

Monte Carlo Rates for Robust I Samples Test for Homogeneity of Variance

5.1 METHODOLOGY

A Monte Carlo study was carried out to evaluate each of the tests with respect to type I error stability and power. The type I error stability study assessed the relationship between observed rejection rates and the nominal rejection rates (α) when the HOV hypothesis (1.1) was true. The power study evaluated the ability of each test to detect differences among sample variances when the HOV hypothesis (1.1) was false.

FORTRAN programs were written to perform the Monte Carlo study. The program used to simulate the randomization tests is in the Appendix. The program had to be modified slightly to evaluate different numbers of populations and different sample sizes. The non-randomization tests used 10,000 replications, and the randomization tests used 2,000 replications with 1,000 shuffles per replication.

Since the focus of this paper was on robust tests, six of the seven distributions were nonnormal in the type I error stability study and power study. Three common and four special distributions were used in each study. The three common distributions were $N(0,1)$, $\chi^2(1)$ and an exponential distribution with $I = 1$, $\text{Exp}(1)$. The four special

distributions were (i) a symmetric distribution with kurtosis of 6 (generated using a method devised by Fleishman (1978) and employing tables from Barnes (1981)) to obtain

$$KUR(6) = 0.66268N(0,1) + 0.10189N^3(0,1) \quad (5.1)$$

(ii) a distribution with no kurtosis and skewness of 3 (generated using Fleishman (1978) and Barnes (1981)) to obtain

$$SKW(3) = -0.05134 - 2.91756N(0,1) + 0.05134N^2(0,1) + 0.87133N^3(0,1) \quad (5.2)$$

(iii) Gamma $\left(\frac{4}{9}, 1\right)$ and (iv) a 50:50 mixture of two normals where one was $N(-2,1)$, and the other was $N(2,1)$.

The majority of the power study was conducted with the $KUR(6)$ and $SKW(3)$; however, the power of each test was evaluated for each of the distributions in the type I error stability study.

5.2 VARIANCE CONFIGURATIONS STUDIED

Two variance configurations were studied:

- Configuration 1, where $I - 1$ variances were equal and the last was larger. This configuration was of the form $(1, \dots, 1, r)$.
- Configuration 2, where $I - 2$ variances were equal, the first was smaller, and the last was larger. This configuration was of the form $(1, m, \dots, m, r)$. Two variations of Configuration 2 were studied. One arrangement was $m = (0.5)r$, and the other arrangement was $m = (0.75)r$. The latter variation of Configuration 2

was studied to see the performance of the tests on a configuration thought to produce lower power.

Configuration 1 was chosen because it was a common circumstance of interest, and it was a favorable LFC. Configuration 2 was selected because it was thought to be an unfavorable LFC.

5.3 TYPE I ERROR STABILITY

The type I error rates are given in the case where each of the I populations have the same variance. The ideal test will be robust across different underlying distributions. That is, a robust test will have a type I error rate that is consistent from distribution to distribution. Similar to Conover, Johnson and Johnson (1981), Wludyka and Nelson (1999) gave guidelines on assessing the degree of robustness. A test was deemed to have “good robustness” if its rejection rate was less than twice the nominal rate and “adequate robustness” if its rejection rate was less than three times the nominal rate. Tests may also be classified as either “conservative” or “liberal.” “Conservative” tests are those in which the empirical rejection rate is less than the nominal rate. “Liberal” tests are those in which the empirical rejection rate is greater than the nominal rate.

Tables 4 and 5 contain results from the Monte Carlo study performed to evaluate the empirical type I error stability. Table 4 holds information from the study conducted using five samples, and Table 5 holds similar information for 10 samples.

Levene and ANOMV-LEV were the only nonrandomization tests that displayed “good robustness” for each of the six nonnormal distributions, although ANOMV-LEV slightly exceeded the criterion for SKW(3). ANOMV-JK exhibited “good robustness” for the mixture, but the test was “inadequate” for SKW(3) and $\mathbf{c}^2(1)$. ANOMV-TR demonstrated “good robustness” for all distributions except Exp(1) and $\mathbf{c}^2(1)$. This test was “inadequate” for $\mathbf{c}^2(1)$. F-K showed “good robustness” for parent distributions of KUR(6), SKW(3) and the mixture. The test was “inadequate” for $\mathbf{c}^2(1)$. Levene’s test proved to be conservative for all parent distributions. None of the other nonrandomization tests were consistently conservative or liberal. ANOMV-LEV and ANOMV-JK leaned toward conservative and liberal, respectively. ANOMV-TR and ANOMV-FK jumped around the nominal rejection rate.

