of the ANOVA itself.

Three configurations of population means were used for the two-group, three-group as well as four-group independent samples designs respectively. When k=2, the first group always have the positive extreme mean and the second group has a mean of zero (e.g. μ_1=0.8, μ_2=0). When k=3, the pattern is that the group means are equally spaced with the adjacent means differing by the same amount and the middle group with a mean of zero (e.g. μ_1=0.4, μ_2=0, μ_3= -0.4). When k=4, a "two in the middle" pattern was used, whereby the two middle groups with means of zero were halfway between two extreme groups (e.g. μ_1=0.4, μ_2=0, μ_3=0, μ_4= -0.4). The ordering of the means paralleled with the ordering of groups, with the first group associated with the positive mean and the last group associated with the smaller mean. The computation of population mean difference was verified using Mathematica 5.0 (2003).

Chapter IV

Results

Non-convergence

Rates of non-convergence for the T_{ML}, T_{ADF} and T_{SB} test statistics were tracked for all cells in the Type I error portion of the study, which were summarized in Table 2. Since the test statistics T_{YB1}, T_{YB2} are derived from the T_{ADF} and the test statistic T_{BC} is based on T_{ML}, these test statistics should produce the same non-convergence rates correspondingly. Overall, almost no nonconvergence occurred when degrees-of-freedom was more than one, that is, when $k \geq 2$. When $k = 2$ with moderate and large sample sizes, no nonconvergence occurred. Only when k=2 with small sample sizes, nonconvergence occurred with the highest rate of 1.7%. The ML and SB estimations mirrored each other closely, which yielded non-convergence rates of no more than 1.7%, 0.6% and 1.0% across distributions with skewness and kurtosis of (0, 0), (0, 3), and (3, 21) respectively. With respect to the ADF estimation, the T_{ADF} yielded 0% nonconvergence across all conditions.

Table 2.
Rates of Non-convergence (%) for k=2

n_1/n_2	σ_1/σ_2	(0, 0)			(0, 3)			(3, 21)		
		T_{ML}	T_{ADF}	T_{SB}	T_{ML}	T_{ADF}	T_{SB}	T_{ML}	T_{ADF}	T_{SB}
4:16	1	0.3	0.0	0.3	0.2	0.0	0.2	0.3	0.0	0.3
	2.5	0.0	0.0	0.0	0.1	0.0	0.1	0.9	0.0	0.9
	4	0.0	0.0	0.0	0.4	0.0	0.4	0.7	0.0	0.7
16:4	1	0.5	0.0	0.5	0.2	0.0	0.2	0.9	0.0	0.9
	2.5	1.4	0.0	1.4	0.5	0.0	0.5	0.3	0.0	0.3
	4	1.7	0.0	1.7	0.6	0.0	0.6	0.7	0.0	0.7
10:10	1	0.0	0.0	0.0	0.5	0.0	0.5	0.0	0.0	0.0
	2.5	0.1	0.0	0.1	0.0	0.0	0.0	0.3	0.0	0.3
	4	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0

Note. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using SB Scaled Chi-square statistic.

Type I errors

There are three different situations, which are the "positive condition" (unequal sample sizes with larger sample sizes paired with larger variances), the "negative condition" (unequal sample sizes with larger sample sizes paired with smaller variances), and the condition of equal sample sizes. Since there are three different distributional forms, three different levels of significance and three different group situations, 27 (3×3×3) tables of Type I error rates in total were created for each group situation at each nominal level with each distributional form.

Table 3 to Table 11 present the results of the Type I error rates under three group variance ratios and under three different sample size ratios for the normal distribution. Among the nine tables, Tables 3, 4 and 5 present the results of the Type I error rates for k=2 at the nominal levels .01, .05 and .10. Tables 6, 7, and 8 present the results of the Type I error rates for k=3 at nominal levels .01, .05 and .10, while Tables 9, 10, and 11 present the results of the Type I error rates for k=4 at nominal levels .01, .05 and .10.

Table 12 to Table 20 present the results of the Type I error rates under three group variance ratios and under three different sample size ratios for the elliptical distribution with univariate skew of 0 and kurtosis of 3. Among the nine tables, Tables 12, 13 and 14 present the results of the Type I error rates for k=2 at the nominal levels .01, .05 and .10. Tables 15, 16, and 17 present the results of the Type I error rates for k=3 at nominal levels .01, .05 and .10, while Tables 18, 19, and 20 present the results of the Type I error rates for k=4 at nominal levels .01, .05 and .10.

Table 21 to Table 29 present the results of the Type I error rates under three group variance ratios and under three different sample size ratios for the nonnormal distribution with univariate skew of 3 and kurtosis of 21. Among the nine tables, Tables 21, 22 and 23 present the results of the Type I error rates for k=2 at the nominal levels .01, .05 and .10. Tables 24, 25, and 26 present the results of the Type I error rates for k=3 at nominal levels .01, .05 and .10, while Tables 27, 28, and 29 present the results of the Type I error rates for *k*=4 at nominal levels.01, .05 and .10.

Normal distribution

K=2

$\underline{\alpha = .01}$. At the "positive conditions" when sample size was small with moderate variance heterogeneity, the test statistics of T_{ADF} and T_{YB2} provided robust Type I error rates, while the rest of the methods all fell below the lower boundary of the robust range; when the variance heterogeneity became large, the test statistics of U, T_{ML}, T_{ADF}, T_{SB} and T_{YB2} provided robust Type I error rates. When sample size became moderate, all ANOVA alternative methods and SMM approaches performed well. With large sample sizes at the "positive conditions", the ANOVA *F* alternative methods continued to provide robust Type I error rates; the SMM approaches, however, delivered inflated Type I error rates.

At the "negative conditions", the ANOVA alternative methods were robust across different sample sizes. The T_{ADF}, T_{YB1}, and T_{YB2} test statistics provided inflated model rejection rates that decreased with increasing sample sizes and became robust when sample sizes were large. The T_{ML}, and T_{SB} test statistics delivered most robust results, while the test statistic T_{BC} provided Type I error rates lower than the robust range when sample sizes were small.

When sample sizes were equal, all ANOVA-based methods and most RMM approaches were robust across different sample sizes and variance heterogeneity ratios. However, the T_{ADF} test statistic provided inflated Type I error rates when sample size is small; the T_{ML}, T_{SB} and T_{YB2} test statistics provided some control but presented inflated Type I error rates at small sample size with homogeneous variances. The T_{YB1} and T_{BC} test statistics were robust with equal sample sizes.

Table 3.
Type I Error Rates (%) for Normal Distribution with k=2 at α=.01.

n_1: n_2	σ_1/σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	0.6	0.7	0.7	0.9	0.8	0.8	**3.1**	0.8	1.3	**2.0**	0.7
	2.5	**7.2**	0.6	0.6	1.0	0.6	0.7	**4.5**	0.7	**2.9**	**3.1**	**0.4**
	4	**11.3**	0.5	0.5	1.0	0.6	0.7	**4.7**	0.7	**2.9**	**3.4**	**0.4**
16:4	2.5	**0.0**	**0.2**	**0.2**	**0.3**	**0.2**	**0.2**	1.3	**0.2**	**0.2**	0.5	**0.1**
	4	**0.0**	**0.3**	**0.3**	**0.4**	0.5	0.5	1.5	0.5	**0.2**	0.6	**0.3**
10:10	1	1.0	0.9	0.9	0.8	1.0	**1.6**	**2.7**	**1.6**	0.6	**1.6**	1.1
	2.5	1.4	0.9	0.9	1.2	1.3	1.0	**2.7**	1.0	0.6	1.5	0.7
	4	1.5	0.9	0.9	1.0	1.0	0.8	**2.7**	0.8	0.5	1.3	0.5
20:80	1	0.8	1.1	1.1	1.1	1.1	1.4	**1.9**	1.4	1.5	1.5	1.3
	2.5	**11.3**	1.3	1.3	1.4	1.3	1.5	**2.1**	1.5	**2.0**	**2.0**	1.5
	4	**16.7**	1.2	1.2	1.5	1.5	**1.7**	**2.2**	**1.7**	**2.0**	**2.0**	**1.6**
80:20	2.5	**0.0**	0.7	0.7	0.7	0.7	1.3	1.3	1.3	1.2	1.2	1.3
	4	**0.0**	0.6	0.6	0.6	0.6	0.9	1.4	0.9	0.7	0.7	0.8
50:50	1	1.2	1.2	1.2	1.3	1.3	1.2	1.2	1.2	1.1	1.1	1.2
	2.5	1.1	1.0	1.0	1.1	1.1	0.7	0.9	0.7	0.7	0.7	0.7
	4	1.4	1.3	1.3	1.4	1.4	0.6	0.8	0.6	0.5	0.5	0.5
100:400	1	1.0	1.0	1.0	1.3	1.3	1.4	1.5	1.4	1.4	1.3	1.3
	2.5	**10.9**	1.3	1.3	1.3	1.3	1.2	1.4	1.2	1.2	1.2	1.2
	4	**15**	1.2	1.2	1.2	1.2	1.3	1.4	1.3	1.4	1.3	1.3
400:100	2.5	**0.0**	0.9	0.9	0.9	0.9	**2.2**	**2.2**	**2.2**	**2.1**	**2.0**	**2.2**
	4	**0.0**	1.0	1.0	1.0	1.0	**2.2**	**2.2**	**2.2**	**2.2**	**2.2**	**2.2**
250:250	1	0.5	0.5	0.5	0.6	0.6	1.2	1.2	1.2	1.2	1.1	1.2
	2.5	0.6	0.6	0.6	0.6	0.6	0.9	0.9	0.9	0.8	0.7	0.8
	4	0.6	0.6	0.6	0.7	0.7	0.8	1.0	0.8	0.8	0.6	0.8

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Overall, most of the ANOVA-based methods and RMM approaches provided reasonable control of Type I error rates across conditions. The Type I error rates of the ANOVA *F* test were inflated at all "negative conditions", but pushed to near zero at all "positive conditions". Specifically, the James second-order *U* statistic provided the best control of Type I error rates, following by Alexander and Govern *A,* Welch v_w and Brown and

Forsythe F^*. The T_{ML}, and T_{SB} test statistics provided the best control over Type I error rates among the RMM approaches when k=2 and α=.01.

$\underline{\alpha = .05}$. The Type I error rates of the ANOVA F test were inflated at all "negative conditions", but pushed to near zero at all "positive conditions". All the ANOVA alternative methods and RMM methods were robust when sample sizes were moderate or large across different variance heterogeneity ratios. When sample sizes were small and unequal, the ANOVA alternative methods delivered Type I error rates that mostly fell below the lower boundary of the robustness range. Excitingly, the T_{ML}, T_{SB}, as well as T_{BC} test statistics were robust across all conditions. The T_{YB1} and T_{YB2} test statistics also controlled the Type I error rate well with only one or two inflated rejection rates when sample sizes were small at the "negative conditions". The T_{ADF} test statistic, however, provided inflated Type I error rates for the "negative conditions" with small sample sizes and conditions with small but equal sample sizes.

Table 4.
Type I Error Rates (%) for Normal Distribution with *k*=2 at α =.05

n_1: n_2	σ_1/σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	4.1	**1.5**	**1.5**	**2.0**	**1.7**	4.2	**8.4**	4.2	6.1	6.9	3.9
	2.5	**1.7**	**1.1**	**1.1**	**2.1**	**1.5**	4.3	**8.8**	4.3	6.8	**7.6**	3.5
	4	**2.4**	**1.0**	**1.0**	**2.0**	**1.5**	4.2	**9.6**	4.2	**7.9**	**8.3**	3.4
16:4	2.5	**0.1**	**1.4**	**1.4**	**1.3**	**1.6**	3.9	6.2	3.9	3.7	4.5	3.4
	4	**0.0**	**2.1**	**2.1**	**2.2**	2.5	4.4	5.8	4.4	3.5	4.5	3.4
10:10	1	4.2	3.8	3.8	3.7	4.1	7.0	**8.5**	7.0	5.6	6.9	5.9
	2.5	5.6	4.3	4.3	4.3	4.6	6.3	**8.7**	6.3	5.4	6.7	5.3
	4	7.0	5.1	5.1	5.5	5.7	5.9	**8.3**	5.9	5.2	6.4	4.7
20:80	1	4.9	5.3	5.3	5.3	5.3	4.8	5.3	4.8	4.9	4.9	4.6
	2.5	**22.8**	6.5	6.5	6.5	6.4	5.0	5.7	5.0	5.5	5.5	5.0
	4	**28.6**	6.8	6.8	6.9	7.0	5.2	6.1	5.2	5.7	5.7	5.1
80:20	2.5	**0.2**	4.7	4.7	4.7	4.7	4.6	4.8	4.6	4.4	4.4	4.4
	4	**0.0**	4.2	4.2	3.8	4.2	4.7	4.9	4.7	4.4	4.4	4.4
50:50	1	4.6	4.6	4.6	4.5	4.6	5.2	5.3	5.2	4.7	4.7	4.9
	2.5	6.1	6.1	6.1	6.1	6.1	5.3	5.7	5.2	5.2	5.1	5.1
	4	7.1	6.9	6.9	6.9	6.9	6.0	6.2	5.9	6.0	6.0	6.0
100:400	1	4.5	4.9	4.9	4.9	4.9	5.6	5.6	5.6	5.6	5.6	5.6
	2.5	**22.7**	5.3	5.3	5.3	5.3	4.7	5.0	4.7	5.0	4.7	4.7
	4	**29.2**	5.2	5.2	5.2	5.2	5.0	5.0	5.0	5.0	4.9	4.9
400:100	2.5	**0.0**	3.7	3.7	3.7	3.7	6.2	6.2	6.2	6.2	6.2	6.2
	4	**0.0**	3.8	3.8	4.0	4.0	5.0	5.1	5.0	4.9	4.9	4.9
250:250	1	4.7	4.7	4.7	5.0	5.0	5.7	5.7	5.7	5.6	5.5	5.7
	2.5	4.9	5.0	5.0	5.1	5.1	6.0	6.2	6.0	6.0	6.0	6.0
	4	4.9	5.0	5.0	5.2	5.2	6.2	6.2	6.2	6.2	6.1	6.2

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .10. The behavior of the ANOVA-based methods was similar to that at the nominal level of .05 for the same distributional shape. Again, the Type I error rates of the ANOVA *F* test were inflated at all "negative conditions", but pushed to near zero at all "positive conditions". Excitingly, all the RMM approaches were robust across all conditions, except one cell provided by the T_{ADF} test statistics. The ANOVA alternative methods were also robust when sample sizes were moderate or large across different variance heterogeneity ratios. When sample sizes were small and unequal, the ANOVA alternative methods delivered Type I error rates that mostly fell below the lower boundary of the robustness range.

Table 5.
Type I Error Rates (%) for Normal Distribution with k=2 at α=.10

n_1: n_2	σ_1/ σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	7.6	**2.1**	**2.1**	**3.5**	**2.7**	8.4	11.9	8.4	10.1	10.9	8.1
	2.5	**25.2**	**1.5**	**1.5**	**3.4**	**2.5**	8.3	13.1	8.3	11.6	12.1	7.6
	4	**32.6**	**1.1**	**1.1**	**2.6**	**2.5**	8.4	13.2	8.4	11.4	12.1	7.8
16:4	2.5	**0.8**	**3.2**	**3.2**	**3.5**	**3.6**	7.8	9.3	7.8	7.7	8.2	7.1
	4	**0.5**	**4.1**	**4.1**	**4.1**	**4.7**	8.3	9.4	8.3	7.6	7.8	7.5
10:10	1	8.7	7.8	7.8	7.6	8.7	14.0	**15.3**	14.0	12.7	13.4	12.6
	2.5	10.4	9.4	9.4	9.2	9.7	11.6	13.8	11.6	10.8	11.9	10.6
	4	12.1	9.7	9.7	10.2	10.5	11.2	13.2	11.2	11.0	11.4	10.1
20:80	1	9.9	10.3	10.3	10.2	10.3	9.0	9.7	9.0	9.2	9.2	8.8
	2.5	**29.8**	11.1	11.1	11.1	11.1	9.1	10.2	9.1	9.4	9.4	9.1
	4	**37.4**	12.1	12.1	12.2	12.2	9.4	10.5	9.4	10.0	10.0	9.3
80:20	2.5	**0.6**	10.3	10.3	10.2	10.5	9.2	9.6	9.2	9.1	9.1	9.1
	4	**0.3**	10.2	10.2	10.2	10.2	8.9	9.2	8.9	8.9	8.8	8.9
50:50	1	10.1	10.2	10.2	10.2	10.3	10.4	10.6	10.4	10.3	10.3	10.4
	2.5	9.9	10.0	10.0	10.0	10.1	10.6	10.9	10.5	10.5	10.4	10.4
	4	10.3	10.2	10.2	10.3	10.3	10.7	11.4	10.6	10.7	10.7	10.6
100:400	1	10.4	10.3	10.3	10.5	10.4	10.8	10.9	10.8	10.9	10.7	10.8
	2.5	**31.5**	10.4	10.4	10.5	10.4	11.1	11.2	11.1	11.2	10.9	11.0
	4	**37.3**	10.1	10.1	10.1	10.1	10.8	11.1	10.8	11.1	10.8	10.8
400:100	2.5	**1.0**	8.7	8.7	8.8	8.8	12.2	12.3	12.2	12.2	12.0	12.2
	4	**0.3**	8.0	8.0	8.1	8.1	11.4	11.5	11.4	11.3	11.2	11.4
250:250	1	8.9	8.9	8.9	9.2	9.2	11.5	11.6	11.5	11.5	11.4	11.5
	2.5	10.3	10.4	10.4	10.5	10.5	11.5	11.5	11.5	11.5	11.5	11.5
	4	10.2	10.3	10.3	10.5	10.3	11.0	11.0	11.0	11.0	11.0	11.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U.* T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

K=3

<u>*α*= *.01*</u>. All RMM approaches except the T_{ADF} test statistic were robust at all conditions, superior to all ANOVA-based methods. The Welch v_w, the Brown and Forsythe *F**, the Alexander and Govern *A*, the James second-order *U* as well as the T_{ADF} test statistics delivered inflated Type I error rates at “negative conditions” when sample size was small. The Brown and Forsythe *F** test generally provided inflated Type I error rates at both “negative conditions” and “positive conditions” across different sample sizes; so did conditions with equal sample sizes and heterogeneous variances. Again, the Type I error rates of the ANOVA F test were inflated at all “negative conditions”, but were pushed to near zero at all “positive conditions”. Even when sample sizes were equal, the ANOVA *F* test provided inflated Type I error rates with heterogonous variances.

Table 6.
Type I Error Rates (%) for Normal Distribution with *k*=3 at *α* =.01

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	**3.0**	1.5	**2.7**	1.5	1.5	0.4	0.8	0.4	0.4	0.4	0.3
	1:1:1	0.9	1.1	0.9	1.2	1.2	0.6	0.8	0.6	0.6	0.6	0.6
12:30:48	4:2.5:1	**9.9**	**1.8**	**1.8**	**1.8**	**1.7**	1.0	**1.7**	1.0	1.1	1.1	0.8
	1:2.5:4	**0.2**	1.3	**1.7**	1.3	1.3	0.6	1.1	0.6	0.7	0.7	0.6
	1:1:1	0.9	**1.8**	1.4	**1.7**	1.8	0.7	1.1	0.7	0.8	0.8	0.7
60:60:60	1:2.5:4	1.3	0.9	1.3	0.9	0.9	1.0	1.3	1.0	1.0	1.1	0.9
	1:1:1	0.8	1.1	0.8	1.2	1.2	0.7	0.7	0.7	0.7	0.7	0.7
24:60:96	4:2.5:1	**9.2**	0.9	**1.7**	0.9	0.9	0.9	1.5	0.9	1.2	1.3	0.9
	1:2.5:4	0.5	0.9	**2.1**	0.9	0.9	0.7	0.8	0.7	0.7	0.7	0.7
	1:1:1	0.5	0.6	0.7	0.8	0.5	1.0	1.2	1.0	1.1	1.1	0.9
300:300:300	1:2.5:4	**1.8**	0.5	**1.8**	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	1:1:1	0.6	0.6	0.6	0.7	0.7	1.1	1.1	1.1	1.1	1.1	1.1
120:300:480	4:2.5:1	10.1	1.1	**1.9**	1.2	1.2	0.8	0.9	0.8	0.9	0.9	0.8
	1:2.5:4	**0.2**	1.0	1.5	1.2	1.2	1.0	1.0	1.0	1.0	1.0	1.0
	1:1:1	1.1	1.3	1.3	1.3	1.3	1.1	1.2	1.1	1.1	1.2	1.1

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U.* T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler’s two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Type I error rates are computed across 1000 replications at the nominal level of .01.
Bolded values indicate rejection rate falling outside of the Bradley’s robustness interval.

α= .05 and .10. The ANOVA-based methods and the RMM approaches were robust across the conditions of sample sizes and variance ratios. As usual, the Type I error rates of the ANOVA F test were inflated at all "negative conditions", but pushed below the lower boundary of the robustness range at all "positive conditions".

Table 7.
Type I Error Rates (%) for Normal Distribution with k=3 at α =.05

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	6.4	5.6	6.2	5.4	5.6	5.0	5.4	5.0	4.9	4.9	4.7
	1:1:1	5.5	5.9	5.5	6.1	6.2	5.2	6.4	5.2	5.0	5.0	5.0
12:30:48	4:2.5:1	**20.8**	6.4	7.2	6.2	6.3	5.0	7.2	5.0	6.0	6.1	4.9
	1:2.5:4	**1.5**	3.6	6.1	3.6	3.7	4.8	5.5	4.8	4.4	4.6	4.5
	1:1:1	4.0	4.6	4.5	5.1	4.6	5.0	6.4	5.0	5.4	5.5	4.7
60:60:60	1:2.5:4	5.7	5.0	5.6	5.3	5.4	4.4	4.7	4.4	4.4	4.4	4.2
	1:1:1	3.9	4.6	3.9	4.6	4.7	4.4	5.1	4.4	4.4	4.4	4.4
24:60:96	4:2.5:1	**20.6**	4.6	6.0	4.6	4.6	6.4	7.4	6.4	7.0	7.1	6.2
	1:2.5:4	**1.7**	4.9	7.2	4.9	5.0	5.8	6.0	5.8	5.6	5.6	5.7
	1:1:1	4.1	4.8	4.1	5.1	5.0	6.3	6.9	6.3	6.6	6.6	6.3
300:300:300	1:2.5:4	5.8	4.8	5.8	5.0	5.0	4.6	4.7	4.6	4.6	4.4	4.6
	1:1:1	4.5	3.9	4.5	4.6	4.7	4.7	4.9	4.7	4.6	4.1	4.7
120:300:480	4:2.5:1	**18.8**	4.6	6.3	4.5	4.6	4.7	4.9	4.7	4.7	4.5	4.6
	1:2.5:4	**1.2**	4.4	5.2	4.6	4.6	4.7	4.8	4.7	4.7	4.5	4.7
	1:1:1	4.4	4.6	4.5	4.6	4.7	4.9	5.0	4.9	4.9	4.4	4.9

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Table 8.
Type I Error Rates (%) for Normal Distribution with k=3 at α=.10

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	10.6	10.6	10.4	10.4	10.6	10.1	11.0	10.1	9.7	10.0	9.5
	1:1:1	10.1	10.5	10.2	10.4	10.5	11.3	12.5	11.3	10.9	10.9	11.0
12:30:48	4:2.5:1	**30.0**	12.1	11.4	11.5	11.9	1.0	12.4	1.0	11.6	11.6	9.5
	1:2.5:4	**3.3**	8.1	10.6	8.0	8.1	8.9	10.3	8.9	8.9	9.1	8.7
	1:1:1	9.4	11.5	10.4	11.2	11.4	10.4	12.0	10.4	10.7	10.8	10.0
60:60:60	1:2.5:4	11.2	10.0	11.4	10.4	10.3	8.8	9.2	8.8	8.9	8.9	8.7
	1:1:1	8.6	9.5	8.7	9.5	9.5	9.6	10.1	9.6	9.4	9.5	9.4
24:60:96	4:2.5:1	**28.5**	10.2	9.4	10.1	10.1	10.8	11.2	10.8	11.0	11.0	10.6
	1:2.5:4	**3.9**	10.8	11.2	10.9	10.8	10.9	11.1	10.9	10.8	10.8	10.8
	1:1:1	10.4	10.4	10.3	10.7	10.7	11.3	11.9	11.3	11.6	11.6	11.1
300:300:300	1:2.5:4	10.6	10.5	10.6	10.5	10.5	10.6	10.6	10.6	10.6	10.3	10.6
	1:1:1	10.6	10.8	10.6	11.2	11.2	9.2	9.3	9.2	9.2	8.7	9.2
120:300:480	4:2.5:1	**27.5**	9.5	1.1	9.5	9.5	10.4	10.7	10.4	10.6	10.2	10.4
	1:2.5:4	**2.2**	8.7	1.0	8.9	8.9	10.5	10.5	10.5	10.5	10.2	10.5
	1:1:1	9.3	9.9	1.3	9.9	10.0	10.0	10.0	10.0	10.0	9.7	10.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

K=4

α= .01. The Brown and Forsythe F^* statistic generally provided inflated Type I error rates when variances were heterogeneous, while the T_{ADF} test statistic delivered inflated Type I error rates when sample size was small. The rest of the test statistics all seemed to provide robust Type I error rates with only a couple of sporadic non-robust cells. The Type I error rates of the ANOVA F test were inflated at all "negative conditions", but were pushed to near zero at all "positive conditions". Even when sample sizes were equal and variances were heterogeneous, the ANOVA F test provided inflated Type I error rates. Overall, the T_{YB2}, the Welch v_w, and the James second-order U statistics best controlled the Type I error rates, followed by the T_{ML}, T_{SB}, T_{YB1} and T_{BC} test statistics.

Table 9.
Type I Error Rates (%) for Normal Distribution with k=4 at α =.01

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	**2.2**	1.4	**1.8**	**1.6**	1.5	1.0	**1.8**	1.0	1.0	1.0	1.0
	1:1:1:1	1.2	1.4	1.2	1.5	1.5	1.0	**1.6**	1.0	1.0	1.0	0.9
10:20:40:50	4:3:2:1	**11.2**	1.5	1.5	1.1	1.5	0.5	**2.3**	0.5	0.8	0.8	0.5
	1:2:3:4	**0.4**	0.9	**1.8**	1.1	0.8	1.2	**2.2**	1.2	1.5	1.5	1.2
	1:1:1:1	0.6	1.5	0.6	1.4	1.1	1.0	**2.3**	1.0	1.2	1.2	0.8
60:60:60:60	1:2:3:4	**1.7**	1.1	1.4	1.1	1.1	0.8	0.9	0.8	0.8	0.8	0.7
	1:1:1:1	0.9	1.0	0.9	1.1	1.1	0.7	0.9	0.7	0.7	0.7	0.7
20:40:80:100	4:3:2:1	**13.7**	1.0	**2.4**	0.8	1.0	0.6	**1.7**	0.6	1.4	1.4	0.6
	1:2:3:4	**0.1**	1.2	**2.7**	1.1	1.3	0.7	0.8	0.7	0.7	0.7	0.7
	1:1:1:1	0.9	1.1	0.7	1.3	1.2	0.9	1.5	0.9	1.2	1.0	0.9
300:300:300:300	1:2:3:4	**1.6**	0.5	**1.6**	0.6	0.6	1.0	1.0	1.0	1.0	0.9	1.0
	1:1:1:1	0.6	0.5	0.6	0.6	0.6	1.3	1.5	1.3	1.3	1.3	1.3
100:200:400:500	4:3:2:1	**10.4**	1.0	**1.9**	1.1	1.1	1.0	1.1	1.0	1.1	0.7	1.0
	1:2:3:4	**0.2**	0.5	1.3	0.8	0.8	**1.6**	**1.6**	**1.6**	**1.6**	1.2	**1.6**
	1:1:1:1	0.8	0.9	0.8	1.0	1.0	1.2	1.2	1.2	1.2	1.1	1.2

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .05 and .10. The ANOVA alternative methods and the RMM approaches were almost robust across all the conditions of sample sizes and variance ratios. As usual, the Type I error rates of the ANOVA F test were inflated at all "negative conditions", but pushed below the lower boundary of the robustness range at all "positive conditions".

Table 10.
Type I Error Rates (%) for Normal Distribution with k=4 at α =.05

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	7.4	5.2	6.8	5.5	5.5	4.4	6.0	4.4	4.3	4.2	4.2
	1:1:1:1	5.1	5.5	5.2	5.7	5.9	6.0	7.5	6.0	5.6	5.4	5.6
10:20:40:50	4:3:2:1	**21.8**	6.8	6.2	4.8	5.8	4.1	6.7	4.1	5.2	5.1	4.0
	1:2:3:4	**1.5**	5.3	7.4	5.3	5.2	4.8	5.8	4.8	4.9	4.8	4.8
	1:1:1:1	4.1	6.3	4.3	5.6	5.9	5.0	6.9	5.0	5.3	5.2	4.8
60:60:60:60	1:2:3:4	5.2	4.2	4.8	4.4	4.4	4.8	5.0	4.8	4.7	4.6	4.7
	1:1:1:1	4.0	4.1	4.0	4.2	4.2	4.9	6.1	4.9	4.7	4.7	4.7
20:40:80:100	4:3:2:1	**24.4**	7.0	**7.8**	6.3	7.1	5.8	6.9	5.8	6.3	6.2	5.7
	1:2:3:4	**1.9**	4.7	7.4	5.1	5.1	5.6	6.3	5.6	5.6	5.6	5.6
	1:1:1:1	4.9	5.9	4.3	6.0	5.9	6.5	**7.7**	6.5	6.8	6.7	6.4
300:300:300:300	1:2:3:4	6.4	4.2	6.4	4.4	4.4	5.5	5.6	5.5	5.5	5.0	5.5
	1:1:1:1	4.5	4.3	4.5	4.6	4.5	5.3	5.3	5.3	5.3	4.9	5.3
100:200:400:500	4:3:2:1	**23.7**	4.5	6.4	4.5	4.6	6.3	6.5	6.3	6.4	5.8	6.3
	1:2:3:4	**0.7**	3.8	6.2	4.3	4.4	6.0	6.1	6.0	6.0	5.6	6.0
	1:1:1:1	4.1	4.3	4.4	4.6	4.7	6.2	6.5	6.2	6.4	5.8	6.2

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Table 11.
Type I Error Rates (%) for Normal Distribution with k=4 at α =.10

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	11.8	10.1	11.4	10.1	10.3	10.4	11.8	10.4	10.6	10.2	9.8
	1:1:1:1	9.5	10.0	9.7	10.1	10.5	11.8	12.6	11.8	11.8	11.7	11.7
10:20:40:50	4:3:2:1	**30.8**	10.6	10.8	10.3	10.5	9.4	13.1	9.4	11.7	11.2	9.0
	1:2:3:4	**3.4**	9.2	11.6	9.4	9.6	8.7	11.0	8.7	9.2	9.0	8.3
	1:1:1:1	8.7	10.2	9.2	10.0	10.2	9.6	12.0	9.6	10.5	10.5	9.4
60:60:60:60	1:2:3:4	9.1	9.9	8.8	9.9	10.1	9.2	10.1	9.2	9.1	9.1	9.2
	1:1:1:1	8.2	9.4	8.3	9.5	9.7	9.9	10.8	9.9	9.9	9.7	9.7
20:40:80:100	4:3:2:1	**32.1**	12.6	13.0	12.5	11.3	10.9	12.6	10.9	11.8	11.6	10.6
	1:2:3:4	**3.9**	11.6	11.3	11.5	12.0	10.4	11.4	10.4	10.7	10.6	10.3
	1:1:1:1	11.1	13.5	11.5	13.4	13.5	10.6	12.0	10.6	11.2	11.0	10.5
300:300:300:300	1:2:3:4	10.5	9.8	10.5	10.2	10.2	10.7	10.9	10.7	10.7	9.8	10.6
	1:1:1:1	10.5	10.0	9.5	10.5	10.5	10.5	10.6	10.5	10.5	10.0	10.5
100:200:400:500	4:3:2:1	**32.4**	10.0	10.4	10.4	10.5	11.3	11.4	11.3	11.4	11.2	11.3
	1:2:3:4	**2.7**	9.7	11.2	10.4	10.5	11.8	11.9	11.8	11.8	11.6	11.7
	1:1:1:1	9.4	10.9	9.8	11.1	11.3	12.0	12.0	12.0	12.0	11.7	12.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Elliptically Symmetric Nonnormal Distribution (0,3)

K=2

α= .01. All the ANOVA alternative methods as well as the RMM approaches were robust when sample sizes were moderate or large across different variance heterogeneity ratios. When sample sizes were small and unequal, the ANOVA alternative methods as well as the RMM approaches all delivered inflated Type I error rates at the "negative conditions"; most of which also pushed the rejection rates even lower to near zero at the "positive conditions". All ANOVA alternative methods and the RMM approaches except the T_{ADF} test statistic were robust when sample sizes were small but equal. The T_{ADF} test statistic also provided inflated Type I error rates when sample sizes were small and variances were heterogeneous. As usual, the Type I error rates of the ANOVA *F* test were inflated at all "negative conditions", but pushed to near zero at all "positive conditions". Overall, the Alexander and Govern *A* and the T_{YB2} test statistics best controlled the Type I error rates, followed by the test statistics T_{ML}, T_{SB}, T_{YB1}, Welch v_w, Brown and Forsythe *F** and James second-order *U* statistic.