Two of the four randomization tests demonstrated “good robustness” for the nonnormal parent distributions. RANDANOMV-D and RANDANOMV-R were the two tests that were “good” for all distributions. RANDANOMV-DD and RANDANOMV-RD were “inadequate” for $\mathbf{c}^2(1)$, Exp(1) and Gamma $\left(\frac{4}{9}, 1\right)$ when permutation shuffling was used. RANDANOMV-DD exhibited “good” robustness when bootstrap shuffling was used. None of the four tests were consistently conservative or liberal; however, the tests were more conservative when bootstrap shuffling was used.

5.4 POWER

Tables 6 – 12 contain results from the power study. Tables 6 – 9 hold power information for KUR(6). Tables 10 and 11 contain power results for SKW(3), and Table 12 contains the results from a power study using one variance configuration for the other five distributions used in the type I error stability study.

Each of the tests had greater power for Configuration 1 than for variance Configuration 2. The power of each test increased with n . Among the randomization tests RANDANOMV-D and RANDANOMV-R appear to be the best. When looking at the nonrandomization tests, ANOMV-LEV was the best. There were no significant differences in power when comparing the nonrandomization group as a whole to the randomization group as a whole when permutation shuffling was used. Bootstrap shuffling had a negative impact on the power of the randomization tests. This type of shuffling had the greatest impact on RANDANOMV-D and RANDANOMV-DD while the other randomization tests were impacted minimally. There were instances where individual tests were inappropriate for the distribution under study.

5.5 UNEQUAL MEANS

It was noted in section 3.3 that many non-parametric tests of variability assume a common center of location (Baker 1995). For this reason RANDANOMV-DD was developed. The following example will illustrate the problem that arises with RANDANOMV-D when populations have unequal means.

Consider the sample below for which $I = 3$, $n_i = 2$, $s_1^2 = 0.5$, $s_2^2 = 0.5$, $s_3^2 = 1250$ and $\overline{s^2} = 417$.

$$X = \begin{pmatrix} 2 & 20 & 100 \\ 3 & 21 & 150 \end{pmatrix}$$

From the initial sample $AD_{\max} = 833$ and $AD_{\min} = -416.5$. There are six distinct permutations of the data that produce different test statistics. That is, the original sample produces the same test statistics as

$$X = \begin{pmatrix} 20 & 100 & 2 \\ 21 & 150 & 3 \end{pmatrix}.$$

Two such distinct permutations are

$$X^{(1)} = \begin{pmatrix} 2 & 3 & 100 \\ 20 & 21 & 150 \end{pmatrix} \text{ and } X^{(2)} = \begin{pmatrix} 2 & 20 & 21 \\ 3 & 100 & 150 \end{pmatrix}.$$

For the first permutation $s_1^2 = 162$, $s_2^2 = 162$, $s_3^2 = 1250$, $\overline{s^2} = 524.67$, $AD_{\min}^1 = -362.67$ and $AD_{\max}^1 = 725.33$. The test statistics from the permuted data do not exceed the test statistics from the original sample. For the second permutation $s_1^2 = 0.5$, $s_2^2 = 3200$, $s_3^2 = 8320.5$, $\overline{s^2} = 3840.33$, $AD_{\min}^2 = -3839.83$ and $AD_{\max}^2 = 4480.17$. The test statistics from the permuted data exceed the test statistics from the original sample.

This was done for the remaining four permutations. Four of the six distinct permutations produced test statistics that exceeded the initial test statistics. Thus, one would conclude that there was no difference in variability among the samples. This example illustrates a problem with RANDANOMV-D: this test may have difficulty detecting differences

among variances if the populations have very different means. Thus, caution should be exercised when using this test when the means are known to be very different.

A Monte Carlo study was carried out using an unequal means case. The study used a situation where the means were not “too different.” The type I error results are in Table 10. The results from the power study are in Table 11.

The example presented earlier in this section showed that large differences in the mean could lead to problems with RANDANOMV-D. The Monte Carlo results show that small differences in the mean may not have a big impact on RANDANOMV-D or any of the randomization tests.

From the Monte Carlo study it was seen that bootstrap shuffling is slightly more robust and less powerful than permutation shuffling. The small increase in robustness was offset by the decrease in power. It was also shown that small differences in the mean may not greatly impact the performance of the randomization tests, but care should be exercised when using these tests when the means greatly differ. Because of the previous two issues, the user should employ randomization tests with permutation shuffling when it is known that the means are not very different. When the randomization tests are used in these situations, RANDANOMV-D and RANDANOMV-R are the best, especially with larger sample sizes. These two tests control the type I error rate for all of the distributions in the study, and these tests are as robust as commonly used nonrandomization tests. With smaller sample sizes, RANDANOMV-D and

RANDANOMV-R are still robust tests, but the user sacrifices some power. The user should be hesitant about using RANDANOMV-DD and RANDANOMV-RD when the data is skewed as the Monte Carlo study shows inflated type I error rates for these tests.

Table 1: Empirical Type I Error Rates for $I = 5, n = 10$ (20 in parentheses), $\alpha = 0.05$ left of semi-colon, $\alpha = 0.10$ right of semi-colon. * Denotes Bootstrap Shuffling

Test	Distribution						
	N(0,1)	KUR(6)	SKW(3)	Chi-sq(1)	Exp(1)	Gamma(4/9,1)	50:50 Mixture
Levene	.029 (.033) ; .061 (.071)	.032 (.037) ; .073 (.083)	.015 (.017) ; .039 (.044)	.049 (.045) ; .095 (.094)	.046 (.049) ; .092 (.096)	.044 (.042) ; .088 (.082)	.018 (.014) ; .035 (.033)
ANOMV-LEV	.032 (.034) ; .069 (.074)	.044 (.047) ; .090 (.092)	.028 (.037) ; .066 (.083)	.063 (.055) ; .116 (.103)	.057 (.055) ; .103 (.100)	.056 (.046) ; .102 (.089)	.019 (.016) ; .042 (.036)
ANOMV-JK	.051 (.052) ; .088 (.091)	.099 (.100) ; .174 (.170)	.246 (.313) ; .344 (.430)	.136 (.122) ; .217 (.201)	.119 (.115) ; .195 (.185)	.121 (.110) ; .196 (.186)	.024 (.030) ; .043 (.060)
ANOMV-TR	.034 (.035) ; .072 (.073)	.041 (.039) ; .085 (.082)	.018 (.020) ; .044 (.048)	.154 (.180) ; .245 (.282)	.087 (.098) ; .145 (.171)	.081 (.087) ; .146 (.160)	.014 (.009) ; .032 (.025)
F-K	.031 (.033) ; .066 (.073)	.038 (.038) ; .080 (.081)	.017 (.018) ; .040 (.045)	.162 (.198) ; .254 (.305)	.087 (.106) ; .153 (.177)	.085 (.092) ; .149 (.164)	.012 (.009) ; .027 (.022)
RAND-DI	.057 (.052) ; .108 (.091)	.053 (.058) ; .099 (.091)	.046 (.066) ; .098 (.099)	.050 (.066) ; .090 (.110)	.054 (.048) ; .113 (.104)	.043 (.047) ; .096 (.089)	.050 (.046) ; .095 (.093)
RAND-D	.061 (.043) ; .108 (.083)	.059 (.046) ; .102 (.091)	.049 (.046) ; .089 (.113)	.058 (.064) ; .110 (.107)	.050 (.047) ; .101 (.093)	.048 (.050) ; .092 (.101)	.044 (.044) ; .090 (.090)
RAND-DD	.082 (.058) ; .145 (.099)	.114 (.072) ; .177 (.120)	.160 (.111) ; .231 (.165)	.410 (.272) ; .501 (.357)	.267 (.169) ; .367 (.235)	.258 (.207) ; .350 (.275)	.061 (.047) ; .107 (.082)
RAND-R	.060 (.043) ; .111 (.083)	.059 (.047) ; .107 (.090)	.047 (.051) ; .089 (.088)	.053 (.055) ; .102 (.102)	.054 (.042) ; .097 (.088)	.049 (.046) ; .096 (.103)	.044 (.046) ; .086 (.086)
RAND-RD	.082 (.058) ; .145 (.099)	.114 (.072) ; .177 (.120)	.160 (.111) ; .231 (.165)	.410 (.272) ; .501 (.357)	.267 (.169) ; .367 (.235)	.258 (.207) ; .350 (.275)	.061 (.047) ; .107 (.082)
*RAND-DI	.043 ; .094	.042 ; .103	.030 ; .094	.031 ; .105	.031 ; .105	.048 ; .108	.041 ; .091
*RAND-D	.026 ; .067	.018 ; .047	.015 ; .048	.012 ; .034	.012 ; .038	.018 ; .053	.030 ; .076
*RAND-DD	.027 ; .078	.021 ; .055	.025 ; .051	.038 ; .087	.045 ; .099	.050 ; .106	.046 ; .094
*RAND-R	.032 ; .071	.039 ; .074	.023 ; .065	.022 ; .058	.029 ; .064	.027 ; .072	.027 ; .055
*RAND-RD	.055 ; .112	.068 ; .147	.070 ; .160	.340 ; .442	.212 ; .322	.209 ; .302	.054 ; .098