Table 12.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=2 at α=.01

n_1: n_2	σ_1/ σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	1.2	**1.7**	**1.7**	**1.9**	**1.8**	**2.1**	**3.1**	**2.1**	**1.6**	**2.4**	**1.9**
	2.5	**7.1**	**2.6**	**2.6**	**3.5**	**3.1**	**2.5**	5.0	**2.5**	**3.6**	**4.1**	**2.5**
	4	**12.7**	**2.0**	**2.0**	**3.5**	**2.9**	**3.2**	**5.9**	**3.2**	**4.6**	**5.3**	**3.0**
16:4	2.5	**0.0**	**0.3**	**0.3**	**0.4**	**0.4**	**0.2**	**0.3**	**0.2**	**0.2**	**0.2**	**0.1**
	4	**0.0**	**0.1**	**0.1**	**0.1**	**0.3**	**0.1**	**0.3**	**0.1**	**0.0**	**0.1**	**0.1**
10:10	1	1.0	0.5	0.5	0.6	0.8	1.2	1.5	1.2	0.8	1.1	0.9
	2.5	**1.6**	1.0	1.0	1.2	1.4	0.7	**2.0**	0.7	0.5	0.9	0.5
	4	**1.8**	1.0	1.0	1.1	1.1	0.5	**1.8**	0.5	0.5	0.9	0.5
20:80	1	0.8	0.9	0.9	1.1	0.8	0.5	0.6	0.5	0.5	0.5	**0.4**
	2.5	**9.7**	1.0	1.0	1.0	1.0	0.6	1.2	0.6	0.9	0.9	0.6
	4	**15.8**	1.0	1.0	1.0	1.0	0.8	1.2	0.8	0.9	1.0	0.6
80:20	2.5	**0.0**	1.0	1.0	1.1	1.2	1.0	1.5	1.0	0.8	0.9	0.9
	4	**0.0**	1.0	1.0	1.1	1.4	1.0	1.0	1.0	0.8	0.9	0.9
50:50	1	0.8	0.8	0.8	0.9	1.0	0.9	0.9	0.9	0.9	0.9	0.9
	2.5	1.2	1.3	1.3	1.3	1.3	0.6	0.8	0.6	0.5	0.5	0.6
	4	1.1	0.9	0.9	0.9	0.9	0.8	1.2	0.8	0.7	0.8	0.7
100:400	1	0.5	0.5	0.5	0.5	0.5	0.9	1.1	0.9	1.0	0.9	0.9
	2.5	**10.6**	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.8
	4	**14.8**	0.9	0.9	1.0	1.0	0.9	0.9	0.9	0.9	0.9	0.9
400:100	2.5	**0.0**	0.9	0.9	0.9	1.0	0.8	0.8	0.8	0.8	0.8	0.8
	4	**0.0**	1.1	1.1	1.2	1.2	0.9	1.0	0.9	0.9	0.9	0.9
250:250	1	0.8	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9
	2.5	1.5	1.5	1.5	1.5	1.5	1.1	1.2	1.1	1.1	1.0	1.1
	4	1.3	1.3	1.3	1.3	1.3	1.1	1.1	1.1	1.0	0.9	0.9

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .05. The Type I error rates fell below the range of the robustness for all the ANOVA alternatives and RMM approaches except the T_{ADF} test statistic at the "positive conditions" when sample size was small, and were robust at all other conditions. The T_{ADF} and T_{YB2} test statistics provided inflated Type I error rates at the "negative conditions" when sample size was small, and the T_{ADF} test statistic was robust at all other conditions. Similar to the results based on the normal distribution, the Type I error rates of the ANOVA *F* test were inflated at all “negative conditions”, but pushed to near zero at all “positive conditions”.

Table 13.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), *k*=2 at α =.05

n_1: n_2	σ_1/σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	4.2	3.3	3.3	3.7	3.7	3.5	5.9	3.5	4.7	5.3	3.1
	2.5	**17.9**	3.4	3.4	4.9	4.1	5.4	**9.2**	5.4	7.5	**8.1**	5.1
	4	**24.3**	3.3	3.3	5.2	4.3	57.0	**9.5**	57.0	8.2	**8.7**	5.6
16:4	2.5	**0.4**	**1.5**	**1.5**	**1.6**	**1.8**	**2.3**	3.3	**2.3**	**1.8**	**2.4**	**1.6**
	4	**0.2**	**1.9**	**1.9**	**2.3**	2.6	**2.3**	3.4	**2.3**	**1.9**	**2.2**	**2.0**
10:10	1	5.2	4.4	4.4	4.4	4.7	4.1	5.9	4.1	3.3	4.2	3.3
	2.5	6.2	5.2	5.2	4.9	5.7	4.7	6.3	4.9	4.1	5.1	4.3
	4	7.2	5.2	5.2	5.7	6.2	5.0	7.0	5.0	4.7	6.0	4.4
20:80	1	5.2	5.5	5.5	5.5	5.5	4.6	5.3	4.6	4.7	4.7	4.2
	2.5	**21.6**	4.7	4.7	4.7	4.7	3.9	5.0	3.9	4.7	4.6	3.8
	4	**30.3**	4.7	4.7	4.7	4.7	4.2	4.9	4.2	4.6	4.5	3.8
80:20	2.5	**0.2**	5.3	5.3	5.2	5.3	4.8	5.0	4.8	4.6	4.6	4.6
	4	**0.0**	5.8	5.8	5.8	5.8	5.3	5.4	5.3	5.2	5.2	5.3
50:50	1	5.4	5.4	5.4	5.4	5.4	5.0	5.2	5.0	4.8	4.8	5.0
	2.5	5.4	5.4	5.4	5.4	5.4	4.0	4.8	4.6	4.5	4.5	4.5
	4	5.4	5.0	5.0	5.1	5.2	4.6	4.9	4.6	4.6	4.6	4.6
100:400	1	4.3	4.6	4.6	4.8	4.8	4.9	5.1	4.9	5.0	4.9	4.9
	2.5	**21.3**	4.3	4.3	4.3	4.3	5.2	5.2	5.2	5.2	5.2	5.2
	4	**26.7**	4.6	4.6	4.6	4.6	4.8	4.9	4.8	4.9	4.8	4.8
400:100	2.5	**0.2**	5.9	5.9	6.3	6.4	5.4	5.5	5.4	5.4	5.2	5.4
	4	**0.1**	5.7	5.7	6.0	6.0	5.9	5.9	5.9	5.8	5.6	5.8
250:250	1	5.5	5.5	5.5	5.5	5.8	6.0	6.0	6.0	6.0	5.8	6.0
	2.5	5.7	5.7	5.7	6.0	6.0	5.5	5.5	5.5	5.5	5.2	5.5
	4	6.0	6.0	6.0	6.2	6.2	6.0	6.2	6.0	5.9	5.5	5.9

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler’s two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley’s robustness interval.

$\underline{\alpha = .10}$. The RMM approaches provided robust Type I error rates across almost all sample sizes and variance ratios conditions. The ANOVA alternatives also yielded robust Type I error rates when sample sizes were moderate, large or equal. However, the Type I error rates fell below the range of the robustness for the Welch v_w and Brown and Forsythe F^* statistics when sample sizes were small and unequal. In addition, the Type I error rates tended to be small below the range of the robustness for the Alexander and Govern A and James second-order U statistics for unequal small samples, especially at the "positive conditions". As usual, the Type I error rates of the ANOVA F test were inflated at all "negative conditions", but pushed to near zero at all "positive conditions".

Table 14.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=2 at α=.10

n_1: n_2	σ_1/ σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	8.9	**4.2**	**4.2**	5.8	**4.8**	7.2	10.0	7.2	8.3	8.7	6.4
	2.5	**25.9**	**4.2**	**4.2**	6.3	5.9	8.9	13.6	8.9	11.3	11.8	8.2
	4	**33.7**	**4.3**	**4.3**	6.5	6.0	8.9	14.2	8.9	12.7	12.4	8.4
16:4	2.5	**1.4**	**3.4**	**3.4**	**3.9**	**4.1**	5.9	7.2	5.9	5.4	5.9	**4.7**
	4	**0.7**	**4.7**	**4.7**	5.0	5.3	5.7	6.7	5.7	5.3	5.6	5.1
10:10	1	10.3	9.7	9.7	9.6	10.0	9.1	10.3	9.1	8.0	9.0	7.9
	2.5	11.9	10.6	10.6	10.7	11.0	8.6	10.9	8.6	8.1	8.9	7.4
	4	13.2	10.9	10.9	11.2	11.4	8.8	11.0	8.8	8.6	9.5	8.2
20:80	1	9.1	10.0	10.0	10.1	10.1	10.0	10.8	10.0	10.4	10.4	9.8
	2.5	**31.7**	8.9	8.9	9.1	9.0	8.6	10.0	8.6	9.4	9.4	8.5
	4	**37.9**	9.5	9.5	9.6	9.6	8.4	10.2	8.4	9.5	9.5	8.3
80:20	2.5	**1.3**	9.1	9.1	9.1	9.1	10.0	10.3	10.0	9.9	9.8	9.9
	4	**0.4**	9.3	9.3	9.3	9.5	10.5	10.6	10.5	10.1	10.1	10.2
50:50	1	10.8	10.8	10.8	10.8	10.9	9.9	10.2	9.9	9.7	9.6	9.7
	2.5	9.9	9.8	9.8	9.8	9.9	9.5	10.1	9.5	9.5	9.5	9.4
	4	10.4	10.4	10.4	10.4	10.4	9.9	10.4	9.9	9.9	9.9	9.5
100:400	1	9.5	9.4	9.4	9.5	9.5	9.8	9.9	9.8	9.8	9.8	9.8
	2.5	**30.0**	10.4	10.4	10.4	10.4	9.8	9.9	9.8	9.9	9.7	9.8
	4	**36.8**	10.5	10.5	10.5	10.5	10.2	10.2	10.2	10.2	10.1	10.2
400:100	2.5	**0.9**	10.9	10.9	10.9	10.9	10.8	10.9	10.8	10.8	10.6	10.8
	4	**0.5**	10.8	10.8	10.9	10.9	11.3	11.3	11.3	11.3	10.8	11.3
250:250	1	11.8	11.8	11.8	11.8	11.8	10.4	10.4	10.4	10.3	10.3	10.4
	2.5	11.3	11.3	11.3	11.6	11.6	11.3	11.4	11.3	11.3	11.0	11.3
	4	11.9	11.9	11.9	12.1	12.1	11.4	11.6	11.4	11.4	11.3	11.4

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

K=3

<u>*α*= *.01*</u>. The ANOVA-based methods generally paralleled with their behavior when the distribution was normal. Some sporadic non-robust cells were observed for the ANOVA alternatives and the RMM approaches, all falling below the robustness range. Overall, the test statistics the Welch v_w, Alexander and Govern *A,* James second-order *U*, T_{ADF}, and T_{YB2} best controlled the Type I error rates, followed by the test statistics T_{ML}, T_{SB}, T_{YB1} and T_{BC}. Again, the Type I error rates of the ANOVA F test were inflated at all "negative conditions", but pushed to near zero at all "positive conditions".

Table 15.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), *k*=3 at *α* =.01

n_1: n_2: n_2	$\sigma_1/\sigma_2/\sigma_3$	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	**1.8**	1.1	**1.8**	1.1	1.1	0.7	**1.6**	0.7	0.7	11.1	0.6
	1:1:1	1.2	1.4	1.2	1.5	**1.8**	0.6	0.9	0.6	**0.3**	0.5	0.5
12:30:48	4:2.5:1	**12.4**	1.1	**2.2**	0.9	1.0	0.5	1.4	0.5	0.8	0.8	0.5
	1:2.5:4	**0.1**	0.9	**2.4**	1.0	0.9	0.8	0.9	0.8	0.6	0.7	0.8
	1:1:1	1.1	1.2	1.5	1.1	1.2	0.7	1.2	0.7	0.8	0.8	0.7
60:60:60	1:2.5:4	**1.9**	1.2	**1.8**	1.4	1.5	0.6	0.9	0.6	0.6	0.6	0.6
	1:1:1	1.2	1.3	1.2	1.3	1.3	0.7	0.7	0.7	0.7	0.7	0.7
24:60:96	4:2.5:1	**9.2**	0.9	1.0	0.8	0.9	**0.3**	1.0	**0.3**	0.7	0.7	**0.3**
	1:2.5:4	**0.3**	1.0	1.5	0.9	1.0	0.6	0.9	0.6	**0.4**	**0.4**	**0.4**
	1:1:1	0.9	0.7	0.7	0.9	0.7	**0.4**	0.7	**0.4**	0.6	0.6	**0.4**
300:300:300	1:2.5:4	1.4	1.2	1.4	1.3	1.2	1.5	1.6	1.5	1.5	1.6	1.5
	1:1:1	1.5	1.5	1.5	1.5	1.5	1.4	1.4	1.4	1.4	1.4	1.4
120:300:480	4:2.5:1	**10.0**	0.9	1.3	1.0	1.0	0.8	0.9	0.8	0.8	0.9	0.8
	1:2.5:4	**0.3**	0.6	**1.9**	0.8	0.8	1.0	1.0	1.0	1.0	1.0	1.0
	1:1:1	0.7	0.9	0.9	0.9	0.9	0.7	0.9	0.7	0.8	0.9	0.7

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U.* T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .05 and .10. The ANOVA alternative methods and the RMM approaches were almost robust across all the conditions of sample sizes and variance ratios. As usual, the Type I error rates of the ANOVA F test were inflated at all “negative conditions”, but pushed below the lower boundary of the robustness range at all “positive conditions”.

Table 16.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), *k*=3 at α=.05

n_1: n_2: n_2	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	**7.7**	6.2	7.5	6.0	6.2	5.0	5.9	5.0	4.9	4.9	4.8
	1:1:1	5.1	5.7	5.1	5.6	5.7	4.8	5.8	4.8	4.3	4.5	4.2
12:30:48	4:2.5:1	**24.4**	5.9	6.7	5.4	5.9	4.3	7.0	4.3	5.6	5.9	4.0
	1:2.5:4	**2.2**	5.9	**7.8**	5.9	61.0	4.7	5.6	4.7	4.6	4.7	4.5
	1:1:1	6.4	5.9	5.8	6.0	5.8	3.9	5.9	3.9	4.2	4.3	3.7
60:60:60	1:2.5:4	6.5	5.9	6.1	6.3	6.2	5.3	5.4	5.3	5.3	5.3	5.3
	1:1:1	5.6	6.3	5.6	6.3	6.3	4.5	5.0	4.5	4.4	4.4	4.4
24:60:96	4:2.5:1	**18.2**	3.4	4.3	4.1	3.4	4.9	5.5	4.9	5.1	5.1	4.8
	1:2.5:4	**1.4**	3.9	6.5	4.0	3.9	4.5	5.0	4.5	4.1	4.3	4.2
	1:1:1	4.0	3.8	3.6	3.9	3.9	4.3	5.3	4.3	4.5	4.5	4.0
300:300:300	1:2.5:4	4.4	4.4	4.4	4.7	4.6	5.3	5.4	5.3	5.3	5.1	5.3
	1:1:1	5.1	5.0	5.1	5.2	5.2	6.1	6.4	6.1	6.1	5.6	6.1
120:300:480	4:2.5:1	**19.2**	4.6	5.2	4.5	4.8	3.6	3.8	3.6	3.7	3.5	3.6
	1:2.5:4	**1.5**	4.8	6.9	5.1	5.1	4.7	4.8	4.7	4.7	4.5	4.7
	1:1:1	4.6	5.0	4.6	4.9	5.2	4.1	4.3	4.1	4.0	4.0	4.1

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Type I error rates are computed across 1000 replications at the nominal level of .05.
Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Table 17.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), K=3 at α =.10

n_1: n_2: n_2	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	12.2	11.5	11.6	11.0	11.4	11.1	12.1	11.1	11.1	11.1	10.7
	1:1:1	10.0	9.8	10.3	9.8	10.0	10.3	11.1	10.3	10.0	10.0	9.9
12:30:48	4:2.5:1	**33.1**	11.8	12.2	12.3	11.8	9.9	12.3	9.9	11.3	11.3	9.5
	1:2.5:4	**4.2**	12.4	11.5	12.4	12.7	9.0	9.6	9.0	8.6	8.6	8.6
	1:1:1	11.9	12.2	12.4	11.8	12.1	9.6	11.2	9.6	10.2	10.2	9.2
60:60:60	1:2.5:4	10.5	10.7	10.5	11.1	11.2	10.2	10.8	10.2	10.1	10.2	10.1
	1:1:1	10.7	10.8	10.7	10.7	10.8	10.4	11.0	10.4	10.4	10.4	10.4
24:60:96	4:2.5:1	**28.7**	9.0	8.7	8.9	9.0	9.6	11.0	9.6	10.4	10.4	9.5
	1:2.5:4	**3.2**	10.8	11.3	10.6	11.0	9.8	10.2	9.8	9.6	9.7	9.5
	1:1:1	9.3	8.9	8.0	9.3	9.2	8.8	9.8	8.8	9.0	9.0	8.7
300:300:300	1:2.5:4	8.5	8.1	8.6	8.5	8.5	10.7	10.8	10.7	10.7	10.4	10.7
	1:1:1	8.6	8.7	8.7	9.2	9.2	12.5	12.8	12.5	12.4	12.1	12.4
120:300:480	4:2.5:1	**27.6**	9.1	10.9	9.3	9.2	8.4	8.7	8.4	8.4	8.1	8.4
	1:2.5:4	**3.4**	9.0	10.9	9.3	9.5	9.3	9.5	9.3	9.3	9.0	9.3
	1:1:1	9.4	9.2	9.2	9.1	9.2	9.4	9.5	9.4	9.4	9.2	9.4

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

K=4

<u>*α= .01*</u>. The T_{ADF} and Brown and Forsythe *F** statistic generally provided most cells of inflated Type I error rates, especially when variances were heterogeneous. The rest of the test statistics all seemed to provide quite robust Type I error rates with only a couple of sporadic non-robust cells. Overall, the Alexander and Govern *A*, James second-order *U*, and T_{YB2} statistics best controlled the Type I error rates, followed by the Welch v_w, T_{YB1} T_{ML}, T_{SB}, and T_{BC} test statistics. Again, the Type I error rates of the ANOVA *F* test were inflated at all "negative conditions", but were pushed to near zero at all "positive conditions". Even when sample sizes were equal and variances were heterogeneous, the ANOVA *F* test tended to provide inflated Type I error rates.

Table 18.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=4 at α =.01

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	**2.2**	1.1	**2.2**	1.3	1.3	**1.6**	**1.6**	1.3	**1.6**	1.1	**1.6**
	1:1:1:1	1.0	0.9	1.0	1.1	1.1	**1.6**	**1.6**	0.8	**1.6**	1.4	**1.6**
10:20:40:50	4:3:2:1	**10.8**	0.9	**1.6**	1.1	1.0	0.8	0.8	0.6	0.8	0.5	0.8
	1:2:3:4	**0.3**	1.2	**1.9**	1.2	1.2	0.9	0.9	1.2	0.9	0.8	0.9
	1:1:1:1	0.6	0.8	0.6	0.9	1.0	0.8	0.9	0.8	0.9	0.6	0.8
60:60:60:60	1:2:3:4	**1.8**	1.0	**1.6**	0.9	1.0	1.2	**1.7**	1.2	1.0	1.0	1.2
	1:1:1:1	1.1	1.0	1.1	1.2	1.2	1.1	1.5	1.1	1.0	1.0	1.1
20:40:80:100	4:3:2:1	**10.9**	1.3	**1.8**	1.2	1.4	0.8	0.6	0.8	0.7	0.7	0.6
	1:2:3:4	**0.2**	1.1	2.7	1.3	1.1	1.0	1.0	1.0	1.0	1.0	1.0
	1:1:1:1	**1.6**	1.4	1.6	1.5	1.4	1.0	0.9	1.0	1.2	1.1	0.9
300:300:300:300	1:2:3:4	1.4	0.6	1.3	0.7	0.6	1.3	**1.8**	**1.6**	1.3	1.3	1.2
	1:1:1:1	0.6	0.5	0.6	0.8	0.7	0.8	1.5	**1.6**	0.7	0.7	0.7
100:200:400:500	4:3:2:1	**11.6**	**1.8**	1.8	1.3	1.4	0.6	**1.7**	0.8	1.2	1.2	0.6
	1:2:3:4	**0.3**	0.6	**2.1**	0.8	0.6	1.2	**2.0**	0.9	1.2	1.2	1.1
	1:1:1:1	1.3	1.5	1.2	1.3	1.4	0.8	**1.6**	0.8	0.9	0.9	0.7

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .05 and .10. The ANOVA alternative methods and the RMM approaches were almost robust across all the conditions of sample sizes and variance ratios. As usual, the Type I error rates of the ANOVA F test were inflated at all "negative conditions", but pushed below the lower boundary of the robustness range at all "positive conditions".

Table 19.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), *k*=4 at α =.05

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	6.3	5.6	6.3	5.9	5.9	5.6	5.7	4.8	5.6	5.0	5.6
	1:1:1:1	5.4	5.6	5.5	6.0	6.0	5.9	5.9	5.2	5.8	5.3	5.9
	4:3:2:1	**24.1**	5.3	6.3	5.3	5.3	4.8	4.9	5.1	4.9	4.3	4.8
10:20:40:50	1:2:3:4	**1.4**	5.3	5.7	5.8	5.8	4.5	4.7	5.7	4.6	4.1	4.5
	1:1:1:1	4.4	5.1	4.5	5.3	5.2	4.4	4.8	5.5	4.5	4.2	4.4
60:60:60:60	1:2:3:4	6.8	5.8	6.4	5.9	5.9	5.6	6.3	5.6	5.7	5.7	5.5
	1:1:1:1	6.2	5.0	4.2	5.6	5.4	5.0	5.7	5.0	5.1	5.0	4.9
	4:3:2:1	**22.3**	4.4	6.4	5.2	4.5	4.2	4.2	4.2	4.9	4.7	4.2
20:40:80:100	1:2:3:4	**2.3**	5.2	6.2	5.2	5.2	4.8	4.8	4.8	4.8	4.7	4.8
	1:1:1:1	5.6	5.2	5.3	5.4	5.2	4.2	4.1	4.2	4.4	4.2	4.1
300:300:300:300	1:2:3:4	7.0	4.4	6.3	5.1	4.7	4.8	6.3	5.6	4.7	4.6	4.6
	1:1:1:1	5.2	5.9	5.2	6.0	6.1	5.2	6.5	5.9	5.2	5.1	5.1
	4:3:2:1	**22.1**	5.6	5.6	4.8	5.0	5.1	**7.7**	4.8	6.2	6.2	5.1
100:200:400:500	1:2:3:4	**1.6**	4.3	7.1	4.5	4.5	5.7	7.3	4.5	6.0	6.0	5.6
	1:1:1:1	5.2	5.5	4.4	5.2	5.1	5.5	**8.3**	4.4	6.7	6.5	5.2

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Table 20.
Type I Error Rates (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), *k*=4 at α =.10

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	11.1	8.9	11.1	9.4	9.5	10.8	11.1	10.6	10.9	9.9	10.8
	1:1:1:1	9.7	9.8	9.7	10.2	10.5	11.9	12.0	10.8	11.9	11.4	11.8
	4:3:2:1	**31.6**	10.6	10.9	10.5	10.8	9.5	9.6	10.2	9.4	9.1	9.4
10:20:40:50	1:2:3:4	3.3	9.5	11.6	9.6	10.0	9.9	9.9	10.4	9.8	9.2	9.7
	1:1:1:1	10.3	10.0	10.5	10.4	10.6	9.0	9.3	11.2	9.3	8.9	9.0
60:60:60:60	1:2:3:4	10.9	10.4	10.5	10.6	11.1	10.5	11.1	10.5	10.5	10.4	10.4
	1:1:1:1	10.4	10.8	10.6	11.2	11.3	10.4	10.9	10.4	10.5	10.4	10.4
	4:3:2:1	**31.6**	8.4	10.3	8.6	8.8	9.6	9.2	9.6	9.8	9.8	9.2
20:40:80:100	1:2:3:4	**3.5**	10.0	11.0	10.0	10.4	10.5	10.3	10.5	10.6	10.4	10.3
	1:1:1:1	9.4	9.0	8.3	9.1	9.1	9.7	9.7	9.7	10.7	10.6	9.7
300:300:300:300	1:2:3:4	12.8	10.5	12.1	11.1	11.6	10.6	12.1	10.8	10.4	10.3	10.4
	1:1:1:1	10.3	10.5	10.2	10.6	11.0	10.8	11.7	11.9	10.6	10.4	10.4
	4:3:2:1	**31.4**	11.0	9.9	0.9	10.9	10.2	13.7	9.5	12.3	12.2	10.0
100:200:400:500	1:2:3:4	**3.8**	10.4	10.4	10.7	10.5	10.4	12.3	9.9	10.5	10.4	10.1
	1:1:1:1	9.8	10.3	9.2	10.6	10.3	11.2	14.2	9.0	12.4	12.0	11.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Asymmetric Nonnormal Distribution (3,21)

K=2

α= .01. The results for the asymmetric nonnormal distribution were a little messier when the distribution shape departed farther from normality. A lot more non-robust cells were observed. Generally speaking, the ANOVA-based methods provided VERY inflated Type I error rates increasing with sample size when variances were heterogeneous. The behavior of those ANOVA methods tended to be robust when the variances were equal and sample sizes increased. The RMM approaches also delivered inflated rejection rates in many cells, which were, however, much smaller than those from the ANOVA methods, especially when sample size increased.

Table 21.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=2 at α =.01

n_1: n_2	σ_1/ σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	0.5	**0.3**	**0.3**	**0.3**	0.5	0.6	2.4	0.6	0.5	1.0	0.6
	2.5	**11.2**	1.2	1.2	**2.1**	1.4	**2.8**	**9.1**	**2.9**	**4.7**	**7.2**	**2.2**
	4	**26.9**	1.0	1.0	**2.2**	1.3	**3.2**	**11.2**	**3.2**	**7.3**	**9.1**	**2.8**
16:4	2.5	0.8	0.6	0.6	0.6	0.8	**0.0**	**0.4**	**0.0**	**0.0**	**0.0**	**0.0**
	4	0.6	**3.1**	**3.1**	**3.9**	**4.1**	**0.2**	1.0	**0.2**	**0.0**	0.5	**0.1**
10:10	1	0.5	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**	1.2	**0.3**	**0.1**	**0.4**	**0.2**
	2.5	**6.8**	**5.5**	**5.5**	**5.7**	**6.4**	**1.9**	**3.3**	**1.9**	1.5	**2.3**	1.4
	4	**11.6**	**9.7**	**9.7**	**10.2**	**10.7**	**3.1**	**5.0**	**3.1**	**2.7**	**3.5**	**2.8**
20:80	1	0.9	1.2	1.2	1.2	1.2	**1.6**	**2.1**	**1.6**	1.4	1.4	1.4
	2.5	**28.1**	**11.4**	**11.4**	**11.7**	**11.7**	**2.6**	**3.1**	**2.6**	**3.0**	**3.0**	**2.3**
	4	**42.8**	**15.4**	**15.4**	**16.1**	**16.1**	**3.1**	**3.8**	**3.1**	**3.6**	**3.6**	**2.7**
80:20	2.5	**2.1**	**12.5**	**12.5**	**12.7**	**10.7**	0.7	1.0	0.7	0.8	0.8	0.7
	4	**3.9**	**33.0**	**33.0**	**33.3**	**33.7**	0.9	1.2	0.9	1.0	1.0	0.9
50:50	1	0.8	0.8	0.8	0.9	0.9	0.6	0.7	0.6	**0.3**	**0.3**	**0.4**
	2.5	**46.2**	**16.4**	**16.4**	**16.6**	**16.8**	**1.8**	**2.2**	**1.8**	**1.6**	**1.6**	**1.6**
	4	**28.9**	**28.8**	**28.8**	**28.8**	**28.8**	**3.2**	**3.6**	**3.2**	**2.8**	**3.0**	**3.0**
100:400	1	1.1	1.2	1.2	1.3	1.3	**2.0**	**2.0**	**2.0**	**2.0**	**2.0**	**2.0**
	2.5	**63.0**	**32.9**	**32.9**	**33.1**	**33.1**	**2.2**	**2.4**	**2.2**	**2.3**	**2.2**	**2.2**
	4	**84.1**	**50.8**	**50.8**	**50.9**	**50.9**	**2.2**	**2.4**	**2.2**	1.5	1.4	**2.2**
400:100	2.5	**22.9**	**77.0**	**77.0**	**77.2**	**77.3**	**1.6**	**1.7**	**1.6**	**1.6**	**1.6**	**1.6**
	4	**51.3**	**96.7**	**96.7**	**96.8**	**96.8**	**1.8**	**1.8**	**1.8**	**1.8**	**1.7**	**1.8**
250:250	1	**0.3**	**0.3**	0.5	0.5	**0.3**	1.2	1.2	1.2	1.2	1.1	1.2
	2.5	**66.9**	**67.9**	**67.9**	**68.1**	**68.2**	1.4	1.5	1.4	1.4	1.2	1.4
	4	**87.5**	**88.1**	**88.1**	**88.1**	**88.1**	**1.6**	**1.7**	**1.6**	1.5	1.4	1.5

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

\α= .05. Again, the ANOVA-based methods provided VERY inflated Type I error rates when variances were heterogeneous with moderate and large sample sizes. Excitingly, the RMM approaches performed much superior to the ANOVA-based methods. For the RMM approaches, the Type I error rates were inflated at the "negative conditions" when sample sizes were small and moderate, and sporadically when sample sizes were small and variances

were heterogeneous. Overall, the adjusted T_{ADF} test statistic, T_{YB1} and T_{YB2} provided the most robust cells among all the examined methods, followed by the T_{ML}, T_{SB} and T_{BC} test statistics.

Table 22.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), *k*=2 at α =.05

n_1: n_2	σ_1/ σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	**2.2**	**1.4**	**1.4**	**1.7**	**1.5**	4.5	7.0	4.5	4.7	5.2	4.0
	2.5	**30.8**	2.8	2.8	5.7	3.6	**9.8**	**15.0**	**9.8**	**12.6**	**13.7**	**8.6**
	4	**44.6**	**1.8**	**1.8**	5.4	3.6	**10.3**	**16.8**	**10.3**	**13.6**	**14.4**	**9.3**
16:4	2.5	**2.0**	3.1	3.1	4.1	3.6	**1.9**	3.5	**1.9**	**1.8**	**2.3**	**1.6**
	4	3.3	**9.4**	**9.4**	**10.7**	10.6	3.1	4.9	3.1	3.3	4.0	**2.3**
10:10	1	3.4	**2.4**	**2.4**	**2.4**	2.5	3.7	6.0	3.7	3.0	3.6	2.8
	2.5	**15.7**	**14.6**	**14.6**	**14.6**	**15.0**	7.1	**9.3**	7.1	6.7	7.4	6.3
	4	**22.4**	**9.6**	**19.6**	**20.0**	**20.7**	**9.1**	**11.3**	**9.1**	**8.1**	**9.6**	**7.9**
20:80	1	5.4	5.6	5.6	5.4	5.5	5.1	5.7	5.1	5.1	5.1	4.8
	2.5	**42.1**	**22.6**	**22.6**	**22.8**	**22.6**	**8.7**	**9.6**	**8.7**	**9.3**	**9.2**	**8.5**
	4	**54.5**	**27.3**	**27.3**	**27.5**	**27.5**	**9.5**	**10.5**	**9.5**	**10.0**	**9.9**	**9.3**
80:20	2.5	**11.2**	**32.2**	**32.2**	**31.9**	**32.3**	4.4	4.8	4.4	4.3	4.3	4.2
	4	**11.3**	**54.7**	**54.7**	**54.2**	**54.7**	5.5	5.6	5.5	5.7	5.7	5.5
50:50	1	3.9	4.3	4.3	4.1	4.3	4.5	4.6	4.5	4.4	4.4	4.4
	2.5	**32.7**	**32.7**	**32.7**	**32.7**	**32.7**	6.5	6.9	6.5	6.4	6.4	6.3
	4	**46.5**	**45.9**	**45.9**	**45.9**	**45.9**	7.1	7.5	7.1	7.1	7.1	7.1
100:400	1	5.0	5.4	5.4	5.4	5.5	5.4	5.6	5.4	5.4	5.0	5.4
	2.5	**77.1**	**53.1**	**53.1**	**53.2**	**53.2**	**7.6**	**7.8**	**7.6**	**7.8**	7.5	7.5
	4	**90.0**	**68.3**	**68.3**	**68.3**	**68.3**	7.0	**7.6**	7.0	6.3	6.1	7.0
400:100	2.5	**57.3**	**91.3**	**91.3**	**91.6**	**91.7**	6.0	6.0	6.0	6.0	6.0	6.0
	4	**84.0**	**99.0**	**99.0**	**99.0**	**99.0**	6.5	6.5	6.5	6.4	6.4	6.5
250:250	1	4.2	4.3	4.3	4.4	4.4	6.2	6.2	6.2	6.2	6.1	6.2
	2.5	**83.1**	**83.8**	**83.8**	**84.0**	**84.0**	6.2	6.6	6.2	6.2	6.1	6.2
	4	**94.9**	**95.0**	**95.0**	**95.0**	**95.0**	6.3	6.3	6.3	6.3	6.1	6.3

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .10. The results at the nominal level of 0.10 were very similar to the results at the significant level of 0.05 for the asymmetric nonnormal distribution. Fewer non-robust cells

were observed by the RMM approaches. The RMM approaches only provided inflated Type I error rates at the "negative conditions" when sample size was small. Overall, the T_{BC} test statistic provided the most robust cells among all the examined methods, followed by the T_{ML}, T_{SB} and T_{YB1} test statistics.