Table 2: Empirical Type I Error Rates for $I = 10, n = 10$ (20 in parentheses), $\alpha = 0.05$ left of semi-colon, $\alpha = 0.10$ right of semi-colon. * Denotes Bootstrap Shuffling

Test	Distribution						
	N(0,1)	KUR(6)	SKW(3)	Chi-sq(1)	Exp(1)	Gamma(4/9,1)	50:50 Mixture
Levene	.022 (.034) ; .053 (.076)	.032 (.046) ; .071 (.093)	.013 (.023) ; .032 (.050)	.047 (.053) ; .094 (.105)	.050 (.051) ; .092 (.098)	.044 (.046) ; .084 (.090)	.009 (.007) ; .021 (.021)
ANOMV-LEV	.032 (.037) ; .061 (.075)	.074 (.070) ; .125 (.119)	.109 (.127) ; .187 (.202)	.096 (.091) ; .154 (.140)	.082 (.073) ; .134 (.123)	.080 (.071) ; .130 (.114)	.013 (.016) ; .030 (.033)
ANOMV-JK	.051 (.054) ; .086 (.101)	.128 (.137) ; .210 (.228)	.539 (.623) ; .627 (.718)	.157 (.155) ; .246 (.239)	.141 (.131) ; .227 (.214)	.142 (.135) ; .227 (.219)	.035 (.035) ; .050 (.065)
ANOMV-TR	.032 (.036) ; .061 (.075)	.042 (.043) ; .081 (.086)	.020 (.023) ; .042 (.050)	.180 (.236) ; .281 (.359)	.102 (.113) ; .167 (.196)	.092 (.105) ; .155 (.190)	.009 (.010) ; .023 (.025)
F-K	.024 (.034) ; .053 (.076)	.036 (.042) ; .074 (.088)	.011 (.014) ; .026 (.037)	.227 (.329) ; .338 (.454)	.104 (.136) ; .178 (.219)	.098 (.131) ; .168 (.217)	.005 (.005) ; .013 (.015)
RAND-DI	.031 (.046) ; .082 (.096)	.041 (.042) ; .086 (.103)	.063 (.046) ; .114 (.092)	.052 (.054) ; .113 (.101)	.051 (.068) ; .092 (.126)	.031 (.043) ; .076 (.087)	.051 (.044) ; .106 (.092)
RAND-D	.019 (.052) ; .062 (.103)	.028 (.048) ; .068 (.096)	.054 (.050) ; .110 (.094)	.030 (.055) ; .088 (.108)	.052 (.050) ; .092 (.093)	.042 (.044) ; .085 (.096)	.050 (.043) ; .106 (.091)
RAND-DD	.043 (.065) ; .092 (.122)	.072 (.084) ; .136 (.144)	.199 (.109) ; .263 (.173)	.500 (.380) ; .619 (.460)	.311 (.222) ; .423 (.306)	.301 (.242) ; .413 (.329)	.090 (.042) ; .135 (.093)
RAND-R	.024 (.058) ; .069 (.102)	.027 (.052) ; .070 (.101)	.058 (.049) ; .114 (.095)	.030 (.060) ; .076 (.106)	.046 (.056) ; .088 (.094)	.033 (.046) ; .074 (.093)	.052 (.044) ; .103 (.089)
RAND-RD	.043 (.065) ; .092 (.122)	.072 (.084) ; .136 (.144)	.199 (.109) ; .263 (.173)	.500 (.380) ; .619 (.460)	.311 (.222) ; .423 (.306)	.301 (.242) ; .413 (.329)	.090 (.042) ; .135 (.093)
*RAND-DI	.053 ; .113	.043 ; .107	.035 ; .108	.030 ; .115	.036 ; .106	.039 ; .110	.043 ; .096
*RAND-D	.029 ; .071	.018 ; .045	.013 ; .045	.004 ; .031	.014 ; .036	.014 ; .039	.039 ; .078
*RAND-DD	.037 ; .091	.023 ; .072	.018 ; .151	.025 ; .084	.031 ; .090	.033 ; .083	.052 ; .099
*RAND-R	.044 ; .096	.043 ; .087	.032 ; .075	.019 ; .060	.034 ; .077	.032 ; .070	.039 ; .076
*RAND-RD	.071 ; .135	.091 ; .180	.102 ; .213	.453 ; .569	.258 ; .373	.264 ; .385	.066 ; .105