Table 23.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), *k*=2 at α =.10

n_1: n_2	σ_1/ σ_2	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
4:16	1	5.3	**2.4**	**2.4**	**3.3**	**2.9**	9.1	11.4	9.1	9.7	10.0	8.5
	2.5	**41.7**	**3.5**	**3.5**	8.2	6.9	**15.1**	**19.9**	**15.1**	**17.6**	**18.7**	14.4
	4	**54.8**	**3.1**	**3.1**	8.4	7.4	**15.7**	**21.7**	**15.7**	**19.7**	**20.6**	14.6
16:4	2.5	**4.2**	5.9	5.9	7.1	6.9	5.4	8.4	5.4	6.0	6.6	**4.8**
	4	8.0	14.7	**14.7**	**17.2**	**16.6**	7.6	10.0	7.6	7.8	8.2	6.7
10:10	1	7.3	6.6	6.6	6.0	6.9	9.8	11.9	9.8	9.1	10.0	8.8
	2.5	**22.4**	**20.8**	**20.8**	**20.5**	**21.1**	12.1	14.9	12.1	11.6	12.8	11.5
	4	**28.5**	**26.8**	**26.8**	**26.9**	**27.0**	14.8	**16.7**	14.8	14.3	**15.2**	13.7
20:80	1	9.7	10.4	10.4	10.4	10.5	10.0	10.3	10.0	10.0	10.0	9.8
	2.5	**49.5**	**29.6**	**29.6**	**29.6**	**29.6**	13.9	**15.4**	13.9	14.6	14.6	13.8
	4	**62.2**	**37.4**	**37.4**	**37.5**	**37.5**	14.4	**15.4**	14.4	15.0	15.0	14.3
80:20	2.5	**18.9**	**43.9**	**43.9**	**43.9**	**44.5**	9.0	9.3	9.0	8.9	8.9	8.7
	4	**26.6**	**65.1**	**65.1**	**65.2**	**65.3**	9.6	9.7	9.6	11.3	11.3	9.6
50:50	1	9.4	9.4	9.4	9.4	9.7	10.4	10.9	10.4	10.3	10.2	10.2
	2.5	**42.8**	**43.0**	**43.0**	**43.5**	**43.5**	11.9	12.3	11.9	11.8	11.7	11.8
	4	**55.5**	**55.4**	**55.4**	**55.5**	**55.6**	12.5	12.5	12.5	12.5	12.5	12.4
100:400	1	9.5	11.0	11.0	11.1	11.1	11.0	11.4	11.0	11.3	11.1	11.0
	2.5	**82.5**	**62.2**	**62.2**	**62.2**	**62.2**	12.0	12.3	12.0	12.1	11.9	11.9
	4	**93.1**	**76.8**	**76.8**	**76.8**	**76.7**	12.4	12.4	12.4	11.5	11.3	12.4
400:100	2.5	**73.7**	**95.9**	**95.9**	**96.0**	**96.0**	10.6	10.7	10.6	10.6	10.4	10.6
	4	**94.0**	**99.6**	**99.6**	**99.6**	**99.6**	10.4	10.5	10.4	10.4	10.1	10.4
250:250	1	9.4	9.6	9.6	9.6	9.6	10.7	10.7	10.7	10.6	10.6	10.7
	2.5	**89.6**	**89.8**	**89.8**	**89.8**	**89.8**	11.8	11.8	11.8	11.8	11.7	11.8
	4	**96.9**	**97.0**	**97.0**	**97.0**	**97.0**	11.5	11.5	11.5	11.5	11.3	11.5

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

K=3

<u>α= .01</u>. All ANOVA-based methods delivered inflated Type I error rates at all "negative conditions", "positive conditions" and conditions with equal sample sizes but heterogeneous variances. The RMM approaches provided less non-robust cells with less inflated Type I error rates at the "negative conditions" across sample sizes, or equal and small or moderate sample size conditions with heterogeneous variances. The test statistics T_{BC}, T_{ML}, and T_{SB} best controlled the Type I error rates, followed by test statistics, T_{YB1} and T_{YB2}.

Table 24.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), *k*=3 at α =.01

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	**13.2**	**22.1**	**12.9**	**22.7**	**22.2**	**2.6**	**4.0**	**2.6**	**2.7**	**3.0**	**2.5**
	1:1:1	0.4	0.6	0.4	0.6	0.6	0.7	1.1	0.7	0.5	0.6	0.6
12:30:48	4:2.5:1	**32.6**	**21.5**	**15.2**	**20.9**	**20.9**	**3.3**	**5.8**	**3.3**	**4.6**	**4.8**	**3.0**
	1:2.5:4	**4.3**	**16.0**	**10.4**	**15.8**	**16.1**	1.1	1.4	1.5	0.9	1.0	1.0
	1:1:1	0.7	1.4	1.2	1.5	1.3	0.9	1.5	0.9	1.0	1.1	0.8
60:60:60	1:2.5:4	**25.5**	**37.3**	**25.5**	**39.2**	**38.9**	**2.2**	**2.7**	**2.2**	**2.1**	**2.4**	**2.2**
	1:1:1	0.9	0.9	0.9	0.9	0.9	0.9	1.4	0.9	0.8	0.9	0.9
24:60:96	4:2.5:1	**41.4**	**28.1**	**18.5**	**28.6**	**27.8**	**2.7**	**3.6**	**2.7**	**2.8**	**3.0**	**2.6**
	1:2.5:4	**9.6**	**37.2**	**26.8**	**37.4**	**37.6**	1.5	**1.8**	1.5	1.4	1.4	1.5
	1:1:1	0.5	1.0	0.4	1.2	1.0	1.5	**2.0**	1.5	1.5	1.5	1.4
300:300:300	1:2.5:4	**99.0**	**99.9**	**99.1**	**99.9**	**99.9**	1.2	1.3	1.2	1.2	1.3	1.2
	1:1:1	0.8	0.8	0.7	0.8	0.8	**0.4**	0.6	**0.4**	0.4	0.6	0.4
120:300:480	4:2.5:1	**93.6**	**91.5**	**71.4**	**91.7**	**91.5**	**1.8**	**2.0**	**1.8**	**1.9**	**2.0**	**1.8**
	1:2.5:4	**63.3**	**99.4**	**90.6**	**99.5**	**99.5**	0.9	0.9	0.9	0.9	0.9	0.9
	1:1:1	1.1	1.3	0.8	1.4	1.4	1.4	**1.6**	1.4	**1.6**	**1.6**	1.4

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

<u>α= .05</u>. Similar results were observed for all ANOVA methods at the nominal level of .01. The RMM approaches only provided inflated Type I error rates at the "negative conditions" when sample size was small or moderate, or equal sample size conditions with heterogeneous variances and small sizes. Again, the RMM approaches were superior to the ANOVA-based methods and best controlled the Type I error rates.

Table 25.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=3 at α =.05

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	**26.2**	**39.1**	**25.4**	**39.1**	**39.1**	**8.5**	**9.2**	**8.5**	**8.4**	**8.6**	**8.3**
	1:1:1	3.8	4.4	3.7	4.4	4.7	5.2	5.9	5.2	5.0	5.0	5.1
12:30:48	4:2.5:1	**50.6**	**38.0**	**29.3**	**38.7**	**37.8**	**10.5**	**13.2**	**10.5**	**12.3**	**12.3**	**10.1**
	1:2.5:4	**12.0**	**33.3**	**24.3**	**33.2**	**33.5**	5.4	6.5	5.3	5.5	5.5	5.3
	1:1:1	2.9	5.3	4.1	5.7	5.3	5.7	7.3	5.7	5.5	5.6	5.5
60:60:60	1:2.5:4	**41.3**	**57.8**	**41.6**	**59.1**	**58.7**	6.8	6.9	6.8	6.7	6.8	6.7
	1:1:1	3.3	4.2	3.5	4.2	4.2	4.7	5.0	4.7	4.7	4.7	4.7
24:60:96	4:2.5:1	**57.1**	**45.9**	**33.5**	**46.2**	**45.9**	**9.3**	**10.7**	**9.3**	**10.0**	**10.0**	**9.2**
	1:2.5:4	**25.7**	**59.3**	**44.0**	**58.9**	**59.7**	5.3	5.8	5.3	5.3	5.3	5.2
	1:1:1	3.9	4.6	3.5	4.6	4.7	5.9	6.4	5.9	5.9	6.0	5.9
300:300:300	1:2.5:4	**96.8**	**99.6**	**91.0**	**99.7**	**99.7**	4.4	4.5	4.4	4.4	4.1	4.3
	1:1:1	4.8	4.9	5.3	5.0	5.0	4.5	4.6	4.5	4.4	3.9	4.3
120:300:480	4:2.5:1	**97.5**	**96.6**	**87.0**	**95.7**	**96.8**	6.9	6.9	6.7	6.7	6.7	6.7
	1:2.5:4	**85.8**	**99.8**	**97.9**	**99.8**	**99.8**	5.4	5.5	5.4	5.4	5.0	5.4
	1:1:1	4.1	4.4	3.8	4.3	4.4	5.9	5.9	5.9	5.9	5.4	5.9

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

<u>α= .10</u>. Similar results were observed for all ANOVA-based methods at the nominal level of .01 or .05. Fewer non-robust cells were found for the RMM approaches, which

provided inflated Type I error rates at the "negative conditions" when sample size was small. The RMM approaches best controlled the Type I error rates.

Table 26.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), *k*=3 at α=.10

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30	1:2.5:4	**35.7**	**49.2**	**35.4**	**49.1**	**49.2**	13.8	**15.5**	13.8	13.7	13.8	13.7
	1:1:1	8.5	9.5	8.9	9.4	9.5	9.9	11.9	9.9	9.8	9.8	9.7
12:30:48	4:2.5:1	**57.6**	**48.2**	**38.5**	**48.4**	**48.1**	**16.9**	**19.4**	**16.9**	**17.8**	**17.8**	**16.3**
	1:2.5:4	**18.9**	**44.9**	**34.6**	**43.9**	**44.9**	11.3	12.3	11.9	11.1	11.1	10.9
	1:1:1	6.0	10.2	8.7	10.9	10.0	11.4	12.9	11.4	12.0	12.0	11.3
60:60:60	1:2.5:4	**52.6**	**68.5**	**53.0**	**69.5**	**69.1**	12.1	12.8	12.1	12.1	12.1	12.1
	1:1:1	7.0	8.8	7.1	8.8	8.8	9.2	9.8	9.2	9.0	9.1	9.1
24:60:96	4:2.5:1	**66.0**	**57.2**	**43.8**	**57.7**	**57.2**	14.3	**15.4**	14.3	15.0	**15.1**	14.0
	1:2.5:4	**34.7**	**70.2**	**55.2**	**70.2**	**70.5**	11.9	12.5	11.9	11.8	11.8	11.7
	1:1:1	8.2	9.8	9.4	10.3	9.9	11.5	12.7	11.5	11.7	11.9	11.2
300:300:300	1:2.5:4	**99.0**	**99.9**	**99.1**	**99.9**	**99.9**	10.4	10.6	10.4	10.3	9.8	10.2
	1:1:1	9.3	10.5	10.2	11.2	11.2	10.7	10.7	10.7	10.6	9.7	10.6
120:300:480	4:2.5:1	**98.5**	**98.0**	**92.4**	**98.0**	**98.0**	12.6	12.9	12.6	12.7	12.3	12.6
	1:2.5:4	**92.8**	**100.0**	**98.9**	**100.0**	**100.0**	10.5	10.6	10.5	10.5	10.2	10.5
	1:1:1	7.9	9.5	9.2	9.8	9.8	11.7	11.7	11.7	11.7	11.3	11.7

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

K=4

α= .01. The results for the asymmetric nonnormal distribution were messy when the distribution shape departed farther from normality. Again, the ANOVA-based methods provided inflated Type I error rates when variances were heterogeneous across sample sizes. The RMM approaches generally provided robust cells at the "negative conditions" across sample sizes. Type I error rates were also inflated by the RMM approaches with equal sample sizes when variances were heterogeneous; so were Type I error rates with equal small or moderate sample sizes but variances were heterogeneous.

Table 27.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), *k*=4 at α =.01

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	**11.8**	**21.8**	**10.8**	**22.9**	**22.6**	1.2	1.2	1.2	1.2	1.1	1.2
	1:1:1:1	0.6	0.6	0.4	0.7	0.7	0.8	0.9	0.8	0.8	0.8	0.8
10:20:40:50	4:3:2:1	**21.6**	**17.9**	**7.8**	**19.8**	**16.3**	**2.3**	**2.6**	**2.3**	**2.5**	**2.2**	**2.3**
	1:2:3:4	**2.7**	**13.4**	**7.9**	**13.4**	**12.8**	1.4	1.4	1.4	1.3	1.2	1.4
	1:1:1:1	0.3	**1.7**	0.4	**1.6**	1.5	**2.1**	**2.3**	**2.1**	**2.1**	**1.7**	**2.3**
60:60:60:60	1:2:3:4	**22.4**	**39.9**	**21.0**	**41.3**	**40.9**	**2.0**	**2.7**	**2.0**	**2.0**	**2.0**	**2.0**
	1:1:1:1	0.8	0.8	0.6	1.0	1.0	0.9	1.2	0.9	0.9	0.9	0.9
20:40:80:100	4:3:2:1	**39.7**	**3.1**	**15.8**	**34.1**	**31.2**	**3.7**	**4.9**	**3.7**	**4.3**	**4.3**	**3.7**
	1:2:3:4	**7.7**	**3.3**	**19.9**	**33.0**	**33.7**	**1.9**	**2.1**	**1.9**	**1.6**	**1.6**	**1.7**
	1:1:1:1	1.1	1.1	0.5	1.1	1.1	**1.8**	**2.3**	**1.8**	**1.9**	**1.9**	**1.7**
300:300:300:300	1:2:3:4	**93.1**	**99.4**	**93.6**	**99.4**	**99.4**	**3.1**	**3.8**	**3.1**	**3.1**	**3.1**	**3.1**
	1:1:1:1	0.6	0.7	0.6	0.8	0.8	**1.6**	**2.0**	**1.6**	**1.6**	**1.6**	**1.6**
100:200:400:500	4:3:2:1	**93.8**	**95.2**	**73.6**	**95.6**	**95.4**	**4.1**	**6.7**	**4.1**	**5.5**	**5.3**	**3.8**
	1:2:3:4	**62.4**	**99.4**	**89.3**	**99.5**	**99.5**	1.5	**1.8**	1.5	1.5	1.4	1.5
	1:1:1:1	0.9	1.2	0.8	1.4	1.4	**1.6**	**2.8**	**1.6**	**1.8**	**1.7**	1.5

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .01. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

<u>*α= .05*</u>. Again, the ANOVA-based methods provided very inflated Type I error rates when variances were heterogeneous across sample sizes. However, the performance of the RMM approaches became much better, providing much less non-robust cells and Type I error rates much closer to the nominal level .05. The RMM approaches only yielded inflated Type I error rates at the "negative conditions" with small or moderate sample sizes, or small equal sample sizes conditions with heterogeneous variances. In addition, when the unequal sample sizes were small or moderate, the test statistic T_{ADF} also delivered inflated Type I error rates. In

sum, the RMM approaches were superior to the ANOVA-based methods and best controlled the Type I error rates.

Table 28.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=4 at α=.05

$n_1: n_2: n_3: n_4$	$\sigma_1/\sigma_2/\sigma_3/\sigma_4$	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	**26.4**	**41.2**	**25.8**	**41.1**	**41.6**	**9.1**	**10.7**	**9.1**	**9.3**	**9.1**	**9.0**
	1:1:1:1	3.9	3.5	3.8	4.2	4.0	6.0	6.9	6.0	5.7	5.6	5.7
10:20:40:50	4:3:2:1	**39.8**	**34.9**	**17.0**	**35.2**	**33.9**	**10.4**	**13.9**	**10.4**	**12.5**	**12.4**	**10.2**
	1:2:3:4	**8.9**	**33.4**	**20.5**	**32.6**	**33.3**	5.4	6.8	5.4	5.3	5.2	5.1
	1:1:1:1	4.7	5.0	3.2	5.2	4.8	6.3	**8.2**	6.3	7.1	6.9	5.9
60:60:60:60	1:2:3:4	**39.2**	**61.6**	**38.8**	**62.3**	**62.0**	6.8	7.1	6.8	6.8	6.8	6.8
	1:1:1:1	3.4	4.0	3.9	4.1	4.0	5.1	5.9	5.1	5.1	5.1	5.1
20:40:80:100	4:3:2:1	**57.4**	**52.6**	**28.7**	**54.2**	**52.9**	**9.1**	**11.4**	**9.1**	**10.0**	**9.9**	**8.9**
	1:2:3:4	**17.4**	**55.3**	**34.6**	**55.0**	**56.4**	6.2	6.7	6.2	6.2	6.2	6.1
	1:1:1:1	4.8	6.1	2.8	6.1	6.2	6.6	**7.6**	6.6	6.8	6.7	6.5
300:300:300:300	1:2:3:4	**98.2**	**100.0**	**98.2**	**100.0**	**100.0**	5.1	5.2	5.1	5.1	5.0	5.1
	1:1:1:1	4.0	4.1	4.0	4.7	4.5	5.3	5.5	5.3	5.3	4.9	5.3
100:200:400:500	4:3:2:1	**98.3**	**98.5**	**87.7**	**98.7**	**98.6**	7.5	7.5	7.5	7.5	7.1	7.5
	1:2:3:4	**84.9**	**99.9**	**97.3**	**100.0**	**100.0**	5.1	5.3	5.1	5.2	5.0	5.1
	1:1:1:1	4.2	4.8	4.7	4.9	4.9	7.1	7.1	7.1	71.0	6.7	7.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .05. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

α= .10. Similar results were observed for the ANOVA-based methods at the nominal level of 05. The RMM approaches only yielded inflated Type I error rates at the "negative conditions" when sample size was small or moderate. The test statistic T_{ADF} yielded inflated Type I error rate when variances were heterogeneous with small sample size. Similarly, the RMM approaches were superior to the ANOVA-based methods and best controlled the Type I error rates.

Table 29.
Type I Error Rates (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=4 at α =.10

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F	W	BF	A	U	T_{ML}	T_{ADF}	T_{SB}	T_{YB1}	T_{YB2}	T_{BC}
30:30:30:30	1:2:3:4	**35.0**	**53.0**	**34.4**	**53.3**	**53.6**	10.9	**16.4**	10.9	15.0	14.5	14.8
	1:1:1:1	7.2	9.2	7.1	9.3	9.3	11.1	12.3	11.1	10.7	10.6	10.8
10:20:40:50	4:3:2:1	**50.4**	**45.5**	**25.1**	**45.7**	**45.1**	**15.5**	**19.8**	**15.5**	**18.0**	**17.8**	**15.1**
	1:2:3:4	**14.7**	**44.0**	**30.0**	**43.0**	**44.2**	10.5	11.6	10.5	10.6	10.4	10.4
	1:1:1:1	9.4	9.1	6.3	9.4	9.1	10.9	13.8	10.9	12.0	11.6	10.8
60:60:60:60	1:2:3:4	**50.4**	**70.6**	**50.3**	**70.5**	**70.9**	12.2	13.3	12.2	12.2	12.0	12.1
	1:1:1:1	7.5	7.8	7.5	7.8	8.0	10.8	11.3	10.8	10.8	10.6	10.8
20:40:80:100	4:3:2:1	**65.3**	**63.3**	**38.8**	**64.5**	**63.8**	**16.6**	**18.9**	**16.6**	**17.7**	**17.4**	**16.3**
	1:2:3:4	**26.1**	**69.5**	**46.3**	**68.2**	**69.8**	10.8	11.3	10.8	11.0	10.9	10.8
	1:1:1:1	9.9	11.5	9.0	11.6	11.9	13.3	14.8	13.3	13.9	13.6	12.9
300:300:300:300	1:2:3:4	**99.3**	**100.0**	**99.3**	**100.0**	**100.0**	9.6	9.6	9.6	9.6	9.3	9.5
	1:1:1:1	9.0	10.0	9.7	10.4	10.5	9.6	9.8	9.6	9.7	9.1	9.6
100:200:400:500	4:3:2:1	**99.0**	**99.5**	**92.9**	**99.5**	**99.5**	12.6	12.7	12.6	12.7	12.1	12.6
	1:2:3:4	**92.2**	**100.0**	**98.8**	**100.0**	**100.0**	9.8	10.0	9.8	10.0	9.2	9.8
	1:1:1:1	9.7	9.3	8.6	9.4	9.6	10.9	11.1	10.9	10.9	10.7	13.3

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Type I error rates are computed across 1000 replications at the nominal level of .10. Bolded values indicate rejection rate falling outside of the Bradley's robustness interval.

Empirical Power

As noted before, the purpose of the power analysis is to make comparisons of the performance of different approaches, which are relative estimates rather than considered to be absolute. Based on the results of Type I error rates, all test statistics were studied for the normal distribution and the elliptically symmetric nonnormal distribution with skewness and kurtosis of (0, 3). However, because of the unsatisfactory performance of the ANOVA-based methods for many conditions, only the RMM approaches were included for the asymmetric nonnormal distribution with skewness and kurtosis of (3, 21) for most of the conditions. Table 30 summarizes the test statistics studied across distributional shapes, levels of significance and number of groups in the power analysis.

Table 30.
Methods Used in the Power Analysis

D	*K*	α = .10	α= .05	α= .01
	2	All	All	All
(0, 0)	3	All	All	All
	4	All	All	All
	2	All	All	All
(0, 3)	3	All	All	All
	4	All	All	All
	2	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}	T_{YB2}, T_{YB1}, T_{BC}, BF, U
(3, 21)	3	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}
	4	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}	T_{YB1}, T_{YB2}, T_{ML}, T_{BC}, T_{SB}, T_{ADF}

Note. All = all ANOVA methods and SMM methods, including F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
D is the distributional shapes. (0, 0) = normal distribution with skewness and kurtosis of (0, 0); (0, 3) = elliptically symmetric nonnormal distribution with skewness and kurtosis of (0, 3); (3, 21) = asymmetric nonnormal distribution with skewness and kurtosis of (3, 21).
K indicates the number of groups and α is the nominal level.

To assess the empirical power estimates, the study crossed the selected test statistics with the conditions used in the study of Type I error rates, including different pairs of sample sizes, three pairs of variance ratios, three groups, three Type I error rate levels as well as two effect size conditions. For the normal distribution and the elliptically symmetric nonnormal distribution with skewness and kurtosis of (0, 3), there were 11×(24+15+15)×3×2=3564 cells respectively. For the for the asymmetric nonnormal distribution with skewness and kurtosis of (3, 21), six test statistics were selected for the nominal levels of .05 and .10 while five test statistics were selected for the nominal levels of .01, across nice pairs of sample sizes, three pairs of variance ratios, as well as two effect size conditions, yielding 6×24×2×2+5×24×2 =816 cells for *k*=2; similarly, 6×15×2×2×2+5×15×2=510 and 6×15×3×2=540 cells were yielded for k=3 and k=4 respectively. In total, 3564×2+(816+510+540)=8994 cells were studied.

Again, there are three different situations, which are the “positive condition”, the “negative condition”, and the conditions of equal sample sizes. Table 31 to Table 57 present the results of analyses assessing the power of the selected test statistics across the same conditions as the analysis of Type I error rates, while including two additional conditions of effect size: d=.2 or .8 when k=2, and f=.1 or .4 when $k \geq 2$.

Table 31 to Table 39 present the results of the power estimates under three group variance ratios, three different sample size ratios and two effect size conditions for the normal distribution. Among the nine tables, Table 31, 32 and 33 present the results of the power estimates for k=2 at the nominal levels .01, .05 and .10. Table 34, 35, and 36 present the results of the empirical power estimates for k=3 at nominal levels .01, .05 and .10, while Table 37, 38, and 39 present the results of the power estimates for k=4 at nominal levels .01, .05 and .10.

Table 40 to Table 48 present the results of the empirical power estimates under three group variance ratios, three different sample size ratios and two effect size conditions for the elliptical distribution with univariate skew of 0 and kurtosis of 3. Among the nine tables, Table 40, 41, and 42 present the results of the power estimates for k=2 at the nominal levels .01, .05 and .10. Table 43, 44, and 45 present the results of the power estimates for k=3 at nominal levels .01, .05 and .10, while Table 46, 47, and 48 present the results of the empirical power estimates for k=4 at nominal levels .01, .05 and .10.

Table 49 to Table 57 present the results of the empirical power estimates under three group variance ratios, three different sample size ratios and two effect size conditions for the nonnormal distribution with univariate skew of 3 and kurtosis of 21. Among the nine tables,

Table 49, 50, and 51 present the results of the empirical power estimates for k=2 at nominal levels .01, .05 and .10. Table 52, 53, and 54 present the results of the empirical power estimates for k=3 at nominal levels .01, .05 and .10, while Table 55, 56, and 57 present the results of the empirical power estimates for k=4 at nominal levels .01, .05 and .10.

Normal distribution

K=2

α= .01. At the "negative conditions", the RMM test statistics generally provided higher empirical power estimates. The test statistics T_{YB2} and T_{YB1} delivered power estimates approximately 0%-3% higher than those from ANOVA-based methods when d=.2 and approximately 1.5%-17% higher when d =.8, while sample size was small. The differences of the power estimates increased from small to moderate sample sizes when d=.2. When sample size was large, the RMM test statistics generally provided 2%-3% higher empirical power estimates when d=.2; when d=.8, the power estimates from all test statistics reached 100%.

The power estimates at the "positive conditions" were comparatively much higher than those at the "negative conditions". At the "positive conditions" when sample size was small, the RMM test statistics generally provided higher empirical power estimates, among which the test statistics T_{ADF} and T_{YB2} delivered power estimates approximately 1%-28% higher than those from ANOVA-based methods when d =.2, and approximately 2%-24% higher when d =.8. When sample size increased, the discrepancy of the power estimates decreased. Especially when sample size became large, the empirical power estimates approached 100%. Overall, the RMM approaches generally provided better power estimates, especially when sample sizes were small or sample size were moderate with small effect size.

The power estimates when sample sizes were equal were comparatively much higher than those at the "negative conditions", but smaller than those at "positive conditions". When sample sizes were equal but small, the RMM test statistics generally provided higher empirical power estimates, among which the test statistics T_{YB2}, T_{ML} and T_{SB} delivered power estimates approximately 1.5% higher than those from ANOVA alternative methods when *d*=.2 and approximately 5%-17% higher when *d*=.8.

When sample size became moderate, similar power estimates were provided when variances were homogeneous and *d*=.2; approximately 4%-10% higher power estimates were yielded by the RMM approaches than those from the ANOVA-based methods, when the variance heterogeneity was moderate or large with *d*=.2. When the effect size increased to be large with moderate or large sample sizes and homogeneous variances, the RMM methods again yielded about power estimates about 5-7% higher than those from the ANOVA-based methods. When variances became heterogeneous with moderate or large sample sizes, the empirical power estimates approached 100% for *d*=.8.

In sum, the RMM test statistics generally provided higher empirical power estimates than the ANOVA-based method across all conditions when the effect size was small or the effect size was large with small or moderate sample sizes. With large sample sizes and large effect sizes, the empirical power estimates were all close to 100%. The empirical power estimates generally increased with sample size, effect size, as well as variance ratios except for the "negative conditions".

α= .05. When the nominal level became .05, similar results were found in the sense that, the SMM test statistics generally provided higher empirical power estimates than the ANOVA alternative methods across all conditions and the empirical power estimates increased with sample sizes, effect sizes as well as variance ratios except the "negative conditions". To be more specific, the RMM test statistics delivered power estimates approximately 2%-7% higher than those from the ANOVA alternatives when *d*=.2 and

Table 31. Empirical Power (%) analysis for Normal Distribution with k=2 at α =.01

n_1: n_2	σ_1/ σ_2	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
4:16	1	0.3	1.8	0.7	1.2	0.7	1.2	0.7	2.1	0.7	1.5	1.0	4.0			1.0	4.0	1.6	8.5	2.0	11.2	0.7	3.1
	2.5			0.7	0.8	0.7	0.8	1.0	1.6	0.7	1.0	0.9	3.1			0.9	3.1	2.6	9.5	3.8	12.4	0.8	2.5
	4			0.6	0.6	0.6	0.6	1.0	1.6	0.7	0.7	1.0	5.0			1.0	5.0	3.2	14.5	4.4	17.7	0.7	3.8
16:4	2.5	**0.2**	**37.1**	**1.8**	**32.1**	**1.8**	**32.1**	**2.7**	**49.3**	**2.4**	**35.8**	4.6	83.0	10.0	98.6	4.5	82.9	4.6	94.3	6.4	96.6	3.8	78.3
	4	**1.1**	**87.3**	**8.2**	**75.9**	**8.2**	**75.9**	**9.5**	**97.0**	10.5	84.7	23.4	100.0	36.6	100.0	23.4	100.0	20.1	100.0	27.5	100.0	20.0	100.0
10:10	1	0.6	11.6	0.5	9.3	0.5	9.3	0.5	9.2	0.7	11.0	2.2	16.8			2.2	16.8	1.4	12.2	2.2	17.2	1.5	14.6
	2.5	4.0	55.8	2.9	42.9	2.9	42.9	2.9	45.9	3.1	46.8	4.5	58.9			4.5	58.9	3.2	55.8	5.5	62.8	3.6	55.5
	4	9.3	91.6	5.9	77.8	5.9	77.8	6.1	80.7	6.6	81.7	8.2	94.9			8.2	94.8	7.1	94.0	9.7	96.6	6.9	93.1
20:80	1	3.4	60.6	3.7	54.5	3.7	54.5	3.7	55.1	3.7	54.6	2.8	64.5			2.8	64.5	3.1	67.3	3.2	67.5	2.8	63.6
	2.5			3.5	40.1	3.5	40.1	3.5	40.8	3.5	40.6	3.1	50.4			3.1	50.4	3.8	55.9	3.8	56.3	3.0	49.5
	4			4.6	58.9	4.6	58.9	4.9	59.4	4.9	59.9											3.9	73.0
80:20	2.5	**7.6**	**100.0**	50.4	100.0	50.4	100.0	50.7	100.0	51.3	100.0	59.1	100.0	61.5	100.0	59.1	100.0	57.6	100.0	57.8	100.0	58.1	100.0
	4	**48.0**	**100.0**	98.7	100.0	98.7	100.0	98.7	100.0	98.9	100.0	99.4	100.0	99.4	100.0	99.3	100.0	99.3	100.0	99.3	100.0	99.3	100.0
50:50	1	5.0	84.3	5.0	84.4	5.0	84.4	5.2	85.3	5.3	85.4	5.7	90.1	6.1	90.8	5.7	90.1	4.9	89.1	5.0	89.2	5.4	89.9
	2.5	21.4	100.0	20.7	100.0	20.7	100.0	21.2	100.0	21.4	100.0	26.2	100.0	28.3	100.0	26.2	100.0	25.5	100.0	25.6	100.0	25.7	100.0
	4	52.6	100.0	51.3	100.0	51.3	100.0	51.5	100.0	51.5	100.0	58.4	100.0	61.8	100.0	58.4	100.0	58.2	100.0	58.4	100.0	58.1	100.0
100:400	1	18.1	100.0	19.0	100.0	19.0	100.0	19.8	100.0	19.8	100.0	19.4	100.0	20.5	100.0	19.4	100.0	19.7	100.0	18.9	100.0	19.2	100.0
	2.5			14.4	100.0	14.4	100.0	14.4	100.0	14.5	100.0	16.4	99.9	17.1	99.9	16.4	99.9	16.9	99.9	16.4	99.9	16.3	99.9
	4	65.4	100.0	22.7	100.0	22.7	100.0	22.7	100.0	22.8	100.0	26.6	100.0	27.9	100.0	26.6	100.0	27.3	100.0	26.6	100.0	26.4	100.0
400:100	2.5	**97.5**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0												
	4	**100.0**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	10.0	100.0												
250:250	1	29.6	100.0	29.6	100.0	29.6	100.0	30.6	100.0	30.6	100.0	36.6	100.0	37.2	100.0	36.6	100.0	36.2	100.0	35.2	100.0	36.5	100.0
	2.5	91.4	100.0	91.4	100.0	91.4	100.0	91.9	100.0	92.0	100.0	95.5	100.0	95.8	100.0	95.5	100.0	95.4	100.0	95.2	100.0	95.4	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .01. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

approximately 9%-20% higher when *d* =.8 while sample size was small at the “negative conditions”. When sample size became moderate for the “negative conditions”, the RMM test statistics provided power estimates less than 3% higher when *d* =.2 and delivered empirical power approximately 7%-12% higher when *d* =.8. When sample size was large, the empirical power provided by the RMM test statistics were about 5% higher than those from the ANOVA alternatives for *d* =.2 and were all 100% when *d* =.8.

At the “positive conditions” when sample size was small, the RMM test statistics generally delivered power estimates approximately 7%-24% higher than those from ANOVA alternatives when *d* =.2 and above 22% higher when *d* =.8. When sample size became moderate, the SMM test statistics provided power estimates approximately 5% higher when *d* =.2 and variance heterogeneity ratio was moderate, and delivered about 100% empirical power estimates when *d* =.8 or variance heterogeneity ratio was large. When sample size was large, the empirical power estimates across all test statistics were about 100% at the “positive conditions”.

When sample sizes were equal and small, the RMM test statistics generally provided empirical power estimates 2-4% higher across all variance ratios when *d*=.2, and 2-8% higher across all variance ratios when *d*=.8, while the discrepancy decreased as variance ratios increased. When sample sizes were equal but became moderate, the RMM test statistics provided empirical power estimates 2-6% higher across all variance ratios when *d* =.2, and approached 100% when *d* =.8. When sample sizes were equal but became large, the RMM test statistics yielded empirical power estimates 7% higher when variances were homogenous and *d* =.2, and above 98% at all variance heterogeneity ratios when *d* =.2 and all conditions when *d* =.8.