Table 3: Empirical Rejection Rates – Configuration 1 for $I = 5$, $n = 10$ (20 in parentheses), $\alpha = 0.05$ left of semi-colon, $\alpha = 0.10$ right of semi-colon, KUR(6) Distribution. * Denotes Bootstrap Shuffling

Test	r=1	r=3	r=4	r=5	r=6	r=9	r=16
Levene	.032 (.037) ; .073 (.083)	.149 (.363) ; .244 (.482)	.242 (.555) ; .360 (.674)	.334 (.702) ; .459 (.799)	.413 (.800) ; .544 (.874)	.596 (.938) ; .716 (.966)	.805 (.993) ; .885 (.997)
ANOMV-LEV	.044 (.047) ; .090 (.092)	.196 (.416) ; .288 (.518)	.313 (.619) ; .421 (.710)	.422 (.761) ; .525 (.829)	.508 (.849) ; .615 (.899)	.700 (.958) ; .780 (.975)	.886 (.997) ; .929 (.998)
ANOMV-JK	.099 (.100) ; .174 (.170)	.202 (.329) ; .299 (.435)	.265 (.457) ; .376 (.564)	.331 (.564) ; .444 (.661)	.390 (.645) ; .502 (.733)	.520 (.790) ; .629 (.851)	.698 (.917) ; .780 (.948)
ANOMV-TR	.041 (.039) ; .085 (.082)	.167 (.369) ; .252 (.478)	.261 (.567) ; .364 (.666)	.352 (.710) ; .461 (.792)	.432 (.807) ; .542 (.873)	.607 (.941) ; .712 (.966)	.813 (.993) ; .877 (.997)
F-K	.038 (.038) ; .080 (.081)	.151 (.333) ; .238 (.456)	.232 (.517) ; .341 (.638)	.311 (.658) ; .427 (.763)	.380 (.759) ; .501 (.845)	.546 (.914) ; .663 (.953)	.749 (.987) ; .839 (.995)
RAND-DI	.053 (.058) ; .099 (.091)	.183 (.389) ; .274 (.486)	.281 (.569) ; .373 (.676)	.365 (.697) ; .469 (.787)	.438 (.808) ; .551 (.850)	0.605 (.903) ; .694 (.940)	.800 (.992) ; .864 (.997)
RAND-D	.059 (.046) ; .102 (.091)	.154 (.320) ; .245 (.428)	.231 (.502) ; .333 (.610)	.302 (.632) ; .410 (.723)	.369 (.746) ; .479 (.830)	0.533 (.878) ; .632 (.915)	.724 (.983) ; .810 (.992)
RAND-DD	.114 (.072) ; .177 (.120)	.256 (.384) ; .353 (.500)	.353 (.577) ; .474 (.668)	.455 (.703) ; .555 (.791)	.522 (.798) ; .632 (.868)	0.681 (.911) ; .767 (.950)	.859 (.994) ; .911 (.996)
RAND-R	.059 (.047) ; .107 (.090)	.166 (.323) ; .251 (.439)	.244 (.511) ; .357 (.620)	.328 (.653) ; .437 (.730)	.395 (.752) ; .516 (.837)	0.579 (.887) ; .677 (.921)	.785 (.988) ; .860 (.994)
RAND-RD	.114 (.072) ; .177 (.120)	.256 (.384) ; .353 (.500)	.353 (.577) ; .474 (.668)	.455 (.703) ; .555 (.791)	.522 (.798) ; .632 (.868)	0.681 (.911) ; .767 (.950)	.859 (.994) ; .911 (.996)
*RAND-DI	.042 ; .103	.160 ; .286	.235 ; .392	.313 ; .470	.372 ; .547	.494 ; .701	.654 ; .854
*RAND-D	.018 ; .047	.087 ; .167	.148 ; .240	.206 ; .320	.251 ; .377	.354 ; .498	.494 ; .657
*RAND-DD	.021 ; .055	.097 ; .177	.153 ; .250	.201 ; .316	.239 ; .377	.350 ; .498	.492 ; .638
*RAND-R	.039 ; .074	.131 ; .231	.216 ; .332	.287 ; .414	.353 ; .488	.513 ; .642	.728 ; .835
*RAND-RD	.068 ; .147	.214 ; .337	.307 ; .439	.384 ; .527	.458 ; .602	.613 ; .744	.814 ; .905

Table 4: Empirical Rejection Rates – Configuration 1 for $I = 10$, $n = 10$ (20 in parentheses), $\alpha = 0.05$ left of semi-colon, $\alpha = 0.10$ right of semi-colon, KUR(6) Distribution. * Denotes Bootstrap Shuffling

Test	r=1	r=3	r=4	r=5	r=6	r=9	r=16
Levene	.032 (.046) ; .071 (.093)	.145 (.352) ; .222 (.458)	.238 (.555) ; .337 (.653)	.336 (.707) ; .444 (.785)	.425 (.809) ; .531 (.869)	.622 (.947) ; .718 (.967)	.841 (.996) ; .893 (.998)
ANOMV-LEV	.074 (.070) ; .125 (.119)	.252 (.457) ; .326 (.539)	.383 (.669) ; .459 (.735)	.497 (.806) ; .573 (.852)	.591 (.887) ; .665 (.918)	.779 (.976) ; .828 (.984)	.929 (.999) ; .950 (.999)
ANOMV-JK	.128 (.137) ; .210 (.228)	.200 (.308) ; .293 (.415)	.251 (.426) ; .352 (.538)	.306 (.531) ; .411 (.639)	.356 (.621) ; .463 (.722)	.488 (.792) ; .595 (.862)	.684 (.938) ; .770 (.965)
ANOMV-TR	.042 (.043) ; .081 (.086)	.170 (.385) ; .244 (.476)	.278 (.590) ; .359 (.678)	.373 (.738) ; .466 (.806)	.462 (.835) ; .555 (.885)	.657 (.956) ; .732 (.971)	.852 (.996) ; .892 (.998)
F-K	.036 (.042) ; .074 (.088)	.123 (.303) ; .202 (.411)	.196 (.485) ; .296 (.591)	.274 (.628) ; .383 (.724)	.348 (.733) ; .455 (.816)	.509 (.905) ; .626 (.941)	.735 (.985) ; .815 (.993)
RAND-DI	.041 (.042) ; .086 (.103)	.147 (.296) ; .227 (.387)	.255 (.477) ; .347 (.578)	.356 (.621) ; .474 (.697)	.451 (.715) ; .544 (.788)	.636 (.883) ; .708 (.926)	.804 (.980) ; .885 (.988)
RAND-D	.028 (.048) ; .068 (.096)	.121 (.263) ; .188 (.352)	.214 (.426) ; .297 (.524)	.292 (.573) ; .391 (.658)	.383 (.677) ; .472 (.745)	.555 (.854) ; .647 (.893)	.759 (.972) ; .810 (.983)
RAND-DD	.072 (.084) ; .136 (.144)	.189 (.322) ; .295 (.427)	.299 (.484) ; .413 (.597)	.404 (.631) ; .506 (.720)	.493 (.728) ; .599 (.802)	.672 (.894) ; .758 (.927)	.890 (.985) ; .920 (.990)
RAND-R	.027 (.052) ; .070 (.101)	.117 (.266) ; .186 (.356)	.227 (.426) ; .302 (.528)	.301 (.582) ; .400 (.666)	.393 (.687) ; .496 (.755)	.575 (.860) ; .679 (.900)	.812 (.976) ; .855 (.987)
RAND-RD	.072 (.084) ; .136 (.144)	.189 (.322) ; .295 (.427)	.299 (.484) ; .413 (.597)	.404 (.631) ; .506 (.720)	.493 (.728) ; .599 (.802)	.672 (.894) ; .758 (.927)	.890 (.985) ; .920 (.990)
*RAND-DI	.043 ; .107	.143 ; .247	.237 ; .359	.328 ; .457	.414 ; .533	.573 ; .703	.767 ; .865
*RAND-D	.018 ; .045	.079 ; .144	.144 ; .238	.211 ; .329	.281 ; .414	.418 ; .573	.590 ; .767
*RAND-DD	.023 ; .072	.087 ; .187	.143 ; .292	.218 ; .391	.283 ; .469	.410 ; .642	.573 ; .847
*RAND-R	.043 ; .087	.122 ; .205	.194 ; .296	.281 ; .397	.363 ; .477	.531 ; .646	.759 ; .835
*RAND-RD	.091 ; .180	.191 ; .311	.272 ; .412	.371 ; .516	.456 ; .595	.636 ; .749	.839 ; .910