Table 32: Empirical Power (%) analysis for Normal Distribution with k=2 at α=.05

n_1: n_2	σ_1/ σ_2	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
	1	2.7	8.3	**1.5**	**3.4**	**1.5**	**3.4**	**1.9**	**4.8**	**1.6**	**4.1**	4.8	17.5			4.8	17.5	6.9	23.2	7.6	25.5	4.0	15.2
4:16	2.5	**13.1**	**19.9**	**1.3**	**1.4**	**1.3**	**1.4**	**2.1**	**3.2**	**1.7**	**2.1**	4.8	14.0			4.8	14.0	8.0	22.0			4.1	12.1
	4	**19.4**	**32.2**	**0.8**	**1.1**	**0.8**	**1.1**	**1.9**	**3.1**	**1.4**	**2.4**	5.2	18.3			5.2	18.3					4.4	17.2
16:4	2.5	**2.2**	**79.4**	**6.5**	**52.0**	**6.5**	**52.0**	**8.5**	**77.1**	**7.6**	**60.5**	17.5	99.3	23.0	99.8	17.4	99.3	17.6	99.5	20.2	99.8	15.7	99.0
	4	**6.4**	**99.5**	**24.9**	**90.2**	**24.9**	**90.2**	**28.0**	**100.0**	27.6	99.2	53.3	100.0	58.3	100.0	53.0	100.0	50.7	100.0	53.9	100.0	49.2	100.0
	1	5.7	34.0	5.5	31.3	5.5	31.3	5.4	30.7	5.9	33.2	9.1	38.7			9.1	38.7	7.9	36.5	8.9	38.9	8.3	37.0
10:10	2.5	12.5	82.0	10.3	76.3	10.3	76.3	10.4	77.4	10.6	78.0	13.4	84.2			13.4	84.2	12.4	83.5	14.1	85.6	12.1	82.3
	4	21.7	98.4	17.2	96.6	17.2	96.6	18.2	97.0	18.5	97.2	21.4	99.5			21.4	99.4	21.2	99.5	23.7	99.6	20.2	99.4
	1	11.8	83.1	11.3	80.8	11.3	80.8	11.3	81.0	11.3	80.9	11.9	87.1	13.9	88.2	11.9	87.1	13.1	87.6	13.0	87.6	11.4	87.1
20:80	2.5			10.5	68.7	10.5	68.7	10.7	69.3	10.7	69.3	10.6	77.5	12.4	80.7	10.6	77.5	11.5	79.8	11.4	79.7	10.5	77.2
	4			12.2	85.3	12.2	85.3	12.4	85.5	12.4	85.5	13.7	91.9	15.5	92.9	13.7	91.9	14.9	92.7	14.8	92.7	13.4	91.8
80:20	2.5	**32.3**	**100.0**	75.4	100.0	75.4	100.0	75.2	100.0	75.4	100.0	80.3	100.0	81.5	100.0	80.3	100.0	79.8	100.0	79.8	100.0	80.0	100.0
	4	**87.4**	**100.0**	100.0	100.0	100.0	100.0	99.9	100.0	10.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	16.4	95.6	16.4	95.6	16.4	95.6	16.4	95.6	16.5	95.6	18.3	97.2	18.8	97.2	18.3	97.2	17.7	97.2	17.6	97.2	18.1	97.2
50:50	2.5	42.9	100.0	42.3	100.0	42.3	100.0	42.1	100.0	42.3	100.0	46.6	100.0	48.0	100.0	46.6	100.0	46.2	100.0	46.2	100.0	46.1	100.0
	4	75.2	100.0	74.6	100.0	74.6	100.0	74.7	100.0	74.8	100.0	79.4	100.0	80.4	100.0	79.4	100.0	79.4	100.0	79.4	100.0	79.2	100.0
	1	39.6	100.0	38.7	100.0	38.7	100.0	39.0	100.0	39.0	100.0	43.3	100.0	43.7	100.0	43.3	100.0	43.3	100.0	43.0	100.0	43.2	100.0
100:400	2.5			32.6	100.0	32.6	100.0	32.6	100.0	32.6	100.0	36.7	100.0	37.2	100.0	36.7	100.0	36.9	100.0	36.6	100.0	36.6	100.0
	4			44.9	100.0	44.9	100.0	44.9	100.0	44.9	100.0	48.8	100.0	49.6	100.0	48.8	100.0	49.3	100.0	48.8	100.0	48.8	100.0
400:100	2.5	**99.8**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	4	**100.0**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	53.5	100.0	53.5	100.0	53.5	100.0	54.0	100.0	54.1	100.0	61.6	100.0	61.6	100.0	61.6	100.0	61.5	100.0	61.0	100.0	61.5	100.0
250:250	2.5	98.5	100.0	98.5	100.0	98.5	100.0	98.5	100.0	98.5	100.0	98.6	100.0	98.6	100.0	98.6	100.0	98.6	100.0	98.5	100.0	98.5	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

α= .10. Once again, the RMM test statistics generally provided higher empirical power estimates than the ANOVA methods across all conditions. To be more specific, the RMM test statistics delivered power estimates approximately 3%-15% higher than those from ANOVA alternatives when d =.2 and approximately 18%-43% higher when d =.8, while sample size was small at the "negative conditions". When sample size became moderate for the "negative conditions", the RMM test statistics provided power estimates about 2% higher when d =.2 and delivered empirical power approximately 3%-6% higher when d =.8. When sample size was large, the empirical power provided by the RMM test statistics were about 3%-5% higher than those from the ANOVA alternatives and were all 100% when d =.8.

At the "positive conditions" when sample size was small, the RMM test statistics generally delivered power estimates approximately 7%-28% higher than those from ANOVA alternatives when d =.2, above 11-37% higher when d =.8 with moderately heterogeneous variances, and less than 6% higher above 11-37% higher when d =.8 with greatly heterogeneous variances. When sample size became moderate, the RMM test statistics provided power estimates approximately 4% higher for d =.2 with moderately heterogeneous variances, and delivered 100% empirical power estimates when d =.8 or the variance heterogeneity ratio was large. When sample size was large, the empirical power estimates across all test statistics were 100% at the "positive conditions".

When sample sizes were equal and small, the RMM test statistics generally provided empirical power estimates about 2-6% higher across all variance ratios when d =.2 or .8, with the empirical power estimates increased as variance ratios increased. When sample sizes were equal but became moderate, the RMM test statistics provided empirical power estimates approximately 4% higher across all variance ratios for d =.2, and approached 100% when d =.8. When sample sizes were equal but became large, the RMM test statistics yielded empirical power estimates 5% higher when variances were homogenous and d =.2, and

Table 33: Empirical Power (%) analysis for Normal Distribution with k=2 at α=.10

n_1: n_2	σ_1/ σ_2	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
4:16	1	6.0	14.1	**2.5**	**5.4**	**2.5**	**5.4**	**3.3**	**8.0**	**2.9**	**6.6**	10.1	30.1	14.8	39.1	10.1	30.1	11.8	35.6	13.2	36.6	8.9	27.1
	2.5			**1.6**	**1.9**	**1.6**	**1.9**	**3.0**	**4.7**	**2.5**	**3.7**	9.5	24.0	14.7	35.2	9.5	24.0	13.2	31.1	13.5	32.6	8.5	22.6
	4			**1.3**	**1.8**	**1.3**	**1.8**	**2.8**	**5.2**	**2.3**	**4.8**	9.7	31.1	16.4	45.3	9.7	31.1	13.9	40.6	15.3	41.7	8.6	29.0
16:4	2.5	**7.2**	**91.9**	**12.3**	**63.0**	**12.3**	**63.0**	**15.1**	**88.0**	**14.6**	**81.0**	29.1	99.8	33.1	99.9	29.0	99.8	30.1	99.9	30.6	99.9	27.1	99.8
	4	**16.3**	**99.9**	**39.4**	**94.5**	**39.4**	**94.5**	**44.7**	**100.0**	**43.3**	**100.0**	68.7	100.0	72.3	100.0	68.4	100.0	67.1	100.0	69.0	100.0	66.3	100.0
10:10	1	12.2	48.2	11.2	45.1	11.2	45.1	10.7	44.6	11.7	46.8	16.3	51.9			16.3	51.9	14.7	50.4	15.7	51.7	14.6	49.7
	2.5	20.2	91.0	18.0	88.5	18.0	88.5	18.2	88.8	18.9	89.2	21.4	91.9	24.1	93.4	21.4	91.9	20.9	91.8	21.6	92.4	19.6	91.0
	4	32.9	99.7	28.2	99.3	28.2	99.3	28.9	99.5	29.5	99.5	33.4	99.7	37.2	99.8	33.4	99.6	33.1	99.7	34.5	99.8	30.9	99.6
20:80	1	18.9	90.8	19.0	89.3	19.0	89.3	18.9	89.2	18.9	89.1	20.1	92.3	21.2	93.0	20.1	92.3	20.3	92.7	20.3	92.6	19.8	92.3
	2.5			16.7	80.6	16.7	80.6	16.7	81.0	16.6	81.0	18.1	87.0	20.0	88.3	18.1	87.0	19.5	87.9	19.3	87.8	17.9	87.0
	4			20.4	93.3	20.4	93.3	20.5	93.4	20.5	93.4	22.4	96.1	23.9	96.7	22.4	96.1	23.1	96.5	23.1	96.5	22.0	96.1
80:20	2.5	**51.1**	**100.0**	85.1	100.0	85.1	100.0	85.1	100.0	85.2	100.0	89.2	100.0	89.4	100.0	89.2	100.0	88.8	100.0	88.6	100.0	88.7	100.0
	4	**95.9**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
50:50	1	24.6	97.9	24.6	97.9	24.8	97.9	24.8	97.9	25.0	97.9	27.6	99.0	27.7	99.2	27.6	99.0	27.4	99.0	27.2	99.0	27.5	99.0
	2.5	54.4	100.0	54.4	100.0	54.4	100.0	54.4	100.0	54.4	100.0	58.9	100.0	59.8	100.0	58.9	100.0	58.8	100.0	58.8	100.0	58.7	100.0
	4	83.9	100.0	83.7	100.0	83.7	100.0	83.7	100.0	83.7	100.0	88.0	100.0	88.3	100.0	88.0	100.0	88.0	100.0	88.0	100.0	87.9	100.0
100:400	1	51.4	100.0	50.5	100.0	50.5	100.0	50.7	100.0	50.7	100.0	55.8	100.0	55.9	100.0	55.8	100.0	55.8	100.0	55.6	100.0	55.8	100.0
	2.5			44.5	100.0	44.5	100.0	44.5	100.0	44.5	100.0	47.6	100.0	48.1	100.0	47.6	100.0	47.9	100.0	47.4	100.0	47.5	100.0
	4			57.4	100.0	57.4	100.0	57.4	100.0	57.4	100.0	62.3	100.0	62.6	100.0	62.3	100.0	62.5	100.0	62.3	100.0	62.3	100.0
400:100	2.5	**100.0**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	4	**100.0**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
250:250	1	66.2	100.0	66.2	100.0	66.2	100.0	66.7	100.0	66.7	100.0	71.6	100.0	71.6	100.0	71.6	100.0	71.6	100.0	71.3	100.0	71.6	100.0
	2.5	99.2	100.0	99.2	100.0	99.2	100.0	99.2	100.0	99.2	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.

Power was computed across 1000 replications at the nominal level of .10. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

yielded empirical power estimates above 98% at all variance heterogeneity ratios when d =.2 and provided empirical power estimates of 100% for condition when d =.8.

K=3

α= .01. At the "negative conditions" when sample size was small, the RMM test statistics provided empirical power estimates at around 2-3.5% for f=.1 and 55-60% for f=.4. At the "negative condition" when sample size became moderate, the RMM test statistics generally provided empirical power estimates between 5-7% for f=.1 and 93-94.5% for f=.4, about 2-4% and 14% higher than those from the ANOVA alternatives respectively. When sample size increased to large, the RMM test statistics again provided empirical power estimates 8-10% higher than those from the ANOVA alternatives, falling between 38-40% when f=.1. When sample size was large and f=.4, the power estimates increased to 100%.

The power estimates at the "positive conditions" were comparatively higher than those at the "negative conditions". When f=.1, the RMM test statistics generally provided empirical power estimates approximately 1%-5%, 7-12% and 5% higher than those from most of the ANOVA alternative methods for small, moderate and large sample sizes respectively. When f=.4, the power estimates were all very high, above 98%, increasing with sample sizes.

When sample sizes were equal and small, the power estimates were close to each other, falling at the range of 2.7-6.5% for f=.1; when f= .4, the RMM test statistics generally provided empirical power estimates approximately 7%-16% higher than those from ANOVA-based methods. When sample sizes were equal and moderate, the power estimates fell at the range of 7-10% for f=.1 with heterogeneous variances, and fell at the range of 5.2-6.5% when f=.1 with homogeneous variances; when f increased to .4, all ANOVA-based methods provided power estimates greater than 96.2%, while the RMM test statistics yielded power estimates greater than 99.8%. When sample sizes were equal and large, the RMM test

Table 34: Empirical Power (%) analysis for Normal Distribution with k=3 at α =.01

n_1: n_2: n_3	$\sigma_1/\sigma_2/\sigma_3$	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4			4.4	81.6			4.5	82.7	4.3	81.5	3.5	92.0	5.6	94.0	3.5	92.0	3.6	91.0	3.8	91.6	3.3	91.3
	1:1:1	3.7	68.2	4.2	65.1	3.7	68.3	4.3	65.6	4.3	65.8	3.4	78.1	4.4	81.5	3.4	78.1	2.7	76.6	3.2	77.9	3.1	77.3
	4:2.5:1											2.2	56.7			2.2	56.7	3.3	58.5	3.5	59.6	2.0	55.4
12:30:48	1:2.5:4	**1.5**	**64.3**	5.9	98.3			5.8	98.3	6.2	98.3	7.2	100.0	10.6	100.0	7.2	100.0	7.2	100.0	7.7	100.0	7.2	100.0
	1:1:1	1.9	59.0			2.4	54.6			2.5	49.0	3.3	74.9	5.2	82.1	3.3	74.9	3.3	77.2	3.3	78.3	3.0	73.3
60:60:60	1:2.5:4	7.4	97.5	8.4	99.5	7.1	97.5	9.5	99.5	9.0	99.5	8.4	100.0	9.6	100.0	8.4	100.0	8.2	100.0	8.5	100.0	8.2	100.0
	1:1:1	6.3	97.3	6.4	97.2	6.4	97.3	6.5	97.2	6.4	97.2	5.4	99.8	6.1	99.8	5.4	99.8	5.2	99.8	5.3	99.8	5.2	99.8
	4:2.5:1			3.5	80.9			3.5	84.5	3.2	80.6	5.6	93.8	6.6	94.4	5.6	93.8	6.3	93.9	6.3	94.0	5.5	93.7
24:60:96	1:2.5:4	3.0	99.2	14.6	100.0			15.2	100.0	15.7	100.0	23.3	100.0	25.8	100.0	23.3	100.0	23.1	100.0	23.4	100.0	23.2	100.0
	1:1:1	4.6	95.9	4.2	93.8	4.8	94.6	5.2	95.2	4.3	94.1	7.8	99.0	9.8	99.1	7.8	99.0	8.3	99.0	8.6	99.0	7.6	98.9
300:300:300	1:2.5:4			59.9	100.0			61.4	100.0	61.1	100.0	73.1	100.0	73.6	100.0	73.1	100.0	73.1	100.0	73.5	100.0	72.9	100.0
	1:1:1	42.7	100.0	42.1	100.0	42.7	100.0	44.1	100.0	44.5	100.0	55.4	100.0	56.3	100.0	55.4	100.0	55.1	100.0	56.2	100.0	55.4	100.0
	4:2.5:1			29.8	100.0			30.5	100.0	29.9	100.0	38.7	100.0	39.6	100.0	38.7	100.0	39.0	100.0	39.4	100.0	38.7	100.0
120:300:480	1:2.5:4	**35.0**	**100.0**	93.2	100.0	73.3	100.0	93.7	100.0	93.8	100.0	97.9	100.0	97.9	100.0	97.9	100.0	97.9	100.0	97.9	100.0	97.9	100.0
	1:1:1	43.4	100.0	44.2	100.0	44.0	100.0	45.3	100.0	45.4	100.0	57.0	100.0	58.0	100.0	57.0	100.0	57.1	100.0	57.8	100.0	57.0	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.

Power was computed across 1000 replications at the nominal level of .01. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

statistics generally provided empirical power estimates approximately 11%-14% higher than those from ANOVA methods for f=.1 and all power estimates increased to 100% .

α= .05. At the "negative conditions" when sample sizes were small, the RMM test statistics provided empirical power estimates at much as 5.1% higher than those from ANOVA alternatives for f=.1; when f=.4, the power estimates for the RMM test statistics fell around 80% while most of the power estimates for the ANOVA alternatives fell around 60%, with the Alexander and Govern A yielding the highest value of 65.5% and BF F^* yielding the lowest value of 39.1%. At the "negative condition" when sample size became moderate, the RMM test statistics generally provided empirical power estimates 3-10% higher than those from the ANOVA alternatives for f=.1; when f=.4, the empirical power estimates all fell above 94%. When sample size increased to large with f=.1, the power estimates for the RMM test statistics fell around 64% while most of the power estimates for the ANOVA alternatives except for the BF F^* fell around 54%; when sample size was large and f=.4, all the power estimates increased to 100%.

At the "positive conditions" when sample size was small, the RMM test statistics provided empirical power estimates up to 6% higher than those from ANOVA alternatives methods for f=.1; when f=.4, the power estimates for all the methods fell above 98.0%. At the "positive condition" when sample size became moderate, the RMM test statistics generally provided empirical power estimates 7-17% higher than those from the ANOVA alternatives for f=.1; when f=.4, the empirical power estimates were all 100%. When sample size increased to large, the power estimates were all above 90% for both f=.1 and f=.4.

When sample sizes were equal and small, the power estimates provided by the RMM test statistics were slightly higher but close to those yielded by the ANOVA methods for both f=.1 and .4. When sample sizes were equal and moderate, the power estimates from all test

Table 35: Empirical Power (%) analysis for Normal Distribution with k=3 at α=.05

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4	12.6	86.0	14.0	93.3	12.1	84.7	14.2	93.7	14.0	93.3	13.3	98.8	15.6	99.0	13.3	98.8	13.3	98.9	13.3	98.9	13.1	98.7
	1:1:1	11.6	85.4	12.0	85.5	11.7	85.4	11.9	85.4	12.0	85.5	12.0	93.6	13.5	94.3	12.0	93.6	11.6	92.9	11.8	93.1	11.6	93.0
	4:2.5:1			9.2	60.0	7.5	39.1	9.5	65.5	9.1	59.8	10.2	79.2	12.6	82.5	10.2	79.2	11.3	81.1	11.6	81.2	9.6	78.6
12:30:48	1:2.5:4	**5.6**	**88.3**	16.6	99.8	16.4	98.0	16.4	99.8	16.6	99.9	23.1	100.0	27.1	100.0	23.1	100.0	22.7	100.0	22.9	100.0	22.6	100.0
	1:1:1	8.9	81.9	9.1	77.3	9.7	79.1	10.1	79.7	9.1	77.2	10.9	92.0	12.6	94.4	10.9	92.0	11.4	92.6	11.4	92.7	10.5	91.7
60:60:60	1:2.5:4	17.7	99.9	22.1	99.9	17.5	99.9	23.3	99.9	23.0	99.9	23.8	100.0	25.4	100.0	23.8	100.0	23.7	100.0	23.8	100.0	23.2	100.0
	1:1:1	19.4	99.4	20.1	99.5	19.6	99.4	20.2	99.5	20.4	99.5	19.5	99.9	20.4	99.9	19.5	99.9	19.0	99.9	19.3	99.9	19.1	99.9
	4:2.5:1			11.8	94.9	7.9	77.4	12.3	99.5	11.8	94.8	16.8	98.1	17.9	98.5	16.8	98.1	17.2	98.2	17.3	98.2	16.5	98.0
24:60:96	1:2.5:4	**10.4**	**100.0**	37.6	100.0	30.0	100.0	37.4	100.0	37.9	100.0	45.4	100.0	47.0	100.0	45.4	100.0	45.4	100.0	45.4	100.0	45.0	100.0
	1:1:1	14.5	98.9	14.1	98.8	14.1	99.0	14.5	98.7	14.5	98.9	20.5	99.7	21.9	99.8	20.5	99.7	21.0	99.8	21.1	99.8	20.2	99.6
300:300:300	1:2.5:4	64.4	100.0	82.1	100.0	64.4	100.0	83.2	100.0	83.1	100.0	88.5	100.0	88.7	100.0	88.5	100.0	88.5	100.0	87.8	100.0	88.5	100.0
	1:1:1	66.9	100.0	67.3	100.0	66.9	100.0	68.3	100.0	68.4	100.0	78.6	100.0	78.8	100.0	78.6	100.0	78.6	100.0	78.2	100.0	78.6	100.0
	4:2.5:1			54.4	100.0	38.2	100.0	54.7	100.0	54.4	100.0	64.2	100.0	64.9	100.0	64.2	100.0	64.3	100.0	63.4	100.0	64.0	100.0
120:300:480	1:2.5:4	**66.6**	**100.0**	97.9	100.0	90.6	100.0	98.1	100.0	98.1	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.5	100.0	99.7	100.0
	1:1:1	66.9	100.0	67.2	100.0	67.3	100.0	67.9	100.0	67.6	100.0	78.4	100.0	78.8	100.0	78.4	100.0	78.6	100.0	77.4	100.0	78.4	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

statistics fell between 17.7-25.4% when f=.1; when f increased to .4, all of the methods provided high power estimates, above 99%. When sample sizes were equal and large, most of the test statistics except the ANOVA F and BF F^* statistic yielded empirical power estimates above 88% with heterogeneous variances; when variances became homogenous, the RMM test statistics generally provided empirical power estimates at around 78.5%, approximately 10% higher than the power estimates provided by most of the ANOVA-based methods.

α= .10. At the "negative conditions" when sample size was small, the power estimates for the RMM test statistics fell around 17% while most of the power estimates for the ANOVA alternatives fell around 14% for f=.1; when f=.4, the power estimates for the RMM test statistics fell around 88% while most of the power estimates for the ANOVA alternatives fell around 75%. At the "negative condition" when f=.1, the power estimates for the RMM test statistics fell around 26% and 75.5% for moderate and large sample sizes respectively, while most of the power estimates for the ANOVA alternatives fell around 19% and 67% for moderate and large sample sizes respectively. When f=.4, all the power estimates increased to above 97%.

At the "positive conditions" when sample size was small, the power estimates for the RMM test statistics fell around 36% while most of the power estimates for the ANOVA alternatives fell around 28% for f=.1. At the "positive condition" when sample size became moderate, the RMM test statistics generally provided empirical power estimates 6% higher than most of the power estimates from the ANOVA alternatives for f=.1; when sample size became large, the empirical power estimates were all above 95% with f=.1. When f=.4, the power estimates increased to 100% across all sample size conditions at the "positive conditions".

Table 36: Empirical Power (%) analysis for Normal Distribution with k=3 at α=.10

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4	18.5	92.2	22.3	97.9	17.6	91.5	22.0	98.0	22.2	97.9	22.9	99.7	25.3	99.8	22.9	99.7	23.0	99.7	23.1	99.7	22.1	99.7
	1:1:1	18.9	91.8	19.1	91.7	18.9	91.8	19.2	91.6	19.2	91.9	21.2	96.3	22.9	96.6	21.2	96.3	21.0	96.2	21.0	96.2	21.0	96.2
12:30:48	4:2.5:1			14.6	74.3	11.5	52.2	15.1	76.7	14.6	74.1	16.8	87.8	19.2	89.8	16.8	87.8	17.8	89.0	17.9	89.1	16.7	87.8
	1:2.5:4	**9.4**	**94.7**	28.2	100.0	23.8	99.2	28.3	100.0	28.4	100.0	36.2	100.0	37.8	100.0	36.2	100.0	35.8	100.0	35.8	100.0	35.6	100.0
	1:1:1	15.4	89.3	16.4	87.5	16.0	88.6	17.2	87.5	16.3	87.4	18.7	96.6	21.7	97.1	18.7	96.6	19.3	96.9	19.5	96.9	18.5	96.6
60:60:60	1:2.5:4	26.5	99.9	33.2	100.0	26.4	99.9	34.3	100.0	34.1	100.0	36.9	100.0	37.7	100.0	36.9	100.0	36.6	100.0	36.6	100.0	36.4	100.0
	1:1:1	28.1	99.9	29.7	99.8	28.3	99.9	29.7	99.8	29.8	99.8	30.4	100.0	31.3	100.0	30.4	100.0	30.3	100.0	30.3	100.0	30.2	100.0
24:60:96	4:2.5:1			19.2	97.4	13.0	89.1	19.8	97.5	19.1	97.2	25.7	99.5	28.0	99.6	25.7	99.5	26.2	99.5	26.3	99.5	25.5	99.4
	1:2.5:4	**19.1**	**100.0**	51.1	100.0	28.9	100.0	51.0	100.0	51.5	100.0	57.1	100.0	58.3	100.0	57.1	100.0	56.6	100.0	56.7	100.0	56.6	100.0
	1:1:1	22.9	99.6	22.3	99.2	21.8	99.4	22.6	99.3	22.7	99.3	31.0	99.9	33.4	99.9	31.0	99.9	31.9	99.9	32.0	99.9	31.0	99.9
300:300:300	1:2.5:4	76.2	100.0	88.9	100.0	76.2	100.0	89.1	100.0	89.1	100.0	94.2	100.0	94.3	100.0	94.2	100.0	94.1	100.0	94.0	100.0	94.2	100.0
	1:1:1	78.5	100.0	77.9	100.0	78.5	100.0	78.3	100.0	78.3	100.0	85.8	100.0	85.9	100.0	85.8	100.0	85.7	100.0	85.5	100.0	85.7	100.0
120:300:480	4:2.5:1			67.3	100.0	50.0	100.0	67.5	100.0	67.3	100.0	75.6	100.0	75.8	100.0	75.6	100.0	75.7	100.0	75.3	100.0	75.6	100.0
	1:2.5:4	**80.5**	**100.0**	98.9	100.0	95.2	100.0	99.0	100.0	99.0	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0
	1:1:1	79.0	100.0	79.0	100.0	78.7	100.0	79.2	100.0	79.5	100.0	86.2	100.0	86.5	100.0	86.2	100.0	86.2	100.0	85.7	100.0	86.1	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .10. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

When sample sizes were equal and small, the power estimates were close to each other, falling within the range of 17.6-25.3% for f=.1; when f became .4, the RMM test statistics generally provided empirical power estimates falling between 96-99.8%, less than 8% higher than those from ANOVA-based methods. When sample sizes were equal and moderate, the power estimates fell between the range of 26.5-37.7% when f=.1; when f increased to .4, all of the methods provided power estimates high above 99.8%. When sample sizes were equal and large, most of the test statistics except the ANOVA *F* and BF *F** statistic yielded empirical power estimates above 88% with heterogeneous variances; when variances became homogenous, the RMM test statistics generally provided empirical power estimates at around 86%, approximately 8% higher than the power estimates provided by the ANOVA-based methods.

K=4

<u>*α= .01*</u>. At the "negative conditions" when sample size was small, the RMM test statistics provided empirical power estimates slightly higher (less than 2%) than the power estimates from the ANOVA alternatives when f=.1; when f increased to .4, the RMM test statistics provided empirical power estimates at about 92%, while the ANOVA alternatives yielded empirical power estimates falling between 30.6-83.8% with the highest value from the Alexander and Govern *A* and lowest value from BF *F**. At the "negative conditions" when sample size became moderate, the RMM test statistics again provided empirical power estimates slightly higher (4%) when f=.1; when f=.4, the power estimates were high, above 99.7%. When sample size increased to large with f=.1, the RMM test statistics as well as the ANOVA alternatives provided empirical power estimates at about 72% and 65% respectively; when f=.4, the power estimates became 100%.

As usual, the power estimates at the “positive conditions” were comparatively higher than those at the “negative conditions”. When sample sizes were small with f=.1, the empirical power estimates were fairly close to each other across the RMM and ANOVA alternative test statistics, falling at around 8%; when sample sizes increased to moderate with f=.1, the RMM test statistics generally provided empirical power estimates approximately 3%-8% higher than those from ANOVA alternatives. The power estimates were all very high above 99% for the RMM and ANOVA alternative test statistics, when f=.4 across sample sizes or when sample sizes became large.

When sample sizes were equal and small and variances were heterogeneous, the power estimates were close to each other across all methods, falling at around 5.5% for f=.1 and within the range of 95.7-99% for f=.4; when variances became homogenous, the power estimates yielded by the RMM methods were slightly higher than those provided by the ANOVA-based methods, falling at around 4% for f=.1 and 83% for f=.4. When sample sizes were equal and moderate and f=.1, the RMM test statistics generally provided empirical power estimates approximately 3%-12% higher than those from ANOVA alternatives, falling at around 16% with heterogeneous variances and 8% with homogenous variances. Also, when sample sizes were equal and large and f=.1, the SMM test statistics generally provided empirical power estimates approximately 3-5% higher than those from ANOVA alternative methods with heterogeneous variances and 10% higher with homogenous variances. When f became .4 with moderate or large sample sizes, the power estimates became high, above 99%.

Table 37: Empirical Power (%) analysis for Normal Distribution with k=4 at α =.01

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4			5.1	95.7					5.2	95.9	5.9	99.0			5.9	99.0	5.4	99.0	5.3	99.0	5.6	99.0
	1:1:1:1	3.0	80.9	2.9	75.1	3.0	80.7	3.2	76.9	3.1	78.3	4.2	83.5			4.2	83.5	4.0	83.1	4.0	83.0	4.1	83.0
10:20:40:50	4:3:2:1			3.8	74.2	2.4	30.6	3.4	83.8	2.9	68.8	4.1	92.5			4.1	92.5	5.0	92.3	4.9	92.2	3.9	92.2
	1:2:3:4	**1.5**	**75.2**	7.9	99.5			8.1	99.1	8.1	99.5	7.7	99.8			7.7	99.8	8.2	99.8	8.1	99.8	7.3	99.8
	1:1:1:1	2.4	80.0	3.7	69.6	2.2	72.0	3.2	74.3	3.2	66.4	3.4	84.2			3.4	84.2	4.6	85.9	4.4	85.6	3.2	83.9
60:60:60:60	1:2:3:4			12.4	100.0	6.1	99.5	13.1	100.0	12.7	100.0	15.7	100.0	17.6	100.0	15.7	100.0	15.3	100.0	15.1	100.0	15.5	100.0
	1:1:1:1	6.1	99.0	5.9	99.0	6.3	99.0	6.1	99.1	6.0	99.1	8.3	99.7	9.3	99.9	8.3	99.7	7.9	99.6	7.9	99.5	8.1	99.6
20:40:80:100	4:3:2:1			7.1	99.7			8.4	99.7	7.1	99.7	9.2	100.0			9.2	100.0	9.5	100.0	9.5	100.0	9.1	100.0
	1:2:3:4	**4.2**	**99.7**	21.1	100.0			21.5	100.0	21.9	100.0	25.0	100.0	29.3	100.0	25.0	100.0	25.9	100.0	25.8	100.0	24.6	100.0
	1:1:1:1	7.8	99.1	6.7	99.1	7.4	98.9	7.7	99.1	6.5	99.0	7.4	99.5	9.8	99.6	7.4	99.5	8.1	99.6	7.9	99.6	7.0	99.5
300:300:300:300	1:2:3:4			84.2	100.0			85.9	100.0	85.8	100.0	90.6	100.0	91.0	100.0	90.6	100.0	90.5	100.0	89.4	100.0	90.6	100.0
	1:1:1:1	55.5	100.0	55.2	100.0	55.6	100.0	56.8	100.0	56.9	100.0	65.5	100.0	65.9	100.0	65.5	100.0	65.6	100.0	63.7	100.0	65.4	100.0
100:200:400:500	4:3:2:1			64.5	100.0			66.4	100.0	65.2	100.0	71.6	100.0	72.0	100.0	71.6	100.0	71.5	100.0	68.8	100.0	71.5	100.0
	1:2:3:4	**45.1**	100.0	96.0	100.0	76.8	100.0	96.4	100.0	96.4	100.0												
	1:1:1:1	54.8	100.0	54.4	100.0	54.3	100.0	56.6	100.0	56.0	100.0	63.9	100.0	64.7	100.0	63.9	100.0	63.9	100.0	61.9	100.0	63.8	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .01. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

$\underline{\alpha = .05}$. At the "negative conditions" when sample size was small, the empirical power estimates from most of the ANOVA alternative and RMM test statistics were similar when f=.1; when f increased to .4, the RMM test statistics provided empirical power estimates about 2-6% higher than those from the ANOVA alternatives. At the "negative conditions" when sample size became moderate, the RMM test statistics generally provided empirical power estimates about 2-6% higher when f=.1; when f=.4, the power estimates were all 100%. When sample size increased to large with f=.1, the RMM test statistics as well as most of the ANOVA alternatives provided empirical power estimates of 88% and 84% respectively; when f=.4, the power estimates became 100%.

As usual, the power estimates at the "positive conditions" were comparatively higher than those at the "negative conditions". When sample sizes were small with f=.1, the empirical power estimates from the RMM test statistics were slightly higher, falling between 23.2-29.5% with the T_{ADF} test statistic yielding the highest power estimate value. When sample sizes increased to moderate with f=.1, the RMM test statistics and most of the ANOVA methods generally provided empirical power estimates at approximately 47% and 42% respectively, except for the T_{ADF} test statistic and BF F^*. In this situation, the T_{ADF} test statistic again provided the highest power estimate of 50.4%; on the contrary, the BF F^* yielded the lowest power estimate of 27.4%. The power estimates were all above 98.5%, when f=.4 across sample sizes.

When sample sizes were equal and small and variances were heterogeneous, the power estimates from the RMM test statistics were most 5% higher than the power estimates from the ANOVA-based methods, falling at the range of 12.6-20.7% for f=.1 and all falling above 99% except the ANOVA F for f=.4; when variances became homogenous, the power estimates fell in the range of 11-15.5% for f=.1 and 91.5-95% for f=.4 with slightly higher values yielded by the RMM methods. When sample sizes were equal and moderate and f=.1 with

Table 38: Empirical Power (%) analysis for Normal Distribution with k=4 at α=.05

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4	13.7	92.8	16.1	99.3	12.6	92.1	16.9	99.3	16.5	99.3	18.3	99.8	20.7	99.9	18.3	99.8	18.1	99.8	17.7	99.8	17.7	99.8
	1:1:1:1	11.2	93.0	11.2	91.7	11.0	92.8	11.9	91.5	12.0	92.2	13.0	94.5	15.5	95.0	13.0	94.5	12.7	94.6	12.7	94.5	12.8	94.4
10:20:40:50	4:3:2:1			13.2	92.5	9.4	58.0	14.7	95.7	12.3	91.3	13.6	97.7	17.8	98.1	13.6	97.7	15.2	97.7	14.7	97.7	13.4	97.5
	1:2:3:4	**5.9**	**92.9**	22.3	100.0	16.7	98.5	21.3	99.8	22.6	100.0	23.8	100.0	29.5	100.0	23.8	100.0	25.4	100.0	25.2	100.0	23.2	100.0
	1:1:1:1	11.9	93.0	12.2	88.9	11.1	89.4	11.6	89.9	11.9	88.3	12.7	94.6	16.2	96.6	12.7	94.6	14.6	95.4	14.2	95.4	12.2	94.5
60:60:60:60	1:2:3:4	16.9	100.0	29.2	100.0	16.2	100.0	30.3	100.0	29.5	100.0	34.4	100.0	36.6	100.0	34.4	100.0	34.1	100.0	33.8	100.0	34.0	100.0
	1:1:1:1	18.0	99.9	17.5	99.9	18.1	99.9	18.2	99.9	18.3	99.9	22.6	100.0	23.6	100.0	22.6	100.0	22.4	100.0	22.4	100.0	22.6	100.0
20:40:80:100	4:3:2:1			20.9	100.0			21.8	100.0	20.9	100.0	23.8	100.0	26.6	100.0	23.8	100.0	24.9	100.0	24.9	100.0	23.7	100.0
	1:2:3:4	**11.8**	**100.0**	42.0	100.0	27.4	100.0	42.4	100.0	42.7	100.0	47.4	100.0	50.4	100.0	47.4	100.0	48.3	100.0	48.2	100.0	47.2	100.0
	1:1:1:1	19.1	100.0	19.1	100.0	19.2	100.0	20.1	100.0	19.1	100.0	20.9	100.0			20.9	100.0	22.0	99.9	22.0	99.9	20.9	99.9
300:300:300:300	1:2:3:4	76.1	100.0	95.2	100.0	76.1	100.0	95.5	100.0	95.5	100.0	97.1	100.0	97.2	100.0	97.1	100.0	97.1	100.0	96.9	100.0	97.1	100.0
	1:1:1:1	77.1	100.0	76.6	100.0	77.1	100.0	77.2	100.0	77.3	100.0	83.2	100.0	83.4	100.0	83.2	100.0	83.2	100.0	82.6	100.0	83.2	100.0
100:200:400:500	4:3:2:1			83.4	100.0	46.8	100.0	84.1	100.0	83.6	100.0	88.2	100.0	88.4	100.0	88.2	100.0	88.2	100.0	87.8	100.0	88.1	100.0
	1:2:3:4	**73.5**	**100.0**	98.9	100.0	90.9	100.0	99.0	100.0	99.1	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.7	100.0	99.8	100.0
	1:1:1:1	78.3	100.0	77.4	100.0	78.3	100.0	77.9	100.0	78.0	100.0	81.5	100.0	81.9	100.0	81.5	100.0	81.7	100.0	80.7	100.0	81.4	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

heterogeneous variances, the RMM test statistics generally provided empirical power estimates of approximately 33%-37% while most of the ANOVA methods yielded power estimates of 16.2-30.3%; when variances were homogenous, the power estimates from the RMM test statistics and the ANOVA methods fell at around 23% and 18% respectively. Also, when sample sizes were equal and large and f=.1, the RMM test statistics generally provided empirical power estimates approximately 2% higher than those from ANOVA alternatives with heterogeneous variances and about 6% higher with homogenous variances. When f=.4, the power estimates yielded by both the RMM approaches and ANOVA alternatives became high, above 98%, for moderate and large sample sizes.

α= .10. At the "negative conditions" across sample sizes, the RMM test statistics provided slightly higher empirical power estimates than those from the ANOVA alternatives with the T_{ADF} test statistic and BF F^* providing the highest and lowest power estimates respectively for f=.1. At the "negative conditions" when sample size was small and f=.4, the RMM test statistics again provided slightly higher empirical power estimates than those from the ANOVA alternatives with the T_{ADF} test statistic and BF F* providing the highest and lowest power estimates respectively; however, when sample sizes increased, most of the power estimates became 100%.

As usual, the power estimates at the "positive conditions" were comparatively higher than those at the "negative conditions". When sample sizes were small with f=.1, the empirical power estimates from the RMM test statistics were slightly higher, falling between 35.6-39.1% with the T_{ADF} test statistic yielding the highest power estimate value of 39.1% and BF F^* yielding the lowest power estimate of 24.4%. When sample sizes increased to moderate with f=.1, the RMM test statistics and most of the ANOVA alternative methods generally provided empirical power estimates at approximately 61% and 56% respectively, except for the T_{ADF} test statistic and BF F^*. It this situation, the T_{ADF} test statistic provided the highest power estimate of 62.9%; on the contrary, the BF F^* yielded the lowest power estimate of 37.2%. The power estimates were all above 99.4%, when f=.4 across sample sizes.

When sample sizes were equal and small and variances were heterogeneous, the power estimates from the RMM test statistics were at maximum 10% higher than the power estimates from the ANOVA alternatives, falling at the range of 20.1-31.9% for f=.1 and falling above 99.9% for f=.4; when variances became homogenous, the power estimates became lower, falling at the range of 19.7-24.6% for f=.1 and 95.8-97.9% for f=.4. When sample sizes were equal and moderate and f=.1 with heterogeneous variances, the RMM test statistics generally provided empirical power estimates of approximately 47% while the ANOVA alternatives except the BF F^* yielded the power estimates at around 29%; when variances were homogenous, the power estimates from the RMM test statistics and the ANOVA methods fell at around 34% and 29% respectively. Also, when sample sizes were equal and large and f=.1, the RMM test statistics and most of the ANOVA alternatives generally provided similar empirical power estimates at approximately 97% with heterogeneous variances, and the RMM test statistics yielded power estimates about 5% higher than those from most of the ANOVA alternatives with homogenous variances. When f=.4, the power estimates became 100% for moderate and large sample sizes.

Table 39: Empirical Power (%) analysis for Normal Distribution with k=4 at α=.10

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4	20.7	96.7	26.9	99.5	20.1	96.4	26.7	99.5	27.3	99.5	29.8	100.0	31.9	100.0	29.8	100.0	28.9	100.0	28.9	100.0	28.8	100.0
	1:1:1:1	20.1	96.7	20.6	95.8	19.7	96.8	21.4	95.8	21.8	96.0	22.7	97.6	24.6	97.9	22.7	97.6	22.5	97.6	22.3	97.5	22.2	97.2
10:20:40:50	4:3:2:1			21.4	97.0	14.8	71.3	22.4	97.6	21.2	97.0	22.9	99.0	26.0	99.2	22.9	99.0	24.0	99.1	23.7	99.0	22.8	99.0
	1:2:3:4	**9.9**	**96.7**	34.1	100.0	24.4	99.4	33.1	100.0	34.2	100.0	35.6	100.0	39.1	100.0	35.6	100.0	36.5	100.0	35.8	100.0	34.9	100.0
	1:1:1:1	19.4	96.7	19.7	94.8	18.9	95.4	21.0	94.8	19.6	94.8	21.8	97.5	24.5	98.2	21.8	97.5	22.5	97.9	22.3	97.9	21.1	97.3
60:60:60:60	1:2:3:4	25.0	100.0	41.5	100.0	24.6	100.0	41.9	100.0	42.0	100.0	47.5	100.0	48.4	100.0	47.5	100.0	47.3	100.0	47.1	100.0	47.1	100.0
	1:1:1:1	29.5	100.0	28.8	100.0	29.6	100.0	29.2	100.0	29.8	100.0	33.6	100.0	35.0	100.0	33.6	100.0	33.7	100.0	33.5	100.0	33.4	100.0
20:40:80:100	4:3:2:1			31.6	100.0	19.5	98.1	33.3	100.0	32.3	100.0	35.5	100.0	38.2	100.0	35.5	100.0	36.5	100.0	36.2	100.0	35.2	100.0
	1:2:3:4	**18.7**	**100.0**	56.8	100.0	37.2	100.0	56.2	100.0	57.3	100.0	61.1	100.0	62.9	100.0	61.1	100.0	61.5	100.0	61.4	100.0	61.0	100.0
	1:1:1:1	30.5	100.0	29.2	100.0	28.8	100.0	29.3	100.0	29.5	100.0	32.5	100.0	34.8	100.0	32.5	100.0	32.8	100.0	32.5	100.0	32.2	100.0
300:300:300:300	1:2:3:4	85.0	100.0	97.6	100.0	85.0	100.0	97.7	100.0	97.7	100.0	98.9	100.0	98.9	100.0	98.9	100.0	98.9	100.0	98.9	100.0	98.9	100.0
	1:1:1:1	85.8	100.0	85.8	100.0	85.8	100.0	86.3	100.0	86.5	100.0	90.1	100.0	90.1	100.0	90.1	100.0	90.1	100.0	89.0	100.0	90.1	100.0
100:200:400:500	4:3:2:1			89.1	100.0	60.5	100.0	89.5	100.0	89.4	100.0	93.3	100.0	93.4	100.0	93.3	100.0	93.3	100.0	92.9	100.0	93.3	100.0
	1:2:3:4	**83.1**	**100.0**	99.7	100.0	95.0	100.0	99.7	100.0	99.7	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0
	1:1:1:1	87.0	100.0	86.6	100.0	86.1	100.0	86.7	100.0	86.9	100.0	89.9	100.0	89.9	100.0	89.9	100.0	89.9	100.0	89.2	100.0	89.8	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .10. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

Elliptically Symmetric Nonnormal Distribution (0,3)

K=2

<u>*α= .01*</u>. At the "negative conditions" when sample size was moderate, the RMM test statistics T_{ADF}, T_{YB1} and T_{YB2} yielded the highest empirical power estimates at both *d*=.2 and *d*=.8. At these conditions, the rest of the test statistics provided power estimates varying less than 1%. When sample size became large for the "negative conditions", the ANOVA alternatives generally yielded the higher power estimates when *d*=.2 and the power estimates were all 100% when *d*=.8 at the "negative conditions".

At the "positive conditions" when sample size was small, the RMM test statistics generally delivered power estimates less than 8% higher than those from ANOVA alternatives when *d*=.2 and the variance heterogeneity ratio was moderate, and about 10-30% higher when *d*=.2 and the variance heterogeneity ratio was large. At the "positive conditions" when sample size was moderate, the ANOVA alternatives generally delivered power estimates approximately 5%-9% higher than those from ANOVA alternatives when *d*=.2 and the variance heterogeneity ratio was moderate, and became approximately the same (above 99%) when the variance heterogeneity ratio increased to large. The power estimates from the RMM approaches and the ANOVA alternatives were all 100% when sample size was large or *d* =.8 with moderate sample sizes.

When sample sizes were equal and small, the test statistic T_{ADF} yielded the best empirical power estimates when variances were homogenous; while the variance heterogeneity ratio became moderate and large, the ANOVA alternatives provided slightly higher power estimates when *d*=.2, but lower (2%-14%) power estimates when *d* =.8. When sample sizes increased, the Welch v_w and Brown and Forsythe *F** yielded the best power estimates when *d*=.2 and the power estimates approached 100% when *d*=.8.

Table 40: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=2 at α =.01

n_1: n_2	σ_1/ σ_2	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
4:16	1	0.8	3.5																				
	2.5																						
	4																						
16:4	2.5	**0.5**	**56.5**	**2.0**	**36.9**	**2.0**	**36.9**	**3.1**	**59.7**	**2.7**	**42.6**	4.6	85.2	9.7	97.2	4.6	85.2	3.6	92.5	6.3	95.1	3.7	81.2
	4	**2.9**	**95.9**	**11.6**	**73.0**	**11.6**	**73.0**	**14.5**	**95.7**	**14.4**	**85.1**	28.2	99.5	42.3	100.0	27.7	99.5	24.1	99.8	32.6	99.9	24.5	99.5
10:10	1	2.1	18.2	1.1	15.1	1.1	15.1	1.1	15.3	1.6	17.3	1.3	14.9	2.5	23.0	1.3	14.9	0.8	11.1	1.4	15.6	0.9	13.1
	2.5			4.1	56.8	4.1	56.8	4.3	60.5	4.6	61.9	2.9	64.5			2.9	64.3	2.0	61.6	3.5	70.7	2.2	62.0
	4			7.7	88.1	7.7	88.1	8.7	89.4	9.2	90.9	7.3	93.4			7.3	93.4	5.7	93.2	9.7	95.8	5.7	92.3
20:80	1	3.9	80.1	4.3	72.8	4.3	72.8	4.4	73.0	4.3	72.6	3.4	67.4	4.9	72.0	3.4	67.4	3.9	69.8	3.9	70.1	3.3	66.7
	2.5			3.3	58.0	3.3	58.0	3.3	58.4	3.3	58.2	2.8	54.7	4.3	62.9	2.8	54.7	3.8	59.4	3.8	59.5	2.6	54.4
	4			4.8	77.3	4.8	77.3	5.2	78.0	5.2	78.6	4.3	76.7	5.9	82.3	4.3	76.7	4.7	80.2	4.7	80.5	4.2	76.5
80:20	2.5	**16.6**	**100.0**	68.6	100.0	68.6	100.0	69.0	100.0	69.5	100.0	61.8	100.0	63.6	100.0	61.8	100.0	60.7	100.0	60.9	100.0	61.3	100.0
	4	**78.4**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.3	100.0	99.4	100.0	99.2	100.0	99.2	100.0	99.3	100.0	99.3	100.0
50:50	1	6.7	94.7	19.8	94.8	19.8	94.8	7.2	94.9	7.5	95.1	6.0	90.1	6.7	90.8	6.0	90.1	5.3	88.3	5.3	88.5	5.6	89.7
	2.5	31.0	100.0	54.8	100.0	54.8	100.0	30.8	100.0	31.0	100.0	27.8	100.0	29.9	100.0	27.8	100.0	27.1	100.0	27.3	100.0	27.2	100.0
	4	70.0	100.0	88.1	100.0	88.1	100.0	69.7	100.0	69.7	100.0	62.1	100.0	64.5	100.0	62.1	100.0	61.7	100.0	62.0	100.0	61.4	100.0
100:400	1	27.1	100.0	50.6	100.0	50.6	100.0	27.2	100.0	27.2	100.0	22.7	100.0	23.5	100.0	22.7	100.0	23.0	100.0	22.5	100.0	22.6	100.0
	2.5			43.7	100.0	43.7	100.0	21.6	100.0	21.6	100.0	17.8	100.0	18.6	100.0	17.8	100.0	18.3	100.0	17.8	100.0	17.8	100.0
	4			59.5	100.0	59.5	100.0	34.2	100.0	34.4	100.0	29.7	100.0	30.8	100.0	29.7	100.0	30.3	100.0	29.7	100.0	29.3	100.0
400:100	2.5	**99.8**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	4	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
250:250	1	46.6	100.0	69.7	100.0	69.7	100.0	47.3	100.0	47.3	100.0	36.9	100.0	37.2	100.0	36.9	100.0	36.4	100.0	35.7	100.0	36.7	100.0
	2.5	97.2	100.0	99.7	100.0	99.7	100.0	97.4	100.0	97.4	100.0	93.2	100.0	93.3	100.0	93.2	100.0	93.1	100.0	92.7	100.0	93.1	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.

Power was computed across 1000 replications at the nominal level of .01. d stands for effect size when k=2. d_1=.2 and d_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

$\underline{\alpha = .05}$. At the "negative conditions" when sample size was small, the RMM test statistics generally delivered power estimates at most 6% higher than those from ANOVA alternatives when d=.2 and about 9-33% higher when d=.8, among which the test statistic T_{YB1} provided the best power estimates. When sample size increased to be moderate, the empirical power estimates from the ANOVA alternatives and the RMM test statistics were close to each other. When sample size became large at the "negative conditions", the ANOVA alternatives generally yielded the power estimates about 5% higher for the moderate variance heterogeneity ratio and about 7% higher for the large variance heterogeneity ratio.

At the "positive conditions" when sample size was small, the RMM test statistics generally delivered power estimates approximately 6%-18% higher than those from ANOVA alternatives when d=.2 and the variance heterogeneity ratio was moderate, and about 18-33% higher when d=.2 and the variance heterogeneity ratio was large. Also, at the "positive conditions" when sample size was small, the RMM test statistics generally delivered power estimates approximately 17%-44% higher than those from ANOVA alternatives when d=.8 with moderately heterogeneous variances, and less than 15% higher when d=.8 with greatly heterogeneous variances. When sample size increased to be moderate with moderately heterogeneous variances at the "positive conditions", the empirical power estimates from the ANOVA alternatives were about 7-8% higher than those from and the RMM test statistics with d=.2. When sample size became large or d=.8 with moderate sample sizes, the empirical power estimates were all 100%.

When sample sizes were equal and small, the empirical power estimates were close while the test statistic T_{ADF} yielded the highest values. When sample sizes were equal and moderate, the ANOVA alternatives provided power estimates approximately 6% higher when d=.2. When sample size became large or d=.8 with moderate sample sizes, the empirical power estimates were all high with values above 98%.

Table 41: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=2 at α =.05

n_1: n_2	σ_1/ σ_2	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
4:16	1	3.5	11.6	3.3	5.8	3.3	5.8	3.7	8.7	3.4	6.3	5.0	19.7	7.6	33.9	5.0	19.7	6.2	27.0	6.3	29.5	4.5	17.5
	2.5			3.9	4.5	3.9	4.5	5.1	6.4	4.6	5.9	5.9	17.3			5.9	17.3	8.2	26.0			5.7	15.6
	4			3.6	4.1	3.6	4.1	5.2	6.9	4.6	6.3	6.8	23.1			6.8	23.1	10.6	36.6			6.7	20.4
16:4	2.5	**5.6**	**90.7**	**8.5**	**55.8**	**8.5**	**55.8**	**10.3**	**79.8**	**9.3**	**66.9**	18.3	98.0	25.3	99.0	18.3	98.0	18.2	98.2	20.3	98.7	16.8	97.1
	4	**14.3**	**100.0**	**31.4**	**85.1**	**31.4**	**85.1**	**37.4**	**99.1**	34.5	97.9	57.9	100.0	64.3	100.0	57.1	100.0	56.1	100.0	59.4	100.0	55.4	99.9
10:10	1	7.8	43.3	7.0	40.9	7.0	40.9	6.4	40.2	7.3	42.3	6.1	39.8	7.0	43.4	6.1	39.8	5.3	36.1	5.9	40.1	5.5	36.9
	2.5	15.3	91.6	13.0	86.8	13.0	86.8	13.1	87.2	13.8	88.2	12.1	87.3	16.2	91.2	12.1	87.1	10.9	86.5	12.7	88.2	10.5	85.4
	4	28.6	99.7	22.0	99.3	22.0	99.3	23.1	99.5	23.6	99.7	24.4	98.7	29.4	99.1	24.4	98.7	24.1	98.7	25.6	98.8	22.3	98.4
20:80	1	13.7	94.7	14.0	91.6	14.0	91.6	13.9	91.6	14.0	91.6	13.6	86.5	14.8	87.9	13.6	86.5	13.9	87.6	13.9	87.5	13.3	86.4
	2.5			12.2	82.4	12.2	82.4	12.6	82.5	12.5	82.5	10.9	80.2	12.8	82.0	10.9	80.2	11.9	81.1	11.9	81.0	10.5	79.4
	4			15.3	94.3	15.3	94.3	15.3	94.4	15.4	94.4	14.8	91.5	17.2	92.9	14.8	91.5	16.4	92.4	16.4	92.3	14.6	91.5
80:20	2.5	**52.5**	**100.0**	88.6	100.0	88.6	100.0	88.3	100.0	88.7	100.0	80.2	100.0	81.2	100.0	80.2	100.0	80.1	100.0	80.0	100.0	80.1	100.0
	4	**96.9**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
50:50	1	19.8	99.1	19.8	99.1	19.8	99.1	19.6	99.1	19.8	99.1	18.3	97.5	19.0	97.6	18.3	97.5	17.8	97.4	17.7	97.4	18.0	97.5
	2.5	55.5	100.0	54.8	100.0	54.8	100.0	54.8	100.0	54.8	100.0	48.0	100.0	49.3	100.0	48.0	100.0	47.5	100.0	47.5	100.0	47.3	100.0
	4	88.9	100.0	88.1	100.0	88.1	100.0	88.1	100.0	88.2	100.0	81.2	100.0	82.6	100.0	81.2	100.0	81.1	100.0	81.1	100.0	81.0	100.0
100:400	1	50.4	100.0	50.6	100.0	50.6	100.0	50.9	100.0	50.9	100.0	44.5	100.0	45.2	100.0	44.5	100.0	44.6	100.0	44.3	100.0	44.5	100.0
	2.5			43.7	100.0	43.7	100.0	43.7	100.0	43.7	100.0	38.0	100.0	38.8	100.0	38.0	100.0	38.6	100.0	37.9	100.0	37.9	100.0
	4			59.5	100.0	59.5	100.0	59.5	100.0	59.5	100.0	52.6	100.0	53.7	100.0	52.6	100.0	53.2	100.0	52.5	100.0	52.5	100.0
400:100	2.5	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	4	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
250:250	1	69.6	100.0	69.7	100.0	69.7	100.0	70.1	100.0	70.2	100.0	62.0	100.0	62.1	100.0	62.0	100.0	61.9	100.0	61.0	100.0	61.9	100.0
	2.5	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	98.3	100.0	98.3	100.0	98.3	100.0	98.3	100.0	98.2	100.0	98.3	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

α= .10. At the "negative conditions" when sample size was small, the RMM test statistics generally delivered power estimates 3-13% higher than those from the ANOVA alternatives when *d*=.2 and about 19-49% higher when *d*=.8. In this situation, the test statistics T_{ADF}, T_{YB2}, and T_{YB1}, among all the methods, yielded the best power estimates. When sample size increased to be moderate, the empirical power estimates from the ANOVA alternatives and the RMM test statistics were closer to each other. The test statistic T_{ADF} yielded the greatest values of power estimates when *d*=.2, but the ANOVA alternatives generally yielded power estimates about 3% higher when *d*=.8. When sample size became large at the "negative conditions", the ANOVA alternatives again provided power estimates about 4-6% higher when *d*=.2; when *d*=.8, the power estimates were all 100%.

At the "positive conditions" when sample size was small, the RMM test statistics generally delivered power estimates approximately 11%-27% higher than those from ANOVA alternatives when *d*=.2 and 2%-36% higher when *d*=.8. When sample size increased to be moderate and variance heterogeneity was moderate, the empirical power estimates from the ANOVA alternatives were about 7% higher than those from and the RMM test statistics with *d*=.2. When sample size became large or *d*=.8 with moderate sample sizes, the empirical power estimates were all 100%.

When sample sizes were equal and small, the empirical power estimates were close with the test statistic T_{ADF} yielding the highest values. When sample sizes were equal and moderate, the ANOVA alternatives provided power estimates approximately 8% higher with moderate variance heterogeneity and about 4% higher with large variance heterogeneity when *d* =.2; when *d*= .8, all power estimates became 100%. When sample size became large, the empirical power estimates were all high with values above 99.4%.

Table 42: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=2 at α =.10

n_1: n_2	σ_1/ σ_2	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
4:16	1	6.7	21.4	**4.3**	**9.1**	**4.3**	**9.1**	5.0	12.7	**4.9**	**11.0**	8.2	34.4	12.2	46.4	8.2	34.4	9.9	40.9	10.9	42.7	7.7	31.8
	2.5			**4.6**	**5.2**	**4.6**	**5.2**	6.1	8.4	5.6	7.8	10.3	28.8	14.5	42.3	10.3	28.8	13.3	37.3	13.8	40.0	9.3	26.7
	4			**4.2**	**5.6**	**4.2**	**5.6**	5.8	9.2	5.6	8.9	11.4	39.3	16.7	54.2	11.4	39.3	15.0	49.2	15.1	50.9	10.8	36.4
16:4	2.5	**12.1**	**97.2**	**14.3**	**63.2**	**14.3**	**63.2**	**18.6**	**90.2**	**16.7**	**84.6**	31.4	99.0	36.0	99.2	31.4	99.0	31.8	99.2	33.3	99.2	29.6	99.0
	4	**27.8**	**100.0**	**48.6**	**94.8**	**48.6**	**94.8**	55.7	99.9	53.9	99.8	71.6	100.0	75.8	100.0	70.8	100.0	70.7	100.0	72.1	100.0	69.3	100.0
10:10	1	13.8	58.1	12.6	56.1	12.6	56.1	12.0	55.6	13.5	57.0	10.3	53.2	13.5	55.0	10.3	53.2	9.3	51.2	10.3	52.9	9.3	50.7
	2.5	24.5	95.5	21.8	94.3	21.8	94.3	21.9	94.7	22.6	94.8	21.4	93.1	24.5	94.3	21.4	92.9	20.4	93.1	21.7	93.3	20.0	92.9
	4	41.3	99.9	36.5	99.9	36.5	99.9	37.1	99.9	37.7	99.9	35.6	99.3	40.2	99.4	35.6	99.3	35.0	99.3	37.4	99.3	33.3	99.3
20:80	1	22.7	96.6	21.9	96.1	21.9	96.1	21.8	96.1	21.8	96.1	21.2	92.8	22.4	93.2	21.2	92.8	22.2	93.0	22.1	93.0	21.2	92.7
	2.5			19.8	90.6	19.8	90.6	20.3	90.7	20.3	90.7	19.8	87.4	21.4	88.3	19.8	87.4	20.8	87.9	20.7	87.9	19.6	87.2
	4			25.5	97.9	25.5	97.9	25.5	98.0	25.5	97.9	23.9	95.0	26.3	95.6	23.9	95.0	25.4	95.3	25.1	95.3	23.7	95.0
80:20	2.5	**70.2**	**100.0**	9.4	100.0	94.2	100.0	94.3	100.0	94.3	100.0	86.9	100.0	87.0	100.0	86.9	100.0	86.8	100.0	86.7	100.0	86.8	100.0
	4	**99.9**	**100.0**	100.0	100.0	10.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
50:50	1	29.6	99.7	29.6	99.7	29.6	99.7	29.9	99.7	30.1	99.7	29.1	98.6	29.5	98.6	29.1	98.6	28.6	98.6	28.5	98.6	28.6	98.6
	2.5	68.2	100.0	67.9	100.0	67.9	100.0	67.9	100.0	68.1	100.0	60.4	100.0	60.9	100.0	60.4	100.0	60.3	100.0	60.1	100.0	60.1	100.0
	4	93.3	100.0	93.3	100.0	93.3	100.0	93.3	100.0	93.3	100.0	89.1	100.0	89.4	100.0	89.1	100.0	89.1	100.0	89.1	100.0	88.9	100.0
100:400	1	64.5	100.0	64.3	100.0	64.3	100.0	64.5	100.0	64.5	100.0	56.5	100.0	56.6	100.0	56.5	100.0	56.6	100.0	56.1	100.0	56.5	100.0
	2.5			56.0	100.0	56.0	100.0	56.3	100.0	56.0	100.0	50.0	100.0	50.5	100.0	50.0	100.0	50.2	100.0	50.0	100.0	50.0	100.0
	4			70.2	100.0	70.2	100.0	70.2	100.0	70.2	100.0	65.5	100.0	65.9	100.0	65.5	100.0	65.8	100.0	65.5	100.0	65.5	100.0
400:100	2.5	**100.0**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	4	**100.0**	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
250:250	1	79.7	100.0	79.7	100.0	79.7	100.0	79.9	100.0	79.9	100.0	72.3	100.0	72.5	100.0	72.3	100.0	72.3	100.0	71.9	100.0	72.3	100.0
	2.5	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.4	100.0	99.4	100.0	99.4	100.0	99.4	100.0	99.4	100.0	99.4	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .10. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

K=3

<u>α= .01</u>. At the "negative conditions" when sample size was small, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other falling between 1%-5% when f=.1; when f increased to .4, the RMM test statistics provided empirical power estimates between 60.1%-67.9%, while the ANOVA alternatives yielded empirical power estimates falling between 51.1-62.5% with the highest values from the T_{ADF} , T_{YB1} and T_{YB2} test statistics. At the "negative conditions" when sample size became moderate, the empirical power estimates from the RMM test statistics and the ANOVA alternatives except the BF F^* statistic were again close to each other for both small and large effect sizes falling between 2.2%-6.3% and at around 95% respectively. At this situation, the BF F^* statistic yielded a low power estimate value of 70.9% for f=.4. When sample size increased to large with f=.1, the ANOVA alternatives except the BF F^* statistic generally delivered power estimates approximately 6% higher than those from the RMM test statistics from when f=.1; when the effect size increased to .4, the empirical power estimates became 100% across the test statistics.

As usual, the power estimates at the "positive conditions" were comparatively higher than those at the "negative conditions". At the "positive conditions", the empirical power estimates were fairly close to each other across the ANOVA alternatives and RMM test statistics and sample sizes, which were about 8-11% and above 99.5%, 25% and 100%, as well as 98% and 100% for f=.1 and .4 respectively across small, moderate and large sample sizes. The ANOVA F test generally provided much lower empirical power estimates when the effect size was small or sample size was small at the "positive conditions".

When sample sizes were equal, the ANOVA alternatives including the Welch v_w, the Alexander and Govern A and the James second-order U tended to provide slightly higher power estimates than those from the RMM test statistics across sample sizes, especially for

Table 43: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=3 at α =.01

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4			6.2	93.5			6.1	94.2	6.1	93.5	5.2	89.7			5.2	89.7	4.6	88.7	5.0	89.3	4.8	89.4
	1:1:1	3.9	85.2	3.5	83.4	3.8	85.2	3.6	83.5			3.6	77.0	4.8	81.0	3.6	77.0	3.3	75.9	3.5	77.0	3.5	76.6
12:30:48	4:2.5:1			2.3	52.9			3.0	62.5	1.0	51.1	2.6	60.6	5.0	67.9	2.6	60.6	3.2	62.8	4.0	63.8	2.3	60.1
	1:2.5:4	**2.6**	**85.6**	10.9	99.9			11.1	99.9	10.9	99.9	8.6	99.6	11.8	99.8	8.6	99.5	8.3	99.5	8.8	99.6	8.5	99.5
	1:1:1	3.1	80.3	3.5	72.4	3.6	73.7	4.4	75.9	3.3	70.4	3.2	75.9	5.6	82.6	3.2	75.9	4.0	78.2	4.0	78.7	3.0	75.0
60:60:60	1:2.5:4			12.4	100.0	8.2	99.7	13.2	100.0	13.1	100.0	11.0	99.9	12.5	99.9	11.0	99.9	11.1	99.9	11.1	99.9	10.9	99.9
	1:1:1	9.8	99.3	9.5	99.4	9.8	99.3	9.6	99.4	10.0	99.4	7.9	98.9	9.3	99.2	7.9	98.9	7.8	98.8	8.0	98.9	7.9	98.9
24:60:96	4:2.5:1			3.8	95.1	2.2	70.9	4.8	95.7	3.8	95.1	4.7	94.0	6.3	95.5	4.7	94.0	5.2	94.6	5.2	94.7	4.4	93.9
	1:2.5:4	**4.9**	**100.0**	26.4	100.0	17.8	100.0	26.3	100.0	26.6	100.0	25.2	100.0	27.4	100.0	25.2	100.0	24.7	100.0	25.4	100.0	25.0	100.0
	1:1:1	6.1	99.6	6.1	99.3	5.7	99.3	7.2	99.3	6.3	99.3	7.6	98.2	8.9	98.5	7.6	98.2	8.0	98.4	8.4	98.4	7.5	98.2
300:300:300	1:2.5:4	65.5	100.0	82.2	100.0	65.5	100.0	82.7	100.0	82.7	100.0	71.6	100.0	72.2	100.0	71.6	100.0	71.6	100.0	72.1	100.0	71.6	100.0
	1:1:1	65.7	100.0	66.4	100.0	65.7	100.0	68.0	100.0	68.1	100.0	55.1	100.0	55.7	100.0	55.1	100.0	54.9	100.0	55.6	100.0	54.9	100.0
120:300:480	4:2.5:1			46.5	100.0	27.3	100.0	47.7	100.0	46.7	100.0	40.1	100.0	40.6	100.0	40.1	100.0	40.2	100.0	40.5	100.0	40.0	100.0
	1:2.5:4	**62.7**	**100.0**	99.1	100.0			99.1	100.0	99.1	100.0	97.7	100.0	97.8	100.0	97.7	100.0	97.6	100.0	97.8	100.0	97.7	100.0
	1:1:1	64.6	100.0	64.1	100.0	64.6	100.0	64.8	100.0	64.8	100.0	55.6	100.0	57.5	100.0	55.6	100.0	55.9	100.0	57.4	100.0	55.6	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .01. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

small effect sizes. The differences were about 0-2% for small and moderate sample sizes, and increased to approximately

α= .05. At the "negative conditions" when sample size was small, the RMM test statistics generally delivered power estimates close to those yielded by the Welch v_w, the Alexander and Govern *A* and the James second-order *U* across effect sizes. When sample size became moderate and *f*=.1, the RMM test statistics generally delivered slightly higher (about 2%) power estimates than those yielded by the Welch v_w, the Alexander and Govern *A* and the James second-order *U* statistics. On the contrary, when sample size increased to large with *f*=.1, the power estimates yielded by the Welch v_w, the Alexander and Govern *A* and the James second-order *U* statistics were about 5% higher than those provided by the RMM test statistics. When the effect size was large with moderate and large sample sizes, the empirical power estimates from the three ANOVA alternatives and all RMM test statistics were all high, above 98%.

At the "positive conditions" when sample size was small with *f*=.1, the Welch v_w, the Alexander and Govern *A*, the James second-order *U* statistics and the test statistic T_{ADF} yielded power estimates approximately 3% higher than those from the rest of the four RMM test statistics, falling at around 27%. When sample sizes increased with *f*=.1, all test statistics except the BF *F** provided power estimates at around 50% and 99.7% for moderate and large sample sizes respectively. When the effect size became large, the power estimates were all very close to or at 100%.