Table 5: Empirical Rejection Rates – Configuration 2 for $I = 5$, $n = 10$ (20 in parentheses), $\alpha = 0.05$ left of semi-colon, $\alpha = 0.10$ right of semi-colon, KUR(6) Distribution. * Denotes Bootstrap Shuffling

Test	$m = 2.5, r = 5$	$m = 3.5, r = 7$	$m = 12, r = 16$	$m = 18.75, r = 25$	$m = 36.75, r = 49$
Levene	.132 (.346); .230 (.497)	.168 (.465); .289 (.628)	.265 (.788); .455 (.916)	.333 (.891); .545 (.969)	.429 (.961); .663 (.993)
ANOMV-LEV	.150 (.330); .243 (.468)	.175 (.439); .286 (.607)	.250 (.855); .463 (.951)	.337 (.945); .577 (.986)	.463 (.986); .727 (.998)
ANOMV-JK	.242 (.398); .349 (.516)	.319 (.544); .435 (.652)	.663 (.900); .755 (.932)	.778 (.953); .849 (.969)	.891 (.985); .928 (.989)
ANOMV-TR	.144 (.356); .246 (.501)	.192 (.528); .324 (.676)	.487 (.967); .686 (.987)	.670 (.996); .831 (.999)	.875 (1.00); .956 (1.00)
F-K	.150 (.372); .252 (.521)	.206 (.536); .331 (.684)	.447 (.940); .636 (.979)	.595 (.989); .777 (.997)	.796 (.999); .918 (1.00)
RAND-D1	.133 (.257); .210 (.376)	.150 (.281); .231 (.396)	.114 (.172); .200 (.365)	.117 (.192); .210 (.404)	.120 (.229); .217 (.442)
RAND-D	.154 (.329); .251 (.456)	.210 (.468); .316 (.609)	.508 (.927); .632 (.962)	.623 (.982); .740 (.998)	.741 (.997); .825 (.999)
RAND-DD	.285 (.386); .401 (.527)	.374 (.579); .492 (.693)	.773 (.969); .852 (.986)	.895 (.995); .938 (.998)	.977 (.998); .989 (1.00)
RAND-R	.170 (.329); .263 (.463)	.242 (.494); .351 (.623)	.652 (.946); .769 (.974)	.819 (.995); .884 (.995)	.948 (.998); .978 (.999)
RAND-RD	.285 (.386); .401 (.527)	.374 (.579); .492 (.693)	.773 (.969); .852 (.986)	.895 (.995); .938 (.998)	.977 (.998); .989 (1.00)
*RAND-D1	.119; .229	.129; .242	.101; .220	.104; .226	.105; .233
*RAND-D	.062; .131	.072; .149	.082; .193	.100; .223	.120; .259
*RAND-DD	.072; .143	.082; .163	.098; .236	.119; .284	.145; .328
*RAND-R	.130; .251	.200; .342	.573; .723	.767; .864	.926; .966
*RAND-RD	.227; .369	.321; .480	.716; .826	.857; .919	.965; .986