When sample sizes were equal with *f*=.1, the ANOVA alternatives including the Welch v_w, Alexander and Govern *A*, James second-order *U* and the test statistic T_{ADF} provided slightly higher (less than 4%) power estimates than those from the rest of the RMM test

Table 44: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=3 at α =.05

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4			18.7	98.7	13.8	94.5	18.9	98.7	18.7	98.7	15.1	97.4	17.3	97.7	15.1	97.4	15.0	97.3	15.1	97.4	14.4	97.4
	1:1:1	14.2	94.4	13.9	94.0	14.3	94.4	13.9	94.0	14.5	94.1	12.1	91.2	13.7	92.5	12.1	91.2	11.4	90.8	11.8	90.9	11.9	90.9
12:30:48	4:2.5:1			9.1	80.3	7.3	57.3	10.1	84.0	9.0	80.1	11.1	81.4	13.8	84.9	11.1	81.4	11.9	82.8	12.1	83.0	10.5	80.9
	1:2.5:4	**7.8**	**97.7**	27.6	100.0			27.4	100.0	27.8	100.0	24.1	99.9	27.0	99.9	24.1	99.8	23.5	99.9	23.9	99.9	23.6	99.9
	1:1:1	12.4	94.3	12.2	91.3	12.2	92.2	12.9	92.0	11.9	90.9	11.5	92.1	14.1	94.0	11.5	92.1	12.1	92.9	12.4	93.1	11.3	91.9
60:60:60	1:2.5:4	23.9	100.0	30.3	100.0	24.1	100.0	30.9	100.0	30.8	100.0	27.6	99.9	29.2	99.9	27.6	99.9	27.4	99.9	27.7	99.9	27.4	99.9
	1:1:1	23.5	99.9	24.4	99.9	23.7	99.9	24.5	99.9	24.6	99.9	23.5	99.7	24.3	99.7	23.5	99.7	22.9	99.7	23.2	99.7	23.1	99.7
24:60:96	4:2.5:1			14.3	99.0	8.1	91.3	14.6	99.4	14.0	99.0	16.9	98.2	18.8	98.4	16.9	98.2	17.6	98.2	17.7	98.2	16.7	98.2
	1:2.5:4	**15.8**	**100.0**	50.4	100.0	38.2	100.0	50.4	100.0	50.8	100.0	49.2	100.0	50.3	100.0	49.2	100.0	49.0	100.0	49.2	100.0	48.9	100.0
	1:1:1	20.1	99.8	18.6	99.8	19.0	99.8	19.8	99.8	18.7	99.8	20.9	99.6	23.8	99.7	20.9	99.6	21.9	99.7	22.0	99.7	20.5	99.6
300:300:300	1:2.5:4	85.5	100.0	94.9	100.0	85.5	100.0	95.2	100.0	95.2	100.0	87.9	100.0	88.0	100.0	87.9	100.0	87.8	100.0	87.4	100.0	87.9	100.0
	1:1:1	85.6	100.0	85.4	100.0	85.6	100.0	85.7	100.0	85.8	100.0	75.7	100.0	76.0	100.0	75.7	100.0	75.6	100.0	74.8	100.0	75.6	100.0
120:300:480	4:2.5:1			70.2	100.0	53.7	100.0	70.5	100.0	70.4	100.0	64.6	100.0	65.0	100.0	64.6	100.0	64.8	100.0	63.7	100.0	64.5	100.0
	1:2.5:4	**88.4**	**100.0**	99.9	100.0	97.8	100.0	99.9	100.0	99.9	100.0	99.7	100.0	99.8	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0
	1:1:1	82.6	100.0	83.3	100.0	83.0	100.0	83.5	100.0	83.7	100.0	79.0	100.0	79.3	100.0	79.0	100.0	79.1	100.0	78.1	100.0	78.8	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

statistics for both small and moderate sample sizes across variance ratios. However, when sample sizes became large, the Welch v_w, Alexander and Govern A, and James second-order U statistics yielded approximately about 8% higher power estimates than those from the RMM test statistics. When sample sizes were equal and small with f=.4, the power estimates from the four ANOVA alternatives RMM test statistics were similar. For the rest of the conditions with f=.4, the power estimates were all high, above 97%.

α= .10. At the "negative conditions" when sample sizes were large, the power estimates by the Welch v_w, the Alexander and Govern A and the James second-order U statistics were about 4% higher than those from the SMM test statistics. The BF F^* test again provided much lower power estimates across sample sizes with f=.1 or for small sample sizes with f=.4. Other than the BF F^* test statistic, the rest of the ANOVA alternatives and the RMM test statistics all yielded similar power estimates across all other conditions.

At the "positive conditions" when f=.1, the Welch v_w, the Alexander and Govern A, the James second-order U statistics and the test statistic T_{ADF} delivered power estimates approximately 1-4% higher than those from the rest of the four RMM test statistics for both small and moderate samples. When sample sizes increased to large with f=.1, all test statistics except the BF F^* statistic provided power estimates at above 99%. When the effect size became large at the "positive conditions", the power estimates were all very close to or at 100% across sample sizes.

When sample sizes were equal and variances were heterogeneous with f=.1, the ANOVA alternatives including the Welch v_w, Alexander and Govern A, and James second-order U provided slightly (0-4%) higher power estimates than those from the RMM test statistics across sample sizes. When sample sizes were equal and variances were homogeneous

Table 45: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=3 at α =.10

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4	23.0	97.6	28.1	99.4	22.4	97.2	27.7	99.4	27.8	99.4	23.8	98.8	24.7	98.8	23.8	98.8	23.6	98.8	23.6	98.8	23.6	98.8
	1:1:1	23.7	96.7	23.4	96.9	23.8	96.7	23.6	96.9	24.0	97.1	20.5	95.2	22.3	96.0	20.5	95.2	20.2	95.1	20.2	95.1	20.0	95.0
12:30:48	4:2.5:1			16.7	89.2	14.2	70.1	18.2	90.2	16.7	89.2	17.5	89.1	19.8	90.4	17.5	89.1	18.7	89.5	18.9	89.6	17.0	88.7
	1:2.5:4	**14.0**	**99.2**	40.5	100.0	31.4	100.0	40.4	100.0	40.7	100.0	36.4	99.9	38.0	99.9	36.4	99.8	36.1	99.9	36.2	99.9	35.6	99.9
	1:1:1	23.0	96.9	20.5	96.6	21.5	96.1	21.3	96.4	20.3	96.6	20.1	96.3	24.0	97.1	20.1	96.3	21.2	96.4	21.4	96.5	19.7	96.0
60:60:60	1:2.5:4	35.1	100.0	43.3	100.0	35.0	100.0	43.6	100.0	43.6	100.0	41.0	99.9	42.2	99.9	41.0	99.9	40.8	99.9	40.8	99.9	40.8	99.9
	1:1:1	35.0	100.0	35.9	100.0	35.2	100.0	36.0	100.0	36.2	100.0	35.2	100.0	36.0	100.0	35.2	100.0	35.1	100.0	35.1	100.0	34.8	100.0
24:60:96	4:2.5:1			25.1	99.8	14.3	96.6	25.9	99.8	24.9	99.8	26.5	99.2	27.4	99.2	26.5	99.2	26.6	99.2	26.7	99.2	26.1	99.2
	1:2.5:4	**25.7**	**100.0**	64.6	100.0	49.5	100.0	64.8	100.0	64.9	100.0	61.8	100.0	62.7	100.0	61.8	100.0	61.8	100.0	61.8	100.0	61.8	100.0
	1:1:1	32.2	99.8	29.7	99.8	30.4	99.8	30.3	99.8	30.0	99.8	33.7	100.0	35.3	100.0	33.7	100.0	34.3	99.9	34.4	99.9	33.1	99.9
300:300:300	1:2.5:4	92.5	100.0	97.6	100.0	92.5	100.0	97.6	100.0	97.6	100.0	93.5	100.0	93.6	100.0	93.5	100.0	93.4	100.0	93.0	100.0	93.4	100.0
	1:1:1	92.3	100.0	92.1	100.0	92.3	100.0	92.3	100.0	92.2	100.0	85.1	100.0	85.1	100.0	85.1	100.0	85.1	100.0	84.6	100.0	85.1	100.0
120:300:480	4:2.5:1			79.8	100.0	65.4	100.0	80.2	100.0	79.9	100.0	75.7	100.0	76.1	100.0	75.7	100.0	76.0	100.0	75.2	100.0	75.7	100.0
	1:2.5:4	**95.1**	**100.0**	100.0	100.0	99.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0
	1:1:1	91.5	100.0	91.6	100.0	91.4	100.0	91.6	100.0	91.6	100.0	85.2	100.0	85.2	100.0	85.2	100.0	85.1	100.0	84.6	100.0	85.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .10. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

with f=.1, the power estimates were similar for both small and large sample sizes across all methods; when sample size became large, the ANOVA methods provided power estimates about 7% higher than the estimates from the RMM methods. When sample sizes were equal with f=.4, the power estimates from the four ANOVA alternatives RMM test statistics were similar., falling above 95%.

K=4

<u>*α= .01*</u>. At the "negative conditions" with f=.1, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other, falling between 3.3%-5.3% and 9.3%-12.3% when sample sizes were small and moderate respectively; when sample sizes increased to large, the Welch v_w, Alexander and Govern A and James U provided power estimates about 10% higher than the estimates yielded by the RMM test statistics. When the effect size increased to be large, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other falling between 89.8%-96% when sample size was small; the power estimates were all 100% when sample sizes were moderate and large.

At the "positive conditions" with f=.1, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other, falling between 9.4%-11.8% and above 98.9% when sample sizes were small and large respectively; when sample sizes were moderate, the Welch v_w , Alexander and Govern A, James U and T_{ADF} provided power estimates about 5% higher than those from the rest of the RMM test statistics. When the effect size increased to large, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were above 99.5% across sample sizes.

Table 46: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=4 at α =.01

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4	6.5	92.9	7.5	99.7	5.5	91.2	7.6	99.8	7.7	99.7	7.3	98.9			7.3	98.9	6.7	98.6	6.3	98.6	7.0	98.8
	1:1:1:1	4.1	93.6	3.7	90.9	3.9	93.4	3.9	91.0	3.9	92.0	4.1	87.0	6.0	89.0	4.1	87.0	4.0	87.2	3.9	87.0	4.0	86.8
	4:3:2:1							4.1	96.0	3.3	89.8	4.9	92.4			4.9	92.4	5.3	92.6	5.3	92.4	4.6	92.3
10:20:40:50	1:2:3:4	**3.1**	**93.3**	11.8	99.9	**9.4**	**99.3**	11.5	99.9	11.8	99.9	10.9	99.5			10.9	99.5	11.7	99.5	11.7	99.5	10.6	99.5
	1:1:1:1	4.6	93.4	4.0	88.1	3.8	88.4	4.4	91.2	3.7	85.0	4.6	85.7			4.6	85.7	4.9	87.0	4.9	87.0	4.1	85.5
60:60:60:60	1:2:3:4			20.5	100.0			21.7	100.0	21.0	100.0	18.3	100.0			18.3	100.0	17.6	100.0	17.4	100.0	18.2	100.0
	1:1:1:1	10.7	99.9	9.9	99.9	10.8	99.9	10.0	99.9	10.2	99.9	10.8	99.5	11.9	99.5	10.8	99.5	10.6	99.5	10.4	99.5	10.6	99.5
	4:3:2:1			9.3	100.0			11.7	100.0	9.4	100.0	10.9	100.0	12.3	100.0	10.9	100.0	11.0	100.0	10.9	100.0	10.9	100.0
20:40:80:100	1:2:3:4	**5.6**	**100.0**	33.0	100.0			32.8	100.0	33.6	100.0	27.8	100.0	32.6	100.0	27.8	100.0	29.4	100.0	29.0	100.0	27.5	100.0
	1:1:1:1			9.4	100.0			9.8	100.0	9.3	100.0	9.4	99.9	11.5	99.9	9.4	99.9	10.4	99.9	10.4	99.9	9.3	99.9
300:300:300:300	1:2:3:4			95.4	100.0			96.3	100.0	96.2	100.0									88.8	100.0		
	1:1:1:1	71.2	100.0	70.7	100.0	71.2	100.0	72.2	100.0	72.3	100.0									61.0	100.0		
	4:3:2:1			83.7	100.0			84.6	100.0	84.4	100.0	74.1	100.0	74.8	100.0	74.1	100.0	74.3	100.0	72.0	100.0	74.0	100.0
100:200:400:500	1:2:3:4	**69.7**	**100.0**	99.9	100.0			100.0	100.0	100.0	100.0	99.0	100.0	99.0	100.0	99.0	100.0	99.0	100.0	98.4	100.0	98.9	100.0
	1:1:1:1	72.0	100.0	72.8	100.0	72.2	100.0	74.0	100.0	73.5	100.0	65.6	100.0	66.5	100.0	65.6	100.0	65.9	100.0	63.8	100.0	65.5	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .01. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

When sample sizes were equal, the ANOVA alternatives and the RMM test statistics provided similar power estimates. When the effect size was small and variances were heterogeneous, the empirical power estimates were between 5.5%-7.6%, 17.4%-21.7% and 88.8%-96.3% across sample sizes; when the effect size was large and variances were heterogeneous, most of the empirical power estimates were all above 98.6% across sample sizes. When the effect size was small and variances were homogeneous, the empirical power estimates were between 3.7%-6.0%, 9.9%-11.9% and 61%-72.3% across sample sizes; when the effect size was large and variances were homogeneous, the empirical power estimates were between 86.8%-93.6%, above 99.5% and 100% across sample sizes.

α= .05. At the "negative condition" with f=.1, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other, falling between 9.5%-18.1% for small sample sizes; when sample sizes became moderate, the Welch v_w, Alexander and Govern A, James U and T_{ADF} provided slightly higher (less than 5%) power estimates than the rest of the RMM test statistics; when sample sizes were large, the Welch v_w, Alexander and Govern A, and James U provided power estimates about 5% higher than the rest of the RMM test statistics. When the effect size increased to large, the empirical power estimates from the RMM test statistics and most of the ANOVA alternatives were above 97.8% across sample sizes.

At the "positive condition" with f=.1, the Welch v_w, Alexander and Govern A, James U and T_{ADF} provided power estimates about 2-5% higher than those from the rest of the RMM test statistics for small and moderate sample sizes; when sample sizes became large, the power estimates were all above 98.4%. When the effect size increased to large, the empirical power estimates were all close to 100% across sample sizes.

When sample sizes were equal and variances were heterogeneous, the empirical power estimates yielded by the Welch v_w, Alexander and Govern A, James U and T_{ADF} were at most 5% higher than the estimates from the rest of the RMM test statistics across sample sizes. When sample sizes were equal and variances were homogeneous with f=.1, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other, falling between 12.8%-16% and 23%-26.2%for small and moderate sample sizes respectively; when sample size became large with f=.1 and homogeneous variances, the ANOVA-based methods provided power estimates about 6% higher than the RMM test statistics. When the effect size increased with equal sample sizes, the power estimates were all above 95.1% across sample sizes and variance ratios.

α= .10. At the "negative condition" with f=.1, the Welch v_w, Alexander and Govern A, James U and T_{ADF} provided power estimates about 2-5% higher than those from the rest of the RMM test statistics across sample sizes. When the effect size increased to be large, the empirical power estimates were all close to 100% across sample sizes except that from the BF F^* statistic.

At the "positive condition" with f=.1, the Welch v_w, Alexander and Govern A, James U and T_{ADF} again provided power estimates about 2-6% higher than those from the rest of the RMM test statistics for small and moderate sample sizes; when sample sizes became large, the power estimates were all above 99%. When the effect size increased to large, the empirical power estimates were all close to 100% across sample sizes.

When sample sizes were equal and variances were heterogeneous, the empirical power estimates yielded by the Welch v_w, Alexander and Govern A, James U and T_{ADF} were as much as 5% higher than the estimates from the rest of the RMM test statistics for small and

Table 47: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=4 at α =.05

n_1: n_2: n_3: n_4	$\sigma_1/\sigma_2/\sigma_3/\sigma_4$	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4	15.7	98.3	23.1	100.0	14.8	97.8	24.1	100.0	23.5	100.0	19.2	99.6	22.2	99.6	19.2	99.6	19.6	99.6	19.2	99.6	18.9	99.6
	1:1:1:1	14.5	98.6	13.8	98.4	14.7	98.4	14.9	98.5	14.6	98.5	13.4	95.2	16.0	95.9	13.4	95.2	13.4	95.2	12.8	95.2	12.9	95.1
10:20:40:50	4:3:2:1			13.5	98.6	9.5	73.4	17.2	99.4	12.8	98.5	14.0	98.0	18.1	98.4	14.0	98.0	15.7	97.9	15.6	97.8	13.8	97.9
	1:2:3:4	**8.1**	**98.9**	30.3	100.0	19.9	99.9	29.4	100.0	30.2	100.0	26.3	100.0	30.8	100.0	26.3	100.0	27.8	100.0	27.5	100.0	26.1	100.0
	1:1:1:1	14.2	98.1	13.8	97.4	13.3	97.0	13.3	97.8	12.8	97.1	14.1	95.0	17.3	96.5	14.1	95.0	15.8	95.6	15.6	95.4	13.7	95.0
60:60:60:60	1:2:3:4	25.9	100.0	42.3	100.0	24.8	100.0	43.2	100.0	42.9	100.0	38.4	100.0	40.5	100.0	38.4	100.0	37.7	100.0	37.6	100.0	37.9	100.0
	1:1:1:1	26.1	100.0	25.6	100.0	26.2	100.0	26.1	100.0	25.9	100.0	23.0	99.9	24.3	99.9	23.0	99.9	23.1	99.9	23.0	99.9	23.0	99.9
20:40:80:100	4:3:2:1			28.1	100.0	14.5	98.7	31.1	100.0	28.3	100.0	26.5	100.0	28.6	100.0	26.5	100.0	26.8	100.0	26.5	100.0	26.2	100.0
	1:2:3:4	**15.9**	**100.0**	56.2	100.0	36.3	100.0	56.1	100.0	57.1	100.0	52.2	100.0	54.8	100.0	52.2	100.0	52.8	100.0	52.5	100.0	51.6	100.0
	1:1:1:1	24.9	100.0	24.3	100.0	24.1	100.0	25.0	100.0	24.4	100.0	22.6	100.0	25.0	100.0	22.6	100.0	23.3	100.0	23.2	100.0	22.4	100.0
300:300:300:300	1:2:3:4	87.7	100.0	98.3	100.0	87.8	100.0	98.5	100.0	98.5	100.0	97.4	100.0	97.5	100.0	97.4	100.0	97.4	100.0	97.2	100.0	97.4	100.0
	1:1:1:1	88.2	100.0	87.8	100.0	88.2	100.0	88.2	100.0	88.2	100.0	82.1	100.0	82.3	100.0	82.1	100.0	82.1	100.0	82.0	100.0	82.1	100.0
100:200:400:500	4:3:2:1			94.6	100.0	63.0	100.0	94.9	100.0	94.8	100.0	89.4	100.0			89.4	100.0	89.3	100.0	88.8	100.0	89.3	100.0
	1:2:3:4	**88.8**	**100.0**	100.0	100.0	98.4	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0
	1:1:1:1	88.7	100.0	89.2	100.0	88.7	100.0	89.9	100.0	89.6	100.0	83.0	100.0			83.0	100.0	83.2	100.0	82.6	100.0	83.0	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

Table 48: Empirical Power (%) for the Elliptically Symmetric Nonnormal Distribution with Skewness and Kurtosis of (0, 3), k=4 at α =.10

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	F		W		BF		A		U		T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB2}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4	22.9	99.4	34.4	100.0	21.8	99.4	35.0	100.0	35.5	100.0	30.7	99.9	32.9	99.9	30.7	99.9	30.8	99.9	30.2	99.9	30.1	99.9
	1:1:1:1	24.5	99.6	23.4	99.4	24.7	99.6	23.6	99.5	24.3	99.4	22.6	97.7	23.8	98.1	22.6	97.7	21.7	97.7	21.7	97.7	21.8	97.6
10:20:40:50	4:3:2:1			25.2	99.5	16.2	85.5	27.8	99.7	24.6	99.5	22.5	99.2	25.7	99.2	22.5	99.2	24.4	99.1	24.1	99.1	22.2	99.1
	1:2:3:4	**12.6**	**99.6**	43.6	100.0	28.6	100.0	41.9	100.0	43.8	100.0	37.4	100.0	41.8	100.0	37.4	100.0	38.8	100.0	38.6	100.0	36.9	100.0
	1:1:1:1	23.2	99.8	22.3	99.1	21.2	99.1	23.2	99.3	22.3	99.1	21.2	98.1	24.6	98.6	21.2	98.1	22.7	98.3	22.4	98.3	21.0	98.1
60:60:60:60	1:2:3:4	37.7	100.0	56.0	100.0	36.8	100.0	56.2	100.0	56.5	100.0	51.1	100.0	52.1	100.0	51.1	100.0	50.5	100.0	50.3	100.0	50.8	100.0
	1:1:1:1	39.1	100.0	37.8	100.0	39.1	100.0	38.0	100.0	38.8	100.0	35.2	99.9	36.3	99.9	35.2	99.9	35.3	99.9	35.0	99.9	35.0	99.9
20:40:80:100	4:3:2:1			40.4	100.0	22.7	99.8	42.3	100.0	41.1	100.0	36.9	100.0	38.5	100.0	36.9	100.0	37.1	100.0	37.0	100.0	36.7	100.0
	1:2:3:4	**24.9**	**100.0**	69.0	100.0	48.3	100.0	68.6	100.0	69.6	100.0	64.5	100.0	66.5	100.0	64.5	100.0	65.0	100.0	64.7	100.0	64.4	100.0
	1:1:1:1	37.7	100.0	36.1	100.0	37.0	100.0	37.0	100.0	36.3	100.0	34.0	100.0	35.7	100.0	34.0	100.0	35.0	100.0	34.8	100.0	33.8	100.0
300:300:300:300	1:2:3:4	92.9	100.0	99.1	100.0	92.9	100.0	99.1	100.0	99.1	100.0	98.8	100.0	98.8	100.0	98.8	100.0	98.8	100.0	98.6	100.0	98.8	100.0
	1:1:1:1	92.5	100.0	92.8	100.0	92.5	100.0	92.8	100.0	93.0	100.0	88.7	100.0	88.7	100.0	88.7	100.0	88.7	100.0	88.1	100.0	88.7	100.0
100:200:400:500	4:3:2:1			97.1	100.0	75.2	100.0	97.2	100.0	97.2	100.0	94.4	100.0	94.5	100.0	94.4	100.0	94.4	100.0	94.1	100.0	94.4	100.0
	1:2:3:4	**95.6**	**100.0**	100.0	100.0	99.3	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0
	1:1:1:1	94.9	100.0	94.9	100.0	94.9	100.0	95.2	100.0	95.2	100.0	89.6	100.0	89.8	100.0	89.6	100.0	89.7	100.0	89.2	100.0	89.4	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

moderate sample sizes; when equal sample sizes became large, the empirical power estimates from the ANOVA alternatives and RMM test statistics were close to each other, all above 98.5%, except for the BF F^* statistic. When sample sizes were equal and small with homogeneous variances and f=.1, the empirical power estimates from the RMM test statistics and the ANOVA alternatives were close to each other falling between 21.7%-24.5%; when sample size became moderate or large with f=.1, the ANOVA methods provided power estimates about 2-4% higher than the RMM test statistics. When the effect size increased with equal sample sizes, the power estimates were all above 97.6% across sample sizes and variance ratios.

Asymmetric Nonnormal Distribution (3,21)

K=2

$\underline{\alpha = .01}$. The empirical power estimates of the following five test statistics are summarized in Table 49---- the Brown and Forsythe F^*, the James second-order U, T_{YB1}, T_{YB2} and T_{BC}, based on the study of Type I error rates. At the "negative conditions" when sample sizes were small, both the BF F^* and the James U provided power estimates less than 1% for d=.2; however, the James U yielded power estimates about 6-12% higher than those from the BF F^* test statistic.

At the "positive conditions" when sample sizes were small, the BF F^* and the James U provided power estimates about 22-28% higher than the three RMM test statistics when d=.2 with moderate variance heterogeneity; however, when the effect size increased to .8, the three RMM test statistics yielded power estimates about 5% higher than the two ANOVA alternatives. When variance heterogeneity became large and sample sizes were small at the "positive conditions", the test statistic T_{YB2} provided the highest power estimates of 45.8% for d=.2 and 99.8% for d=.8. When sample sizes increased to moderate with moderate variance heterogeneity, the RMM test statistics yielded power estimates of about 65% and above 89.4% for *d*=.2 and .4 respectively. When sample sizes increased to moderate with large variance heterogeneity, the RMM test statistics yielded power estimates of about 99.2% and 100% for *d*=.2 and .4 respectively.

When sample sizes were equal and small and variances were moderately heterogeneous, the power estimates from the T_{YB} and T_{BC} yielded power estimates of about .4% and 75.2% as well as .3% and 70.7% for *d*=.2 and .4 respectively. When sample sizes became large and variances were heterogeneous, the power estimates from the RMM statistics provided power estimates above 96.8% for both *d*=.2 and .4. However, it was noted that the empirical power estimates from the two ANOVA alternatives tended to be higher than the estimates from the RMM statistics when sample sizes were equal and variances were homogeneous.

Table 49: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=2 at α=.01

n_1: n_2	σ_1/ σ_2	BF		U		T_{YB1}		T_{Yb2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
	1	1.0	14.8	1.1	20.6	0.4	12.7	0.6	17.0	0.0	5.3
4:16	2.5	0.7	4.4	0.7	10.7						
	4	0.4	2.3	0.6	13.8						
16:4	2.5	33.7	61.1	35.7	86.6	7.2	91.3	11.8	93.3	8.3	82.2
	4					35.5	99.8	45.8	99.8	34.9	98.3
	1	9.6	89.6	11.8	91.2	1.3	25.1	1.6	32.6	1.4	25.7
10:10	2.5					0.4	75.2			0.3	70.7
	4										
	1	52.1	100.0	51.9	100.0	1.7	82.1	1.7	82.3	1.1	76.7
20:80	2.5										
	4										
80:20	2.5					65.1	100.0	65.2	100.0	65.2	100.0
	4					99.2	100.0	99.2	100.0	99.2	100.0
	1	76.2	100.0	76.9	100.0	6.0	87.0	6.1	87.0	6.0	87.4
50:50	2.5										
	4										
	1	100.0	100.0	100.0	100.0						
100:400	2.5										
	4					22.5	100.0	21.4	100.0	21.0	100.0
400:100	2.5										
	4										
	1	100.0	100.0	100.0	100.0	38.4	100.0	37.4	100.0	38.6	100.0
250:250	2.5					97.0	100.0	96.8	100.0	97.0	100.0
	4					100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .01. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

α= .05. The empirical power estimates of all six RMM test statistics are summarized in Table 50. At the “negative conditions” with large sample sizes and *d*=.2, the empirical power estimates from the RMM test statistics were at about 34% and 51% for moderate and large variance heterogeneity respectively; when effect size increased to .8, the power estimates were all 100%.

At the “positive conditions” when sample sizes were small and variances were moderately heterogeneous, the RMM test statistics provided power estimates falling between 26.6-35% and 94.7-97.3% for *d*=.2 and .8 respectively; when variance heterogeneity increased, the power estimates were around 66.8-75.9% and 99.9% with *d*=.2 and .8 respectively. When sample sizes were moderate at the “positive conditions”, the RMM test statistics provided power estimates of around 81% and 91.3% for *d*=.2 with moderately and greatly heterogeneous variances respectively. When sample sizes were moderate with *d*=.8 or large, the power estimates were close to 100% for both effect size conditions.

When sample sizes were equal and variances were homogeneous for *d*=.2, the power estimates were around 8%, 20% and 61% for small, moderate and large sample sizes respectively; when sample sizes were equal and variances were moderately heterogeneous for *d*=.2, the power estimates were around 8%, 49% and 99.7% for small, moderate and large sample sizes respectively; when variance heterogeneity increased to large with *d*=.2, the power estimates were around 88% and 100% for moderate and large sample sizes respectively. In addition, when the effect size increased to .8 with equal and small sample sizes, the power estimates were around 51% and 92% for homogeneous and moderately heterogeneous variances respectively; when sample sizes were equal and moderate, the power estimates were around 95% and 100% for homogeneous and heterogeneous variances; and the power estimates increased to 100% when sample sizes were large and equal.

Table 50: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=4 at α=.05

n_1: n_2	σ_1/σ_2	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{Yb2}		T_{BC}	
		d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
4:16	1	2.3	24.5	5.0	38.8	2.3	24.5	3.2	32.7	3.7	35.0	2.0	22.5
	2.5												
	4												
16:4	2.5	29.1	95.4	35.0	97.3	28.7	95.4	27.7	96.6	30.8	96.8	26.6	94.7
	4	70.0	99.9	75.9	99.9	68.4	99.9	69.4	99.9	72.2	99.9	66.8	99.9
10:10	1	8.5	51.2	10.1	54.3	8.5	51.2	7.6	49.1	8.5	51.6	7.6	48.7
	2.5	8.6	92.7			8.6	92.2	8.4	92.7	10.0	93.3	7.4	91.7
	4												
20:80	1	8.2	93.5	9.6	94.5	8.2	93.5	9.0	94.0	8.8	94.0	7.9	93.2
	2.5												
	4												
80:20	2.5	81.4	100.0	82.3	100.0	81.4	99.8	81.3	100.0	81.3	100.0	81.2	100.0
	4	99.7	100.0	99.8	100.0	99.4	100.0	99.7	100.0	99.7	100.0	99.7	100.0
50:50	1	20.4	95.1	21.2	95.2	20.4	95.1	20.0	95.0	20.0	95.0	20.0	95.1
	2.5	49.0	100.0	50.4	100.0	49.0	99.5	48.9	100.0	48.9	100.0	48.7	100.0
	4	88.7	100.0	89.6	100.0	88.7	99.8	88.5	100.0	88.5	100.0	88.1	100.0
100:400	1	43.1	100.0	43.5	100.0	43.1	100.0	43.2	100.0	42.7	100.0	43.0	100.0
	2.5									34.1	100.0	34.0	100.0
	4	50.7	100.0			50.7	100.0	51.6	100.0	50.8	100.0	50.6	100.0
400:100	2.5	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
250:250	1	61.1	100.0	61.3	100.0	61.1	100.0	60.9	100.0	60.6	100.0	60.9	100.0
	2.5	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Power was computed across 1000 replications at the nominal level of .05. d stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

<u>α= .10</u>. The empirical power estimates of the six RMM test statistics are summarized in Table 51. At the "negative conditions" effect size of .2, the empirical power estimates for moderate and large variance heterogeneity ratios respectively were about 12% and 16%, as well as 48% and 67% for moderate and large sample sizes respectively. At the "positive

conditions" with effect size of .8, the empirical power estimates for moderate and large variance heterogeneity ratios respectively were about 97% and 99.9%, and 100% for moderate and large sample sizes respectively.

At the "positive conditions" with effect size of .2, the empirical power estimates for moderate and large variance heterogeneity ratios were about 42% and 80%, 87% and 99.9%, as well as 98.8% and 100% for small, moderate and large sample sizes respectively. At the "positive conditions" with effect size of .8, the empirical power estimates for moderate and large variance heterogeneity ratios were about 97% and 99.9%, 99% and 100%, as well as 100% for small, moderate and large sample sizes respectively.

When sample sizes were equal and variances were homogeneous for d=.2, the power estimates were around 15%, 29% and 70% for small, moderate and large sample sizes respectively; when sample sizes were equal and variances were moderately heterogeneous for d=.2, the power estimates were around 17%, 63% and 99.7% for small, moderate and large sample sizes respectively; when variance heterogeneity increased to large with d=.2, the power estimates were around 37%, 94% and 100% for small, moderate and large sample sizes respectively. In addition, when the effect size increased to .8, the power estimates were around 61%, 96% and 99.8% when variances were homogeneous, moderately heterogeneous and largely heterogeneous with small equal sample sizes; while the power estimates were around 97.5% and 100% when variances were homogeneous and heterogeneous with equal and moderate sample sizes; and the power estimates increased to 100% when sample sizes were large and equal.

Table 51: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=4 at α=.10

n_1: n_2	σ_1/ σ_2	T_{ML} d_1	T_{ML} d_2	T_{ADF} d_1	T_{ADF} d_2	T_{SB} d_1	T_{SB} d_2	T_{YB1} d_1	T_{YB1} d_2	T_{YB2} d_1	T_{YB2} d_2	T_{BC} d_1	T_{BC} d_2
4:16	1	5.8	40.4	9.9	51.2	5.8	40.4	7.6	46.8	8.2	48.4	4.9	37.9
	2.5											6.6	29.2
	4											6.6	47.5
16:4	2.5	42.5	97.3	46.8	97.9	42.1	97.3	41.5	97.7	43.4	97.8	39.7	96.7
	4	80.2	99.9	82.5	99.9	78.6	99.9	80.0	99.9	80.8	99.9	79.1	99.9
10:10	1	15.8	61.8	17.8	64.5	15.8	61.8	15.4	60.1	16.2	62.0	14.5	59.1
	2.5	17.4	96.2	21.7	97.0	17.4	95.7	17.3	96.0	18.4	96.5	16.1	95.5
	4	37.3	99.8			37.3	99.5	37.2	99.8			34.4	99.8
20:80	1	16.8	97.1	17.7	97.3	16.8	97.1	17.2	97.3	17.2	97.3	16.5	97.1
	2.5	12.1	95.7			12.1	95.7	13.0	96.2	12.9	96.1	11.7	95.7
	4	16.3	99.2			16.3	99.2	17.7	99.3	17.7	99.3	16.0	99.2
80:20	2.5	87.4	100.0	87.9	92.6	87.4	99.8	87.4	92.4	87.3	92.4	87.1	100.0
	4	99.9	100.0	99.9	100.0	99.6	100.0	99.9	100.0	99.9	100.0	99.9	100.0
50:50	1	28.8	97.5	29.3	97.5	28.8	97.5	28.7	97.5	28.7	97.5	28.7	97.5
	2.5	62.9	100.0	64.0	100.0	62.9	99.5	62.9	100.0	62.8	100.0	62.7	100.0
	4	94.4	100.0	94.8	100.0	94.4	99.8	94.6	100.0	94.6	100.0	94.4	100.0
100:400	1	56.5	100.0	56.8	100.0	56.5	100.0	56.6	100.0	56.2	100.0	56.5	100.0
	2.5	47.8	100.0	48.1	100.0	47.8	100.0	47.9	100.0	47.8	100.0	47.8	100.0
	4	67.1	100.0	67.6	100.0	67.1	100.0	67.5	100.0	67.2	100.0	67.1	100.0
400:100	2.5	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
250:250	1	70.4	100.0	70.4	100.0	70.4	100.0	70.4	100.0	70.3	100.0	70.4	100.0
	2.5	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0	99.7	100.0
	4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Power was computed across 1000 replications at the nominal level of .10. *d* stands for effect size when k=2. d_1=.2 and d_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

K=3

α= .01. At the “positive conditions” with effect size of .1, the empirical power estimates were about 18%, 35%, and 97% for small, moderate and large sample sizes respectively. At the “positive conditions” with effect size of .4, the empirical power estimates were about 93%, 99.9%, and 100% for small, moderate and large sample sizes respectively.

When sample sizes were equal and f=.1 with homogeneous variances, the power estimates were around 6%, 8% and 56% for small, moderate and large sample sizes respectively; while the effect size increased to .4, the power estimates were around 82%, 98% and 100% for small, moderate and large sample sizes respectively. When sample sizes were equal and large but variances were heterogeneous, the power estimates were around 72% and 100% for small and large effect sizes. It was also noted that the power estimates were around 2%, 5% and 55% across the three unequal sample sizes with homogeneous variances for f=.1, and about 76.2%, 99.2% and 100% for f=.4.

Table 52: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=3 at α=.01

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB}		T_{BC}	
		$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$
30:30:30	1:2.5:4												
	1:1:1	5.8	82.1	7.7	86.0	5.8	82.1	5.2	82.6	5.5	83.3	5.5	81.8
12:30:48	4:2.5:1												
	1:2.5:4	18.5	93.0	23.1	98.8	18.5	98.1	19.4	98.5	19.8	98.5	17.6	92.8
	1:1:1	1.8	91.5	2.7	82.6	1.8	81.8	2.0	76.2	2.0	77.3	1.7	91.3
60:60:60	1:2.5:4												
	1:1:1	7.8	98.4	9.2	98.8	7.8	98.4	7.6	98.5	7.8	98.5	7.6	98.4
24:60:96	4:2.5:1												
	1:2.5:4	35.1	100.0			35.1	99.9	35.7	100.0	36.2	100.0	34.8	100.0
	1:1:1	5.3	99.2			5.3	98.8	5.3	99.9	5.7	99.9	5.1	99.1
300:300:300	1:2.5:4	72.1	100.0	72.8	100.0	72.1	100.0	72.1	100.0	72.6	100.0	72.0	100.0
	1:1:1	55.9	100.0	56.5	100.0	55.9	100.0	55.7	100.0	56.5	100.0	55.8	100.0
120:300:480	4:2.5:1												
	1:2.5:4	96.8	100.0	96.9	100.0	96.8	100.0	96.8	100.0	96.8	100.0	96.8	100.0
	1:1:1	55.0	100.0			55.0	100.0					55.0	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

α= .05. At the "negative conditions" with large sample sizes, the empirical power estimates were about 63% and 100% for effect sizes of .1 and .4 respectively. At the "positive conditions" with effect size of .1, the empirical power estimates were about 37%, 55%, and 99.5% for small, moderate and large sample sizes respectively. At the "positive conditions" with effect size of .4, the empirical power estimates were above 97.7% across sample sizes.

When sample sizes were equal and *f*=.1 with homogeneous variances, the power estimates were around 15%, 23% and 77.6% for small, moderate and large sample sizes respectively, while the effect size increased to .4, the power estimates were around 93%, 99.7% and 100% for small, moderate and large sample sizes respectively. When variances were heterogeneous with equal and moderate sample sizes, the power estimates were around 40% and 86% for moderate and large effect sizes respectively; when sample sizes increased to large, the power estimates were around 99.5% and 100% for moderate and large effect sizes respectively. It was also noted that the power estimates were around 9%, 20% and 80% across the three unequal sample sizes with homogeneous variances for *f*=.1, and about 94%, 99.9% and 100% for *f*=.4.

Table 53: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=3 at α=.05

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4												
	1:1:1	15.4	93.4	16.5	94.1	15.4	93.4	15.2	93.0	15.3	93.1	15.0	93.3
	4:2.5:1												
12:30:48	1:2.5:4	36.4	97.7	39.7	99.8	36.4	99.4	36.8	99.8	36.8	99.8	35.5	97.6
	1:1:1	8.9	96.1	11.1	94.2	8.9	93.2	9.4	92.8	9.6	93.0	8.6	96.1
60:60:60	1:2.5:4	40.6	99.5	41.3	99.5	40.6	99.4	40.6	99.5	40.7	99.5	40.4	99.5
	1:1:1	23.5	99.7	24.9	99.7	23.5	99.7	23.1	99.7	23.4	99.7	23.4	99.7
	4:2.5:1												
24:60:96	1:2.5:4	54.6	100.0	55.5	100.0	54.6	99.9	54.4	100.0	54.6	100.0	54.3	100.0
	1:1:1	19.6	99.9	21.5	99.9	19.6	99.7	20.2	99.9	20.5	99.9	19.5	99.9
300:300:30	1:2.5:4	86.5	100.0	86.9	100.0	86.5	100.0	86.4	100.0	85.7	100.0	86.4	100.0
	1:1:1	77.8	100.0	77.8	100.0	77.8	100.0	77.7	100.0	76.6	100.0	77.7	100.0
	4:2.5:1	63.4	100.0	63.8	99.8	63.4	100.0	63.5	99.8	62.7	99.8	63.4	100.0
120:300:480	1:2.5:4	99.5	100.0	99.5	100.0	99.5	100.0	99.5	100.0	99.3	100.0	99.5	100.0
	1:1:1	80.0	100.0	80.5	100.0	80.0	100.0	80.2	100.0	79.2	100.0	79.8	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

α= .10. At the "negative conditions" with moderate sample sizes, the empirical power estimates were about 17% and 100% for effect sizes of .1 and .4 respectively; when sample sizes became large, the empirical power estimates were about 76% and 100% for effect sizes of .1 and .4 respectively

At the "positive conditions" with effect size of .1, the empirical power estimates were about 47%, 65%, and 99.8% for small, moderate and large sample sizes respectively. At the "positive conditions" with effect size of .4, the empirical power estimates were above 98.5% across sample sizes.

When sample sizes were equal and f=.1 with homogeneous variances, the power estimates were around 24%, 35% and 87% for small, moderate and large sample sizes respectively; while the effect size increased to .4, the power estimates were around 96%, 99.8% and 100% for small, moderate and large sample sizes respectively. When variances were heterogeneous with equal and small sample sizes, the power estimates were around 40% and 97% for small and large effect sizes respectively; when sample sizes increased to moderate, the power estimates were around 49% and 99.8% for small and large effect sizes respectively; when sample sizes became large, the power estimates were around 92% and 100% for small and large effect sizes respectively. It was also noted that the power estimates were around 17%, 30% and 87% across the three unequal sample sizes with homogeneous variances for f=.1, and 97%, 99.9% and 100% for f=.4.

Table 54: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=3 at α=.10

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30	1:2.5:4	40.4	99.9			40.4	97.3	40.1	97.3	40.1	97.4	39.7	99.9
	1:1:1	23.9	96.0	25.0	96.4	23.9	96.0	23.9	96.0	23.9	96.0	23.4	96.0
	4:2.5:1												
12:30:48	1:2.5:4	46.7	98.5	48.8	99.8	46.7	99.5	46.5	99.8	46.7	99.8	45.6	98.5
	1:1:1	16.6	97.6	19.4	96.8	16.6	97.1	17.0	96.4	17.3	96.4	15.6	97.6
60:60:60	1:2.5:4	48.8	99.8	49.5	99.8	48.8	99.7	48.9	99.8	48.9	99.8	48.7	99.8
	1:1:1	35.0	99.8	36.2	99.9	35.0	99.8	35.0	99.8	35.0	99.8	34.7	99.8
	4:2.5:1	17.1	100.0			17.1	100.0	17.9	100.0			17.1	100.0
24:60:96	1:2.5:4	65.1	100.0	66.1	100.0	65.1	99.9	64.9	100.0	64.9	100.0	64.7	100.0
	1:1:1	30.2	99.9	31.6	99.9	30.2	99.8	30.5	99.9	30.5	99.9	29.7	99.9
300:300:30	1:2.5:4	92.2	100.0	92.2	100.0	92.2	100.0	92.2	100.0	91.9	100.0	92.1	100.0
	1:1:1	87.0	100.0	87.1	100.0	87.0	100.0	87.0	100.0	86.4	100.0	87.0	100.0
	4:2.5:1	76.5	100.0	77.0	99.9	76.5	100.0	76.7	99.9	75.9	99.9	76.5	100.0
120:300:480	1:2.5:4	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0
	1:1:1	87.7	100.0	87.8	100.0	87.7	100.0	87.7	100.0	87.4	100.0	87.7	100.0

Note. F = ANOVA F. W = Welch v_w. BF = Brown and Forsythe F*. A = Alexander and Govern A. U = James second-order U. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Power was computed across 1000 replications at the nominal level of .10. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.

Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

K=4

<u>*α*= *.01*</u>. At the "positive conditions" with effect size of .1, the empirical power estimates were about 17% and 99% for small and large sample sizes respectively. At the "positive conditions" with effect size of .4, the empirical power estimates were above 99% for both small and large sample sizes.

When sample sizes were equal and f=.1 with homogeneous variances, the power estimates were around 11% and 66% for small and large sample sizes respectively; when the effect size increased to .4, the power estimates were all above 99% for small and large sample sizes respectively. When sample sizes were equal and large but variances were heterogeneous, the power estimates were at around 90% and 100% for small and large effect sizes respectively.

Table 55: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=4 at α =.01

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB}		T_{BC}	
		$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$
30:30:30:30	1:2:3:4												
	1:1:1:1												
	4:3:2:1												
10:20:40:50	1:2:3:4	17.5	99.2			17.5	99.1	20.5	99.8	20.4	99.8	17.2	99.0
	1:1:1:1											2.9	90.7
60:60:60:60	1:2:3:4												
	1:1:1:1	11.3	90.8	12.4	99.6	11.3	99.3	11.3	99.3	11.1	99.3	11.2	90.6
	4:3:2:1												
20:40:80:100	1:2:3:4												
	1:1:1:1												
300:300:300:300	1:2:3:4	90.6	100.0	90.8	100.0	90.6	100.0	90.6	100.0	89.7	100.0	90.6	100.0
	1:1:1:1	65.7	99.8	66.5	100.0	65.7	100.0	65.6	100.0	63.6	100.0	65.6	99.8
	4:3:2:1												
100:200:400:500	1:2:3:4	98.8	100.0	98.9	99.0	98.8	100.0	98.9	99.0	98.5	99.0	98.8	100.0
	1:1:1:1												

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe F^*. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .01. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

$\underline{\alpha = .05}$. At the "negative conditions" with large sample sizes, the empirical power estimates were about 88.8% and 100% for effect sizes of .1 and .4 respectively. At the "positive conditions" with effect size of .1, the empirical power estimates were about 36%, 59%, and 99.8% for small, moderate and large sample sizes respectively. At the "positive conditions" with effect size of .4, the empirical power estimates were above 97.8% across sample sizes.

When sample sizes were equal and *f*=.1 with homogeneous variances, the power estimates were around 16%, 26% and 85.7% for small, moderate and large sample sizes respectively; while the effect size increased to .4, the power estimates were around 95%, above 96% and 100% for small, moderate and large sample sizes respectively. When variances were heterogeneous with equal and moderate sample sizes, the power estimates were around 50% and 99.9% for moderate and large effect sizes respectively; when sample sizes increased to large, the power estimates were around 96% and 100% for moderate and large effect sizes respectively. It was also noted that the power estimates were around 10%, 20% and 83% across the three unequal sample sizes with homogeneous variances for *f*=.1, and around 95.8%, 99.9% and 100% for *f*=.4.

Table 56: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), k=4 at α =.05

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4												
	1:1:1:1	16.7	95.4	18.8	96.1	16.7	95.4	16.7	95.7	16.4	95.7	16.5	95.3
	4:3:2:1												
10:20:40:50	1:2:3:4	36.1	99.9	43.4	100.0	36.1	99.8	39.6	99.9	39.3	99.9	35.8	99.9
	1:1:1:1	10.5	95.8			10.5	94.3	10.6	95.7	10.4	95.7	10.1	95.7
60:60:60:60	1:2:3:4	49.7	100.0	51.5	100.0	49.7	100.0	49.7	100.0	49.6	100.0	49.4	100.0
	1:1:1:1	26.4	96.2	27.9	99.9	26.4	99.9	26.4	99.9	26.4	99.9	26.2	96.1
	4:3:2:1												
20:40:80:100	1:2:3:4	59.0	100.0	62.1	100.0	59.0	100.0	60.6	100.0	60.4	100.0	58.8	100.0
	1:1:1:1	20.4	99.9			20.4	100.0	20.6	100.0	20.4	100.0	20.2	99.9
300:300:300:300	1:2:3:4	96.3	100.0	96.4	100.0	96.3	100.0	96.3	100.0	96.2	100.0	96.3	100.0
	1:1:1:1	85.7	99.9	85.7	100.0	85.7	100.0	85.7	100.0	84.5	100.0	85.7	99.9
	4:3:2:1	88.8	100.0	88.8	100.0	88.8	100.0	88.8	100.0	88.5	100.0	88.8	100.0
100:200:400:500	1:2:3:4	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0
	1:1:1:1	83.2	100.0	83.6	100.0	83.2	100.0	83.5	100.0	82.3	100.0	83.2	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic.
Power was computed across 1000 replications at the nominal level of .05. f stands for power when k is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

$\underline{\alpha = .10}$. At the "negative conditions" with large sample sizes, the empirical power estimates were about 94% and 100% for effect sizes of .1 and .4 respectively. At the "positive conditions" with effect size of .1, the empirical power estimates were about 51%, 70%, and 99.8% for small, moderate and large sample sizes respectively. At the "positive conditions" with effect size of .4, the empirical power estimates were all 100% across sample sizes.

When sample sizes were equal and f=.1 with homogeneous variances, the power estimates were around 27%, 37% and 91% for small, moderate and large sample sizes respectively; when the effect size increased to .4, the power estimates were around 97%, above 98.2% and 100% for small, moderate and large sample sizes respectively. When variances were heterogeneous with equal and small sample sizes, the power estimates were around 49% and 99.8% for small and large effect sizes respectively; when sample sizes increased to be moderate, the power estimates were around 62% and 100% for small and large effect sizes respectively; when sample sizes became large, the power estimates were around 98% and 100% for small and large effect sizes respectively. In addition, the power estimates were around 18%, 31% and 90% across the three unequal sample sizes with homogeneous variances for *f*=.1, and 97%, 100% and 100% for *f*=.4.

Table 57: Empirical Power (%) for the Asymmetric Nonnormal Distribution with Skewness and Kurtosis of (3, 21), *k*=4 at α=.10

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	T_{ML}		T_{ADF}		T_{SB}		T_{YB1}		T_{YB}		T_{BC}	
		f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
30:30:30:30	1:2:3:4	48.9	99.9			48.9	99.8	48.8	99.9	48.5	99.9	48.4	99.9
	1:1:1:1	27.7	97.1	30.1	97.6	27.7	97.1	27.5	97.3	27.0	97.3	27.0	97.1
	4:3:2:1												
10:20:40:50	1:2:3:4	50.1	100.0	54.6	100.0	50.1	99.9	52.4	100.0	51.9	100.0	49.6	100.0
	1:1:1:1	18.1	97.5	20.6	98.2	18.1	97.3	18.9	97.4	18.8	97.4	17.8	97.3
60:60:60:60	1:2:3:4	61.2	100.0	62.2	100.0	61.2	100.0	61.0	100.0	60.9	100.0	60.8	100.0
	1:1:1:1	37.0	98.2	38.1	100.0	37.0	100.0	36.9	100.0	36.5	100.0	36.6	98.2
	4:3:2:1												
20:40:80:100	1:2:3:4	69.8	100.0	71.3	100.0	69.8	100.0	70.5	100.0	70.5	100.0	69.5	100.0
	1:1:1:1	30.7	100.0	33.2	100.0	30.7	100.0	31.8	100.0	31.3	100.0	30.5	100.0
300:300:300:300	1:2:3:4	98.1	100.0	98.1	100.0	98.1	100.0	98.1	100.0	98.0	100.0	98.1	100.0
	1:1:1:1	91.5	100.0	91.8	100.0	91.5	100.0	91.4	100.0	90.9	100.0	91.4	100.0
	4:3:2:1	94.0	100.0	94.1	100.0	94.0	100.0	94.0	100.0	93.5	100.0	94.0	100.0
100:200:400:500	1:2:3:4	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0	99.8	100.0
	1:1:1:1	90.6	100.0	90.9	100.0	90.6	100.0	90.6	100.0	89.5	100.0	90.4	100.0

Note. F = ANOVA *F*. W = Welch v_w. BF = Brown and Forsythe *F**. A = Alexander and Govern *A*. U = James second-order *U*. T_{ML} = SMM with ML. T_{ADF} = SMM with ADF. T_{SB} = SMM using Satorra and Bentler Scaled Chi-square statistic. T_{YB1} and T_{YB2} = SMM using Yuan and Bentler's two corrected test statistic. T_{BC} = SMM using Bartlett-corrected statistic. Power was computed across 1000 replications at the nominal level of .10. *f* stands for power when *k* is greater than or equal to 2. f_1=.2 and f_2=.8.
Bolded values indicate the Type I error rates below the lower bound of the Bradley's robustness interval.

Parameter Estimates

Parameter estimates provided by three RMM test statistics of T_{ML}, T_{ADF} and T_{SB} were tracked for all cells in the Type I error portion of the study. As mentioned before, the test statistics T_{YB1}, T_{YB2} are derived from the T_{ADF} and the test statistic T_{BC} is based on T_{ML}. Thus, these test statistics should produce the same parameter estimates correspondingly. The population means across groups and distributional shapes are all zero. The parameter estimates of the population means for each group across distributional shapes and estimation methods are summarized in Table 58, 59 and 60 for k=2, 3 and 4 respectively. When k=2, the parameter estimates of the population variances for each group across distributional shapes and estimation methods are summarized in Table 61. When k=3, the parameter estimates of the population variances for each group across the three estimation methods are summarized in Table 62, 63 and 64 for each distributional shape respectively. When k=4, the parameter estimates of the population variances for each group across the three estimation methods are summarized in Table 65, 66 and 67 for each distributional shape respectively.

Parameter Estimates of Population Means

When the distribution was normal, the parameter estimates of the population means yielded by the three estimation strategies of ML, ADF and SB were all very close to zero, the true population mean, deviating within ± .003 from zero, the true population mean, across different group situations. When the distribution became nonnormal but elliptically symmetric with skewness and kurtosis of (0, 3), the parameter estimates of the population means yielded by the three estimation strategies of ML, ADF and SB were again all very close to zero. However, the deviations from zero were a little larger, within ± .011 from zero. The largest deviations appeared at the sample sizes at the “positive conditions”.

When the distribution was asymmetric nonnormal with skewness and kurtosis of (3, 21), the parameter estimates of the population means yielded by the three estimation strategies deviated from zero more. The absolute values of the deviations at k=2 were between the range of .035 and .139, .008 and .048, as well as .001 and .013 for small, moderate and large sample sizes respectively at this distributional shape; when k=3, the absolute values of the deviations fell between the range of .045 and .102, .018 and .056, as well as .004 and .013 for each sample sizes; when k=4, the absolute values of the deviations were between the range of .048 and .128, .025 and .056, as well as .005 and .018 for each sample sizes. Generally, the parameter estimates deviated from the true population mean the most at the "positive conditions" and the deviations increased as sample sizes decreased.

Table 58: Average model implied population means by three RMM methods when k=2.

n_1: n_2	σ_1/ σ_2	(0,0)			(0,3)			(3,21)		
		ML	ADF	RB	ML	ADF	RB	ML	ADF	RB
4:16	1	0.000	-0.001	0.000	0.001	0.001	0.001	-0.086	-0.091	-0.086
	2.5	0.000	-0.001	0.000	0.000	-0.001	0.000	-0.055	-0.096	-0.055
	4	0.001	0.000	0.001	0.000	-0.002	0.000	-0.035	-0.084	-0.035
16:4	2.5	0.003	-0.001	0.003	0.009	0.010	0.009	-0.139	-0.130	-0.139
	4	-0.001	-0.001	-0.001	0.009	0.011	0.009	-0.152	-0.135	-0.152
10:10	1	0.001	0.001	0.001	0.004	0.004	0.004	-0.090	-0.086	-0.090
	2.5	0.000	0.000	0.000	0.005	0.005	0.005	-0.107	-0.105	-0.107
	4	-0.001	0.000	-0.001	0.004	0.004	0.004	-0.092	-0.094	-0.092
20:80	1	0.001	0.001	0.001	0.002	0.002	0.002	-0.024	-0.024	-0.024
	2.5	0.001	0.001	0.001	0.001	0.001	0.001	-0.013	-0.016	-0.013
	4	0.001	0.001	0.001	0.000	0.001	0.000	-0.008	-0.010	-0.008
80:20	2.5	0.000	0.000	0.000	0.005	0.005	0.005	-0.048	-0.046	-0.048
	4	0.000	0.000	0.000	0.004	0.004	0.004	-0.044	-0.042	-0.044
50:50	1	0.001	0.000	0.001	0.001	0.001	0.001	-0.025	-0.024	-0.025
	2.5	0.001	0.001	0.001	0.000	0.000	0.000	-0.023	-0.023	-0.023
	4	0.001	0.001	0.001	0.000	0.000	0.000	-0.016	-0.016	-0.016
100:400	1	0.000	0.000	0.000	0.001	0.001	0.001	-0.006	-0.006	-0.006
	2.5	0.000	0.000	0.000	0.000	0.000	0.000	-0.003	-0.003	-0.003
	4	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.001	-0.001
400:100	2.5	0.000	0.000	0.000	-0.002	-0.002	-0.002	-0.013	-0.013	-0.013
	4	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.011	-0.010	-0.011
250:250	1	0.000	0.000	0.000	0.000	0.000	0.000	-0.006	-0.006	-0.006
	2.5	0.000	0.000	0.000	0.000	0.000	0.000	-0.005	-0.005	-0.005
	4	0.000	0.000	0.000	0.000	0.000	0.000	-0.003	-0.003	-0.003

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

Table 59: Average model implied population means by three RMM methods when k=3.

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	(0,0)			(0,3)			(3,21)		
		ML	ADF	RB	ML	ADF	RB	ML	ADF	RB
30:30:30	1:2.5:4	-0.001	-0.001	-0.001	0.002	0.002	0.002	-0.059	-0.061	-0.059
	1:1:1	-0.001	-0.001	-0.001	0.001	0.001	0.001	-0.053	-0.052	-0.053
	4:2.5:1	-0.001	-0.001	-0.001	-0.002	-0.001	-0.002	-0.038	-0.045	-0.038
12:30:48	1:2.5:4	-0.003	-0.003	-0.003	0.002	0.001	0.002	-0.102	-0.099	-0.102
	1:1:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.052	-0.054	-0.052
60:60:60	1:2.5:4	0.000	0.000	0.000	0.003	0.003	0.003	-0.030	-0.031	-0.030
	1:1:1	0.000	0.000	0.000	0.001	0.001	0.001	-0.028	-0.028	-0.028
	4:2.5:1	0.000	0.001	0.000	0.000	0.000	0.000	-0.018	-0.020	-0.018
24:60:96	1:2.5:4	0.001	0.001	0.001	0.003	0.003	0.003	-0.056	-0.054	-0.056
	1:1:1	0.001	0.001	0.001	0.000	0.000	0.000	-0.028	-0.028	-0.028
300:300:300	1:2.5:4	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.007	-0.007	-0.007
	1:1:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.006	-0.006	-0.006
	4:2.5:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.004	-0.004	-0.004
120:300:480	1:2.5:4	0.000	0.000	0.000	0.000	0.000	0.000	-0.013	-0.013	-0.013
	1:1:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.006	-0.006	-0.006

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

Table 60: Average model implied population means by three RMM methods when k=4.

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	(0,0)			(0,3)			(3,21)		
		ML	ADF	RB	ML	ADF	RB	ML	ADF	RB
30:30:30:30	1:2:3:4	-0.001	-0.001	-0.001	0.002	0.002	0.002	-0.077	-0.079	-0.077
	1:1:1:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.058	-0.058	-0.058
	4:3:2:1	0.001	0.001	0.001	-0.003	-0.003	-0.003	-0.048	-0.058	-0.048
10:20:40:50	1:2:3:4	0.001	0.000	0.001	0.003	0.002	0.003	-0.128	-0.126	-0.128
	1:1:1:1	0.001	0.001	0.001	-0.001	-0.001	-0.001	-0.056	-0.060	-0.056
60:60:60:60	1:2:3:4	0.000	0.000	0.000	0.002	0.002	0.002	-0.040	-0.041	-0.040
	1:1:1:1	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.031	-0.031	-0.031
	4:3:2:1	0.001	0.001	0.001	-0.001	-0.001	-0.001	-0.025	-0.028	-0.025
20:40:80:100	1:2:3:4	0.002	0.002	0.002	0.001	0.001	0.001	-0.072	-0.071	-0.072
	1:1:1:1	0.001	0.001	0.001	-0.001	0.000	-0.001	-0.031	-0.032	-0.031
300:300:300:300	1:2:3:4	0.000	0.000	0.000	0.000	0.000	0.000	-0.009	-0.009	-0.009
	1:1:1:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.007	-0.007	-0.007
	4:3:2:1	0.001	0.001	0.001	0.000	0.000	0.000	-0.005	-0.005	-0.005
100:200:400:500	1:2:3:4	0.000	0.000	0.000	0.001	0.001	0.001	-0.018	-0.018	-0.018
	1:1:1:1	0.000	0.000	0.000	0.000	0.000	0.000	-0.007	-0.007	-0.007

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach

Parameter Estimates of Population Variances

K=2

When *k*=2, the population variance σ_1^2 has the values of 1, 6.25 and 16 for the standard deviation ratios σ_1/σ_2 of 1, 2.5 or 4 respectively, while the population variance σ_2^2 has the value of 1 across all conditions. When the distribution was normal or elliptically symmetric nonnormal with skewness and kurtosis of (0, 3), the parameter estimates of the population variances yielded by the three estimation strategies deviated from the population variances less than 3.6 and .3 for group one and group two respectively. The parameter estimates were most close to the true population variance values at the "positive conditions", followed by the conditions of "equal sample sizes", then the "negative conditions". In terms of estimation strategies, the ML and SB estimation strategies mirrored each other, providing parameter estimates of population variances slightly greater than the true population variances in most of the conditions. On the contrary, the ADF estimation strategy tended to yield parameter estimates of population variances slightly smaller than the true population variances in most of the conditions.

When the distribution was asymmetric nonnormal with skewness and kurtosis of (3, 21), the parameter estimates of the population variances provided by the ML and SB estimation strategies were a little larger (less than 2) than the true population parameters for small sizes at the "negative conditions" or conditions of equal sample sizes; at all other conditions for the "positive conditions" or when sample sizes were moderate or large, the ML and SB estimation strategies provided parameter estimates of the population variances slightly smaller (less than .2) than the true population parameters. The ADF estimation strategy yielded parameter estimates of the population variances smaller (less than 4) than the true population parameters throughout all sample size and variance ratio conditions, deviating more from the true population parameters than the other two estimations strategies.

Table 61: Average model implied variances by the RMM methods for k=2.

n_1: n_2	σ_1/σ_2	(0,0)						(0,3)						(3,21)					
		ML		ADF		RB		ML		ADF		RB		ML		ADF		RB	
		E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2
4:16	1	1.225	1.010	0.996	0.998	1.225	1.010	1.164	1.038	0.884	0.962	1.164	1.038	1.066	1.059	0.752	0.781	1.066	1.059
	2.5	7.656	1.002	6.213	0.998	7.664	1.002	7.426	1.047	5.509	0.988	7.426	1.047	6.786	1.039	4.673	0.836	6.786	1.039
	4	19.596	1.004	15.923	0.997	19.596	1.004	19.130	1.045	14.107	0.997	19.130	1.045	17.446	1.028	11.952	0.870	17.446	1.028
16:4	2.5	6.400	1.206	6.226	1.019	6.400	1.206	6.306	1.249	5.756	0.990	6.306	1.249	6.180	1.287	4.740	0.877	6.180	1.287
	4	16.560	1.163	15.923	1.021	16.560	1.163	16.310	1.205	14.590	1.006	16.310	1.205	15.858	1.307	12.129	0.892	15.858	1.307
10:10	1	1.065	1.069	1.006	0.998	1.065	1.069	1.028	1.094	0.920	0.951	1.028	1.094	1.039	1.102	0.743	0.727	1.039	1.102
	2.5	6.856	1.012	6.261	0.998	6.856	1.012	6.622	1.052	5.593	0.993	6.622	1.052	6.500	1.102	4.605	0.772	6.500	1.102
	4	17.551	1.000	15.994	0.996	17.551	1.000	17.062	1.039	14.219	1.011	17.062	1.039	16.691	1.091	11.788	0.815	16.691	1.091
20:80	1	1.028	1.004	0.987	1.002	1.028	1.004	1.020	1.008	0.920	0.999	1.020	1.008	0.986	1.037	0.771	0.931	0.986	1.037
	2.5	6.474	1.002	6.168	1.002	6.479	1.002	6.411	1.006	5.691	1.005	6.411	1.006	6.215	1.033	4.775	0.971	6.215	1.033
	4	16.605	1.002	15.803	1.002	16.605	1.002	16.425	1.006	14.546	1.005	16.425	1.006	15.939	1.032	12.205	0.991	15.939	1.032
80:20	2.5	6.280	1.016	6.226	0.991	6.266	1.016	6.213	1.037	6.048	0.968	6.213	1.037	6.185	1.026	5.437	0.809	6.185	1.026
	4	16.080	1.005	15.902	0.992	16.079	1.005	15.944	1.026	15.402	0.985	15.944	1.026	15.853	1.019	13.996	0.849	15.853	1.019
50:50	1	1.002	1.012	0.991	1.000	1.002	1.012	0.991	1.016	0.956	0.979	0.991	1.016	0.984	1.056	0.851	0.881	0.984	1.056
	2.5	6.308	1.004	6.191	1.000	6.313	1.004	6.236	1.008	5.890	0.998	6.236	1.008	6.184	1.050	5.273	0.931	6.184	1.050
	4	16.179	1.002	15.853	1.001	16.188	1.002	15.991	1.007	15.035	1.002	15.991	1.007	15.862	1.047	13.491	0.966	15.862	1.047
100:400	1	1.003	1.001	0.995	1.001	1.003	1.001	0.995	1.000	0.967	0.998	0.995	1.000	0.988	1.007	0.902	0.983	0.988	1.007
	2.5	6.289	1.001	6.226	1.001	6.281	1.001	6.230	0.999	6.028	0.999	6.230	0.999	6.184	1.006	5.624	0.996	6.184	1.006
	4	16.084	1.001	15.916	1.001	16.084	1.001	15.951	0.999	15.428	0.999	15.951	0.999	15.834	1.006	14.393	1.000	15.834	1.006
400:100	2.5	6.238	1.005	6.227	1.000	6.238	1.005	6.219	1.011	6.185	0.997	6.219	1.011	6.133	1.032	5.981	0.944	6.133	1.032
	4	15.977	1.003	15.940	1.000	15.977	1.003	15.929	1.009	15.816	1.002	15.929	1.009	15.707	1.030	15.353	0.965	15.707	1.030
250:250	1	0.998	1.003	0.996	1.001	0.998	1.003	0.997	1.003	0.989	0.997	0.997	1.003	0.982	1.010	0.949	0.965	0.982	1.010
	2.5	6.245	1.002	6.222	1.001	6.245	1.002	6.244	1.001	6.159	1.000	6.244	1.001	6.144	1.008	5.927	0.984	6.144	1.008
	4	15.991	1.001	15.926	1.001	15.991	1.001	15.990	1.001	15.753	1.001	15.990	1.001	15.734	1.008	15.182	0.993	15.734	1.008

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

E1= $\text{var}(\delta_1)$, which is the model implied variance for group one. E2= $\text{var}(\delta_2)$, which is the model implied variance for group two.

K=3 and *K*=4

When *k*=3, the population variances σ_1^2, σ_2^2 and σ_3^2 have the values of 1,1 1, and 1, 6.25, 16, as well as 16, 6.25, 1 for the standard deviation ratios of 1:1:1, 1:2.5:4, or 4:2.5:1. When *k*=4, the population variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ and σ_4^2 have the values of 1,1,1,1 and 1,2,9,16 as well as 16,9,4,1 with the standard deviation ratios following 1:1:1:1, 1:2:3:4, or 4:3:2:1.

The parameter estimates of the population variances yielded by the ML and SB estimation strategies yielded parameter estimates slightly deviating (less than 1.5) from the population variances across the three distributional shapes and sample size and variance ratio conditions. The deviations were largest at the "negative conditions", decreasing with larger sample sizes. However, the ADF estimation strategy tended to provide parameter estimates slightly lower than the true population variances. The deviations were less than .2, 1.8 and 4.3 for the normal distribution, and the two nonnormal distributions with skewness and kurtosis of (0, 3) and (3, 21) respectively. The deviations were largest at the "negative conditions", decreasing as sample sizes increased.

Table 62: Average model implied variances and population means by the RMM methods for *k*=3 and distribution with (0, 0).

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	ML			ADF			RB		
		E1	E2	E3	E1	E2	E3	E1	E2	E3
30:30:30	1:2.5:4	0.993	6.437	16.504	0.986	6.229	16.028	0.993	6.437	16.504
	1:1:1	1.008	1.025	1.026	0.987	0.998	1.002	1.008	1.025	1.026
	4:2.5:1	17.098	6.457	1.005	15.870	6.238	1.003	17.098	6.457	1.005
12:30:48	1:2.5:4	1.030	6.427	16.290	0.994	6.242	16.033	1.030	6.427	16.290
	1:1:1	1.061	1.026	1.012	0.992	0.999	1.003	1.061	1.026	1.012
60:60:60	1:2.5:4	0.994	6.351	16.266	0.991	6.255	16.046	0.994	6.351	16.266
	1:1:1	1.001	1.014	1.013	0.991	1.001	1.003	1.001	1.014	1.013
	4:2.5:1	16.442	6.355	1.003	15.771	6.257	1.002	16.442	6.355	1.003
24:60:96	1:2.5:4	1.002	6.337	16.179	0.986	6.253	16.026	1.002	6.337	16.179
	1:1:1	1.022	1.013	1.007	0.986	1.001	1.002	1.022	1.013	1.007
300:300:300	1:2.5:4	0.996	6.275	16.047	0.996	6.251	16.000	0.996	6.275	16.047
	1:1:1	0.998	1.003	1.002	0.996	1.000	1.000	0.998	1.003	1.002
	4:2.5:1	16.040	6.276	1.000	15.904	6.252	0.999	16.040	6.276	1.000
120:300:480	1:2.5:4	0.998	6.273	16.017	0.994	6.252	15.988	0.998	6.273	16.017
	1:1:1	1.001	1.003	1.000	0.994	1.000	0.999	1.001	1.003	1.000

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

E1= $\mathrm{var}(\delta_1)$, which is the model implied variance for group one. E2= $\mathrm{var}(\delta_2)$, which is the model implied variance for group two.

E3= $\mathrm{var}(\delta_3)$, which is the model implied variance for group three.

Table 63: Average model implied variances and population means by the RMM methods for k=3 and distribution with (0, 3).

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	ML			ADF			RB		
		E1	E2	E3	E1	E2	E3	E1	E2	E3
30:30:30	1:2.5:4	0.989	6.482	16.547	0.966	5.888	14.958	0.989	6.482	16.547
	1:1:1	1.005	1.031	1.026	0.935	0.955	0.950	1.005	1.031	1.026
	4:2.5:1	16.890	6.492	1.001	14.198	5.875	0.993	16.890	6.492	1.001
12:30:48	1:2.5:4	1.013	6.456	16.258	0.929	5.933	15.290	1.013	6.456	16.258
	1:1:1	1.047	1.031	1.009	0.895	0.955	0.973	1.047	1.031	1.009
60:60:60	1:2.5:4	0.986	6.394	16.265	0.974	6.062	15.331	0.972	6.567	16.426
	1:1:1	0.995	1.020	1.012	0.959	0.976	0.968	0.995	1.020	1.012
	4:2.5:1	16.284	6.399	1.001	14.579	6.051	0.995	15.638	6.568	1.010
24:60:96	1:2.5:4	0.995	6.381	16.144	0.945	6.087	15.568	0.965	6.559	16.275
	1:1:1	1.013	1.020	1.005	0.918	0.977	0.983	0.974	1.049	1.015
300:300:300	1:2.5:4	0.995	6.264	16.000	0.992	6.198	15.800	0.982	6.314	16.105
	1:1:1	0.997	1.002	0.999	0.988	0.993	0.990	0.984	1.010	1.006
	4:2.5:1	15.932	6.265	0.997	15.536	6.191	0.996	15.806	6.315	1.000
120:300:480	1:2.5:4	0.991	6.263	15.985	0.980	6.201	15.870	0.984	6.312	16.020
	1:1:1	0.995	1.002	0.998	0.973	0.993	0.994	0.987	1.010	1.001

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

E1= $\mathrm{var}(\delta_1)$, which is the model implied variance for group one. E2= $\mathrm{var}(\delta_2)$, which is the model implied variance for group two.

E3= $\mathrm{var}(\delta_3)$, which is the model implied variance for group three.

Table 64: Average model implied variances and population means by the SMM methods for *k*=3 and distribution with (3,21).

n_1: n_2: n_3	σ_1/ σ_2/ σ_3	ML			ADF			RB		
		E1	E2	E3	E1	E2	E3	E1	E2	E3
30:30:30	1:2.5:4	0.957	6.598	16.163	0.785	4.981	12.836	0.957	6.598	16.163
	1:1:1	0.965	1.056	1.008	0.755	0.776	0.780	0.965	1.056	1.008
12:30:48	4:2.5:1	16.354	6.606	1.008	11.737	5.011	0.881	16.354	6.606	1.008
	1:2.5:4	1.015	6.599	16.284	0.731	4.905	13.385	1.015	6.599	16.284
	1:1:1	1.016	1.056	1.016	0.718	0.773	0.818	1.016	1.056	1.016
60:60:60	1:2.5:4	0.972	6.567	16.426	0.869	5.488	14.175	0.972	6.567	16.426
	1:1:1	0.978	1.050	1.024	0.837	0.861	0.879	0.978	1.050	1.024
24:60:96	4:2.5:1	15.638	6.568	1.010	12.304	5.552	0.951	15.638	6.568	1.010
	1:2.5:4	0.965	6.559	16.275	0.775	5.456	14.421	0.965	6.559	16.275
	1:1:1	0.974	1.049	1.015	0.757	0.867	0.900	0.974	1.049	1.015
300:300:300	1:2.5:4	0.982	6.314	16.105	0.958	6.026	15.524	0.982	6.314	16.105
	1:1:1	0.984	1.010	1.006	0.947	0.962	0.966	0.984	1.010	1.006
120:300:480	4:2.5:1	15.806	6.315	1.000	14.572	6.041	0.987	15.806	6.315	1.000
	1:2.5:4	0.984	6.312	16.020	0.918	6.011	15.604	0.984	6.312	16.020
	1:1:1	0.987	1.010	1.001	0.906	0.963	0.974	0.987	1.010	1.001

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

E1= $\mathrm{var}(\delta_1)$, which is the model implied variance for group one. E2= $\mathrm{var}(\delta_2)$, which is the model implied variance for group two.

E3= $\mathrm{var}(\delta_3)$, which is the model implied variance for group three.

Table 65: Average model implied variances and population means by the RMM methods for k=4 and distribution with (0, 0).

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	ML				ADF				RB			
		E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
30:30:30:30	1:2:3:4	0.996	4.115	9.280	16.595	0.987	3.990	9.015	16.067	0.996	4.115	9.280	16.595
	1:1:1:1	1.012	1.028	1.027	1.032	0.988	1.000	1.002	1.006	1.012	1.028	1.027	1.032
10:20:40:50	4:3:2:1	17.355	9.439	4.091	1.004	15.926	8.956	4.010	1.002	17.355	9.439	4.091	1.004
	1:2:3:4	1.051	4.178	9.183	16.286	0.996	3.989	9.022	16.027	1.051	4.178	9.183	16.286
	1:1:1:1	1.079	1.045	1.019	1.012	0.995	0.998	1.003	1.002	1.079	1.045	1.019	1.012
60:60:60:60	1:2:3:4	0.996	4.062	9.146	16.301	0.991	4.004	9.025	16.045	0.996	4.062	9.146	16.301
	1:1:1:1	1.004	1.014	1.014	1.016	0.991	1.001	1.003	1.003	1.004	1.014	1.014	1.016
20:40:80:100	4:3:2:1	16.612	9.234	4.049	1.002	15.807	8.991	4.006	1.000	16.612	9.234	4.049	1.002
	1:2:3:4	1.015	4.085	9.109	16.150	0.987	3.996	9.017	15.994	1.015	4.085	9.109	16.150
	1:1:1:1	1.034	1.023	1.010	1.006	0.987	1.000	1.002	1.000	1.034	1.023	1.010	1.006
300:300:300:300	1:2:3:4	0.997	4.015	9.026	16.062	0.996	4.001	9.000	16.014	0.997	4.015	9.026	16.062
	1:1:1:1	0.998	1.004	1.002	1.003	0.996	1.000	1.000	1.001	0.998	1.004	1.002	1.003
100:200:400:500	4:3:2:1	16.084	9.047	4.005	1.002	15.914	8.991	3.996	1.001	16.084	9.047	4.005	1.002
	1:2:3:4	1.000	4.018	9.008	16.047	0.995	3.997	8.990	16.019	1.000	4.018	9.008	16.047
	1:1:1:1	1.004	1.005	1.001	1.002	0.995	0.999	0.999	1.001	1.004	1.005	1.001	1.002

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.
E1= $\mathrm{var}(\delta_1)$, which is the model implied variance for group one. E2= $\mathrm{var}(\delta_2)$, which is the model implied variance for group two.
E3= $\mathrm{var}(\delta_3)$, which is the model implied variance for group three. E4= $\mathrm{var}(\delta_4)$, which is the model implied variance for group four.

Table 66: Average model implied variances and population means by the RMM methods for *k*=4 and distribution with (0, 3).

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	ML				ADF				RB			
		E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
30:30:30:30	1:2:3:4	0.992	4.145	9.300	16.462	0.958	3.780	8.424	14.930	0.992	4.145	9.300	16.462
	1:1:1:1	1.007	1.035	1.028	1.023	0.931	0.950	0.944	0.944	1.007	1.035	1.028	1.023
	4:3:2:1	17.133	9.566	4.100	1.002	14.118	8.376	3.801	0.984	17.133	9.566	4.100	1.002
10:20:40:50	1:2:3:4	1.039	4.218	9.211	16.256	0.908	3.763	8.578	15.327	1.039	4.218	9.211	16.256
	1:1:1:1	1.067	1.058	1.020	1.010	0.886	0.937	0.959	0.967	1.067	1.058	1.020	1.010
60:60:60:60	1:2:3:4	0.988	4.089	9.146	16.206	0.970	3.886	8.629	15.400	0.988	4.089	9.146	16.206
	1:1:1:1	0.997	1.021	1.014	1.009	0.956	0.974	0.964	0.970	0.997	1.021	1.014	1.010
	4:3:2:1	16.419	9.258	4.047	1.001	14.533	8.574	3.878	0.994	16.419	9.258	4.047	1.001
20:40:80:100	1:2:3:4	1.007	4.099	9.101	16.127	0.932	3.843	8.731	15.662	1.007	4.099	9.101	16.127
	1:1:1:1	1.023	1.026	1.010	1.005	0.912	0.959	0.974	0.985	1.023	1.026	1.010	1.005
300:300:300:300	1:2:3:4	0.995	4.009	8.999	16.034	0.991	3.968	8.889	15.838	0.995	4.009	8.999	16.034
	1:1:1:1	0.997	1.002	0.999	1.001	0.987	0.992	0.989	0.992	0.997	1.002	0.999	1.001
	4:3:2:1	15.952	9.048	4.000	1.002	15.437	8.901	3.969	1.000	15.952	9.048	4.000	1.002
100:200:400:500	1:2:3:4	0.993	4.019	8.999	16.047	0.974	3.966	8.931	15.936	0.993	4.019	8.999	16.047
	1:1:1:1	0.996	1.005	1.000	1.002	0.966	0.990	0.993	0.998	0.996	1.005	1.000	1.002

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.
E1= $\mathrm{var}(\delta_1)$, which is the model implied variance for group one. E2= $\mathrm{var}(\delta_2)$, which is the model implied variance for group two.
E3= $\mathrm{var}(\delta_3)$, which is the model implied variance for group three. E4= $\mathrm{var}(\delta_4)$, which is the model implied variance for group four.

Table 67: Average model implied variances and population means by the RMM methods for *k*=4 and distribution with (3, 21).

n_1: n_2: n_3: n_4	σ_1/ σ_2/ σ_3/ σ_4	ML				ADF				RB			
		E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
30:30:30:30	1:2:3:4	0.964	4.224	9.090	16.734	0.737	3.112	7.095	12.803	0.964	4.224	9.090	16.734
	1:1:1:1	0.969	1.058	1.010	1.046	0.732	0.758	0.758	0.760	0.969	1.058	1.010	1.046
	4:3:2:1	16.695	9.427	4.101	1.008	11.671	6.807	3.309	0.824	16.695	9.427	4.101	1.008
10:20:40:50	1:2:3:4	1.047	4.195	9.239	16.252	0.681	2.879	7.237	13.187	1.047	4.195	9.239	16.252
	1:1:1:1	1.041	1.048	1.027	1.015	0.705	0.719	0.785	0.789	1.041	1.048	1.027	1.015
60:60:60:60	1:2:3:4	0.975	4.202	9.237	16.304	0.837	3.455	7.893	13.952	0.975	4.202	9.237	16.304
	1:1:1:1	0.980	1.051	1.025	1.018	0.822	0.846	0.858	0.845	0.980	1.051	1.025	1.018
	4:3:2:1	15.945	9.460	4.074	1.016	12.084	7.544	3.593	0.911	15.945	9.460	4.074	1.016
20:40:80:100	1:2:3:4	0.989	4.197	9.172	16.335	0.730	3.232	7.901	14.210	0.989	4.197	9.172	16.335
	1:1:1:1	0.993	1.050	1.018	1.020	0.737	0.810	0.873	0.874	0.993	1.050	1.018	1.020
300:300:300:300	1:2:3:4	0.983	4.040	9.058	16.125	0.950	3.844	8.712	15.425	0.983	4.040	9.058	16.125
	1:1:1:1	0.984	1.010	1.006	1.007	0.944	0.957	0.964	0.958	0.984	1.010	1.006	1.007
	4:3:2:1	15.834	9.140	4.004	1.008	14.356	8.562	3.901	0.984	15.834	9.140	4.004	1.008
100:200:400:500	1:2:3:4	0.986	4.059	9.009	16.153	0.891	3.775	8.704	15.585	0.986	4.059	9.009	16.153
	1:1:1:1	0.989	1.015	1.001	1.009	0.890	0.946	0.967	0.972	0.989	1.015	1.001	1.009

Note: ML=maximum likelihood estimation method; ADF= asymptotically distribution free method; and SB=Satorra and Bentler Scaled Chi-square approach.

E1= $\text{var}(\delta_1)$, which is the model implied variance for group one. E2= $\text{var}(\delta_2)$, which is the model implied variance for group two.

E3= $\text{var}(\delta_3)$, which is the model implied variance for group three. E4= $\text{var}(\delta_4)$, which is the model implied variance for group four.

Chapter V

Summary and Discussion

Type I Error Results Summary

Excitingly, the RMM approaches do provide comparatively robust Type I error rates across different distributional shapes and different sample size and variance conditions. When the distribution is normal, the Type I error rates across all ANOVA alternatives and RMM test statistics are robust across all conditions of sample size and variance ratios when k=3 and 4; but when k=2, the ANOVA alternatives tend to provide Type I error rates smaller than the lower boundary of the robust range at the "positive conditions" with small sample sizes. Thus the RMM test statistics control the Type I error rates better than the ANOVA alternatives when k=2. When the distribution is elliptically symmetric nonnormal with skewness and kurtosis of (0, 3), results were similar to those from the normal distribution. It indicates that the change of kurtosis does not affect the control of Type I error rates much across the ANOVA-based methods and RMM test statistics. When the distribution is asymmetric nonnormal with skewness and kurtosis of (3, 21), the ANOVA-based methods generally provided Type I error rates off the range of robustness when variances are heterogeneous. On the contrary, the RMM approaches provided much better control of Type I error rates.

In terms of test statistics, T_{ML} and T_{SB} mirrored one another closely. Derived from T_{ML}, T_{BC} provided similar Type I error rates. Among all the SMM test statistics, T_{YB1} and T_{YB2} yielded slightly better Type I error rates than T_{ADF}. Among the ANOVA alternatives, the Welch v_w, the Alexander and Govern A as well as James second-order U produced similar Type I error rates across conditions, while BF F^* yielded more non-robust cells than other ANOVA alternatives and RMM test statistics.

Empirical Power Results Summary

Normal distribution

The RMM test statistics generally yielded higher power estimates than the ANOVA-based methods across sample sizes, variance ratios, number of groups and the levels of significance. The advantage, however, decreased with sample sizes. Among all the methods, the T_{ADF} test statistic usually provided the best empirical power estimates, followed by the RMM test statistics of T_{YB1} and $T_{YB2.}$ The rest of the RMM test statistics also produced reasonably good empirical power estimates. Among all the ANOVA-based methods, the Alexander and Govern A statistic often yielded the highest value for the empirical power estimates across most of the conditions; on the contrary, the BF F^* statistic usually provided much lower empirical power estimates across all conditions than those from the rest of the ANOVA alternatives. The proposed RMM methods and the ANOVA alternatives all perform comparatively much better than the ANOVA F test, which generally yielded much smaller empirical power estimates at most of the conditions when the effect size was small, and only provided good power estimates with unequal sample sizes but homogenous variances. When the effect size increased to large, the empirical power estimates were all above 90%, approaching 100%.

Elliptically Symmetric Nonnormal Distribution (0,3)

The ANOVA F test provided comparable empirical power estimates for equal sample sizes across variance ratios, but much lower values at the “positive conditions”. (Note: Since the ANOVA F test usually provided inflated Type I error rates at the “negative conditions”, the power estimates were not produced.) The ANOVA alternatives and the RMM test statistics yielded much higher power estimates at the “positive conditions”.

When the effect size was small, the Welch v_w, the Alexander and Govern A as well as James second-order U tended to provide slightly higher empirical power estimates than the RMM test statistics, especially for large sample size. However, the differences were small, often less than 5%. Among all the methods, the Alexander and Govern A and the T_{ADF} statistics often provided the best empirical power estimates, followed by the Welch $v_{w,}$ James second-order U and the RMM test statistics of T_{YB1} and T_{YB2}. Once again, the BF F^* statistic usually provided much lower empirical power estimates across all conditions.

Asymmetric Nonnormal Distribution (3,21)

Due to the unsatisfactory results of Type I error rates from the ANOVA-based methods, the six RMM test statistics were studied for the power analysis for most of the conditions. Generally, the power estimates increased as sample sizes increased and also were higher at the "positive conditions" than those at the "negative condition" as expected. In sum, the power estimates provided by the RMM methods were all close to each other across conditions, with the T_{ADF} test statistic yielding slightly higher power estimate values and the T_{BC} test statistic yielding slightly lower power estimate values. However, the differences were tiny usually within 2%, which were ignorable.

Discussion

Methods such as ANOVA have been used for the last half-century by researchers in many fields for inference in experimental, quasi-experimental, and non-experimental designs. Foundational to the advancement of research in many fields, these methods, however, rest upon assumptions that are frequently not met. As a result, the inferences they lead may mistakenly declare population differences (e.g., treatment effects) that do not exist, and perhaps worse, fail to detect those effects that do exist. Thus, it is of paramount importance for researchers to use statistical methods for such designs that produce reliable results.

Addressing homogeneity of variance violations has involved the incorporation of weights and degrees of freedom adjustments, while normality violations have been met with suggestions to eliminate offending portions of sample data. More desirable would seem to be a paradigm that makes neither assumption in the first place. Specifically, drawing from the RMM literature within SEM, this dissertation proposes univariate models with no variance constraints to be analyzed with robust estimation strategies. The current study seeks to clarify these methods and understand their behavior empirically. The results show that they are generally superior in the control of Type I error and power, especially when the distribution is asymmetric nonnormal. These results have paradigmatic implications that could reverberate throughout and beyond social sciences.

Our current investigation clearly suggests that the RMM approaches are robust across most of the conditions and the distributional shapes, and are preferable to the ANOVA-based methods based on trimmed means and Winsorized variances. In order to apply the ANOVA-based methods appropriately, researchers must be very familiar with their data such as the population variability for each group, sample sizes as well as the distributional shape of the populations (e.g. the degree of nonnormality) to make the best choice of the statistical method and obtain more accurate inferences. However, the RMM approaches are superior to the ANOVA-based methods across the distributional shapes and most of other conditions of variance and sample size ratios. Even though the researchers are blind to their data, it is pretty safe for the researchers to utilize the RMM approaches, saving much efforts and time from the pre-study of their data.

Some of the key findings from the current study are the following:

- The investigation provides evidence that the ANOVA F test tends to yield inflated Type I error rates at the “negative conditions” and Type I error rates smaller than the low boundary of robust range, when the distribution is normal or the distribution is nonnormal but elliptically symmetric with skewness and kurtosis of (0, 3). Thus, the RMM methods and the ANOVA alternatives are much better in control of Type I error rates across most of the conditions. However, when sample size is small at “positive conditions”, the ANOVA alternatives tend to provide Type I error rates smaller than the lower boundary of the robust range. At this situation, the RMM methods are specially recommended, which tend to provide robust results.
- When the distribution is asymmetric nonnormal with skewness and kurtosis of (3, 21), both the ANOVA *F* test and it alternatives tend to yield nonrobust results when variances are heterogeneous. At this situation, the RMM approaches are highly recommended, which tend to control Type I error rates much better.
- Although criticized to produce high nonconvergence rate with large model and small sample sizes, the ADF estimation strategy performed well for our simple model with just one observed variable, resulting in 0% nonconvergence rate across conditions and distributional shapes. However, if the model grows to be complex, the performance of the ADF estimation strategy may be less satisfactory.
- It is also worth noting that the ML estimation strategy surprisingly works well across different distributional shapes, even under the extreme nonnormal distribution with skewness and kurtosis of (3, 21). This finding shows that the

ML estimation strategy can work fine for a simple model without any latent factor under nonnormality.

- Among all the RMM test statistics, the test statistics of T_{YB1} and T_{YB2} are most recommended by providing most robust cells across conditions and high values for power estimates. The T_{ADF} test statistic tended to provide more nonrobust cells though higher power estimates. The rest three test statistics of T_{ML}, T_{SB} and T_{BC} also performed reasonably, although slightly less well than the test statistics of T_{YB1} and $T_{YB2.}$

The success of the RMM approaches in the study supports the application of SMM to observed variables, and also ignites many future studies as our next step. First, the proposed RMM approaches can be extended to repeated measures design, when there are correlations among the groups, or there are multiple measures per subject. Although widely used many experimental studies involving a "control group", the design has the assumptions of normality and homogeneity of within group variances, in other words, homogeneity of within treatment variances and homogeneity of covariance between pairs of treatment levels. The second assumption is commonly referred to as the compound symmetry assumption. To deal with the violation of the assumption, the F ratio is usually corrected to a new critical value, such as the Geisser-Greenhouse, Box and Huynh and Feldt correction (Vonesh & Chinchilli, 1997). Our next study thus is to propose repeated robust means modeling approaches, and compare them to the repeated measures design and its corrections under a variety of sample sizes, variance ratios and distributional shapes.

Second, as briefly introduced in Chapter I, bootstrap methods are computer-intensive methods of statistical analysis that are widely used in variety of fields. In the field of SEM, Bollen and Stine (1992) proposed a bootstrap method for adjusting the *p* value associated with T_{ML}, which may yield more appropriate *p* values than the unadjusted T_{ML} under nonnormal conditions. Ichikawa and Konishi (1995) showed bootstrap estimated standard errors are less biased than unadjusted ML estimates under nonnormality. Additionally, Yung and Bentler (1996) provided promising evidence for the performance of the bootstrap in SEM using examples from existing data sets. Since the bootstrap methods can be applied to any level of modeling, it is interesting to incorporate the method to the proposed RMM approaches, and compare their performance to the bootstrapped ANOVA-based methods. The study shall add to the literature of the incorporation of the bootstrap methods to the ANOVA between-subjects designs.

Third, since the current study only investigated the any-pairs power rate (that is, the probability of detecting at least one true pairwise difference), future study to examine the error rates for pairwise comparison will be very interesting, providing more detailed guidance for researchers in utilizing the RMM approaches. In addition, instead of using factorial designs to detect the effects of two or more observed variables, the RMM approaches might also be developed to manipulate more observed variables while handling the interactions between variables at the same time. What's more, we have assumed the observed variable is perfectly measured in the study. The issue the effects of the error that might have occurred during survey process and the data collection can also be studied.

A.1.

Sample EQS syntax for SMM with ML estimation on independent groups design with *k* groups

> ***k*** = number of groups
> $\boldsymbol{n_1}$ = sample size for group one
> $\boldsymbol{n_k}$ = sample size for group *k*
> **Path1** = path for the data set of group one. For example, d:\EQS61\rt1.txt.
> **Path*k*** = path for the data set of group *k*. For example, d:\EQS61\rt*k*.txt.

/SPECIFICATIONS
VAR=1; cases= $\boldsymbol{n_1}$; ME=ML; MA=RAW; DATA='**Path1**';
ANAL=MOMENT; GROUPS=***k***;
/EQUATIONS
V1 = *V999 + 1.000 E1;
/VAR
E1= *;
/PRINT
FIT=ALL; COV=YES;
/END
.
.
.
(Repeat the first part for each of the middle groups.)
.
.
.
/SPECIFICATIONS
VAR=1; cases= $\boldsymbol{n_k}$; ME=ML; MA= RAW; DATA=' **Path*k***';
ANAL=MOMENT;
/EQUATIONS
V1 = *V999 + 1.000 E1;
/VAR
E1= *;
/Constraints
(1,V1,V999)=…=(***k***,V1,V999);
/PRINT
FIT=ALL; COV=YES;
/END

A.2.
Sample EQS syntax for SMM with ADF estimation on independent groups design with *k* groups

k = number of groups
$\boldsymbol{n}_{\mathbf{1}}$ = sample size for group one
$\boldsymbol{n}_{\mathbf{k}}$ = sample size for group *k*
Path1 = path for the data set of group one. For example, d:\EQS61\rt1.txt.
Path*k* = path for the data set of group *k*. For example, d:\EQS61\rt*k*.txt.

/SPECIFICATIONS
VAR=1; cases= $\boldsymbol{n}_{\mathbf{1}}$; ME=AGLS; MA=RAW; DATA='**Path1**';
ANAL=MOMENT; GROUPS=***k***;
/EQUATIONS
V1 = *V999 + 1.000 E1;
/VAR
E1= *;
/PRINT
FIT=ALL; COV=YES;
/END
.
.
.
(Repeat the first part for each of the middle groups.)
.
.
.
/SPECIFICATIONS
VAR=1; cases= $\boldsymbol{n}_{\mathbf{k}}$; ME=AGLS; MA= RAW; DATA=' **Path*k***';
ANAL=MOMENT;
/EQUATIONS
V1 = *V999 + 1.000 E1;
/VAR
E1= *;
/Constraints
(1,V1,V999)=…=(***k***,V1,V999);
/PRINT
FIT=ALL; COV=YES;
/END

A.3.
Sample EQS syntax for SMM with SB estimation on independent groups design with *k* groups

> ***k*** = number of groups
> $\boldsymbol{n_1}$ = sample size for group one
> $\boldsymbol{n_k}$ = sample size for group *k*
> **Path1** = path for the data set of group one. For example, d:\EQS61\rt1.txt.
> **Path*k*** = path for the data set of group *k*. For example, d:\EQS61\rt*k*.txt.

```
/SPECIFICATIONS
VAR=1; cases= n1; ME= ML,ROBUST; MA=RAW; DATA='Path1';
ANAL=MOMENT; GROUPS=k;
/EQUATIONS
V1 = *V999 + 1.000 E1;
/VAR
E1= *;
/PRINT
FIT=ALL; COV=YES;
/END
.
.
.
(Repeat the first part for each of the middle groups.)
.
.
.
/SPECIFICATIONS
VAR=1; cases= nk; ME= ML,ROBUST; MA= RAW; DATA=' Pathk';
ANAL=MOMENT;
/EQUATIONS
V1 = *V999 + 1.000 E1;
/VAR
E1= *;
/Constraints
(1,V1,V999)=…=(k,V1,V999);
/PRINT
FIT=ALL; COV=YES;
/END
```

Reference:

Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, 19(2), 91-101.

Algina, J., ashima, T. C., & Lin, W. (1994). Type I error rates for the Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups. *Journal of Educational and Behavioral Statistics*. 19(3), 275-291.

Anderson, J. C.,&Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155–173.

Bartlett, M. S. (1950). Tests of significance in factor analysis. British Journal of Psychology, *Statistical Section*, 3, 77-85.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 34, 144-152. Brownie, C.

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of means. *Technometrics*. 16(1), 129-132.

Browne, M. W. (1982). *Covariance structures. In D. M. Hawkins* (Ed.), Topics in applied multivariate analysis (pp. 72–141). Cambridge, England: Cambridge University Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, 21, 205–229.

Budescu, D. V. (1982). The power of the F test in normal populations with heterogeneous variances. *Educational and Psychological Measurement*, 42, 409-416.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.

Clinch, J. J., & Keselman, H. J. (1982). Parametric alternatives to the analysis of variance, *Journal of Educational Statistics*, 7(3), 207-214.

Cohen, J. (1976). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised edition). New York: Academic Press.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and non-parametric statistics. *The American Statistician*, 35, 124-129.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.

Dijkstra, J. B., & Wetter, P. S. P. J. (1981). Testing the equality of several means when population variances are unequal. *Communications in Statistics-Simulation and Computation*. B10, 557-569.

Dixon, W. J. (1992). *BMDP statistical software manual (Vol. 1).* Los Angeles: University of California Press.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research,* 42, 237-288. Hays, W.

Fleishman, A. I.(1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532.

Finch, J. F., West, S. G.,&MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling*, 4, 87–107.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, 6, 391-398.

Fouladi, R. T. (1998, April). *Covariance structure analysis techniques under conditions of multivariate normality and nonnormality — modified and bootstrap based test statistics.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Fouladi, R. T. (1999, April). *Model fit in covariance structure analysis under small sample conditions — Modified maximum likelihood and asymptotically distribution free generalized least squares procedures*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal,* 7, 356-410.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research*, 42, 237-288.

Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development*, 20, 91-105.

Hancock, G. R. (2003). Fortune cookies, measurement error, and experimental design. *Journal of Modern Applied Statistical Methods*, 2, 293-305.

Hancock, G. R. (2004). *Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables.* In D. Kaplan (Ed.), The SAGE Handbook of Quantitative Methodology for the Social Sciences. Thousand Oaks, CA: SAGE Publications.

Harwell, M. R. (1992). Summarizing Monte Carlo results in methodological research. *Journal of Educational Statistics*, 17, 297-313.

Hill, G. W. (1970). Algorithm 395: Student's t-distribution. *Communications of the ACM*. 13, 619-620.

Hsiung, T., & Olejnik, S. (1996). Type I error rates and statistical power for the James second-order test and the univariate F test in two-way fixed-effects ANOVA models under heteroscedasticity and/or nonnormality. The *Journal of Experimental Education*. 65, 57-71.

Hsu, P. (1938). Contributions to the theory of "Student's" t test applied to the problem of two samples. *Statistical Research Memoirs*, 2,1-24.

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.

Ichikawa, M., & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. *Psychometrika*, 60, 77–93.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38, 324-329.

Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks, CA: Sage Publications.

Kendall, M. & Stuart, A. (1977). *The advanced Theory of Statistics*, vol. 1, 4th ed. New York: Macmillan.

Keselman, H. J., Algina, J., Wilcox, R. R., & Kowalchuk, R. K. (2000).Testing repeated measures hypotheses when covariance matrices are heterogeneous: Revisiting the robustness of the Welch–James test again. *Educational and Psychological Measurement*, 60, 925–938.

Keselman, H. J., Kowalchuk, R. K., & Lix, L. M. (1998). Robust nonorthogonal analyses revisited: An update based on trimmed means. *Psychometrika*, 63, 145–163.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621.

Levy, K. J. (1978). An empirical comparison of the ANOVA F-test with alternatives which are more robust against heterogeneity of variance. *Journal of Statistical Computation and Simulation,* 8, 49-57.

Lix, L. M., & Keselman, H. J. (1995). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. *Psychological Bulletin*, 117, 547-560.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of mean equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement,* 58, 409–429, (58, 853).

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test. *Review of Educational Research*, 66(4), 579-619.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156–166.

Muthén, B. (1989). Multiple-group structural modeling with non-normal continuous variables. . *British Journal of Mathematical and Statistical Psychology,* 42, 55 – 62.

Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30.

Nevitt, J. & Hancock, G. R. (1999). PWRCOEFF & NNORMULT: A set of programs for simulating multivariate nonnormal data. *Applied Psychological Measurement*, 23, 54.

Nevitt, J. & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 8, 353-377.

Nevitt, J. & Hancock, G. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, 39 (3), 439-478.

Oshima, T. C, & Algina, J. (1992a). A SAS program for testing the hypothesis of the equal means under heteroscedasticity: James's second-order test. *Educational and Psychological Measurement*, 52, 117-118.

Oshima, T. C., & Algina, J. (1992b). Type I error rates for James's second-order test and Wilcoxon's Hm test under heteroscedasticity and non-normality. *British Journal of Mathematical and Statistical Psychology*, 45, 255-263.

Overall, J. E., Atlas, R. S., Gibson, J. M. (1995). Tests that are robust against variance heterogeneity in k × 2 designs with unequal cell frequencies. *Psychological Reports*. 76:1011–1017.

Ramberg, J. S., & Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17, 78-82.

Rogan, J. C., & Keselman. H. J. (1977). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal? An investigation via a coefficient of variation. *American Educational Research Journal*, 14(4), 493-498.

SAS Institute, Inc. (1990). *SAS/IML software: Usage and reference, Version 6.* Cary, NC: Author.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika,* 6, 309-316.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin,* 2, 110-114.

Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality & Quantity*, 24, 367-386.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology,* 22, 249-278.

Satorra, A. (1999). *Scaled and adjusted restricted tests in multi-sample analysis of moment structures.* To appear in Innovations in Multivariate Statistical Multivariate Analysis: A Festschrift for Heinz Neudecker (R.D.H. Heijmans, D.S.G. Pollock, and A. Satorra, edts.), Dordrecht: Kluwer Academic Publishers.

Satorra, A. & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association 1988 proceedings of the Business and Economics Sections (pp. 308-313)*. Alexandria, VA: American Statistical Association.

Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp.399-419). Thousand Oaks, CA:Sage.

Sawilowsky, S, Blair, R. C., & Higgins, J. J. (1989). An investigation of Type I error and power properties of the rank transformation procedure in factorial ANOVA. *Journal of Educational Statistics*, 14, 255-267.

Schneider, P. J. , & Penfield, D. A. (1997). Alexander and Govern's approximation: Providing an alternative to ANOVA under variance. *The Journal of experimental education*, 65(3), 271.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology,* 27, 229–239.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*. 99(1), 90-99.

van der Waerden, B. L. (1952). Order tests for the two sample problem and their power. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen (A), 55 (Indagationes Mathematical 14),* 453-458.

Vonesh, E.F., and Chinchilli, V.M. 1997. *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, New York.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, 26, 208–221.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29:350–362.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrik,* 34:29–35.

Welch, B. L. (1951). On the comparison of several means: An alternative approach. *Biometrika*, 38. 330-336.

West, S. G, Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56−75). Thousand Oaks, CA: Sage.

Wilcox, R. R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology,* 41, 109-117.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, 65, 51–77.

Wilcox, R. R. (1996). *Statistics for the Social Sciences*. New York: Academic Press.

Wilcox, R. R., Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA, E W and F[a] statistics. *Communications in Statistics-Simulation*, 15(4), 933-943.

Wilcox, R. R., Keselman, H. J., & Kowalchuk, R. K. (1998). Can tests for treatment group equality be improved?: The bootstrap and trimmed means conjecture. British *Journal of Mathematical and Statistical Psychology*, 51, 123–134.

Wilcox, R. R., Keselman, H. J., Muska, J., & Cribbie, R. (2000). Repeated measures ANOVA: Some new results on comparing trimmed means and means. *British Journal of Mathematical and Statistical Psychology*, 53, 69–82.

Wolfram Research, Inc. (2003). Mathematica, Version 5.0, Champaign, IL.

Yung, Y.-F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 195–226). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yuan, K.-H. & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51, 289-309.

Yuan, K. & Bentler, P. M. (1999). F tests for mean and covariance structure analysis. *Journal of Education and Behavioral Statistics*, 3, 225-243.

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, 61, 165-170.

Zimmerman, D. W. (1996). A note on homogeneity of variance of scores and ranks. *Journal of Experimental Education*, 64, 351-362.

Zimmerman, D. W. (2000).Statistical nominal levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology*, 127(4), 354-364.

Zimmerman, D. W., & Zumbo, B. D. (1993). Rank transformations and the power of the Student t test and Welch t' test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology*, 47, 523-539.