Table 6: Empirical Rejection Rates – Configuration 2 for $I = 10$, $n = 10$ (20 in parentheses), $\alpha = 0.05$ left of semi-colon, $\alpha = 0.10$ right of semi-colon, KUR(6) Distribution. * Denotes Bootstrap Shuffling

Test	$m = 2.5, r = 5$	$m = 3.5, r = 7$	$m = 12, r = 16$	$m = 18.75, r = 25$	$m = 36.75, r = 49$
Levene	.102 (.263); .178 (.385)	.120 (.341); .211 (.478)	.155 (.558); .281 (.739)	.185 (.685); .336 (.843)	.230 (.815); .400 (.931)
ANOMV-LEV	.162 (.270); .233 (.377)	.172 (.317); .249 (.461)	.151 (.755); .278 (.904)	.180 (.907); .347 (.975)	.245 (.983); .480 (.998)
ANOMV-JK	.208 (.334); .316 (.461)	.268 (.476); .384 (.594)	.640 (.907); .735 (.936)	.773 (.961); .841 (.973)	.903 (.989); .936 (.991)
ANOMV-TR	.111 (.255); .181 (.376)	.130 (.380); .219 (.529)	.333 (.953); .530 (.983)	.522 (.995); .727 (.999)	.804 (1.00); .925 (1.00)
F-K	.106 (.275); .179 (.400)	.135 (.395); .231 (.532)	.268 (.823); .434 (.916)	.360 (.940); .549 (.980)	.514 (.997); .717 (1.00)
RAND-D1	.083 (.139); .146 (.234)	.086 (.146); .153 (.247)	.068 (.083); .121 (.158)	.068 (.084); .123 (.160)	.070 (.085); .125 (.164)
RAND-D	.096 (.234); .161 (.342)	.140 (.395); .237 (.514)	.455 (.925); .574 (.958)	.604 (.984); .718 (.989)	.736 (.998); .852 (1.00)
RAND-DD	.173 (.314); .291 (.430)	.258 (.487); .379 (.600)	.739 (.958); .839 (.974)	.897 (.991); .937 (.994)	.987 (.999); .997 (1.00)
RAND-R	.104 (.241); .171 (.345)	.163 (.407); .243 (.527)	.590 (.944); .725 (.963)	.800 (.988); .886 (.991)	.963 (.999); .979 (1.00)
RAND-RD	.173 (.314); .291 (.430)	.258 (.487); .379 (.600)	.739 (.958); .839 (.974)	.897 (.991); .937 (.994)	.987 (.999); .997 (1.00)
*RAND-D1	.082; .181	.088; .187	.066; .144	.067; .145	.067; .146
*RAND-D	.041; .084	.042; .092	.032; .103	.039; .123	.043; .153
*RAND-DD	.056; .123	.058; .131	.053; .168	.064; .205	.082; .249
*RAND-R	.101; .181	.146; .254	.569; .684	.761; .852	.948; .974
*RAND-RD	.188; .311	.264; .394	.712; .808	.860; .925	.978; .988

Table 7: Empirical Rejection Rates – Configuration 2 for $I = 5$, $n = 10$ (20 in parentheses), $\alpha = 0.05$, SKW(3) Distribution. * Denotes Bootstrap Shuffling. *** Denotes Test is Inappropriate for SKW(3).

Test	$m = 6.75, r = 9$	$m = 12, r = 16$	$m = 18.75, r = 25$	$m = 27, r = 36$	$m = 36.75, r = 49$
Levene	.259 (.664)	.397 (.789)	.487 (.846)	.551 (.874)	.596 (.887)
ANOMV-LEV	.302 (.704)	.471 (.819)	.574 (.864)	.636 (.890)	.675 (.904)
ANOMV-JK	***	***	***	***	***
ANOMV-TR	.443 (.955)	.683 (.994)	.824 (.999)	.900 (.999)	.942 (1.00)
F-K	.379 (.926)	.590 (.987)	.727 (.997)	.824 (.999)	.881 (.999)
RAND-D1	.357 (.366)	.415 (.407)	.437 (.436)	.456 (.461)	.463 (.469)
RAND-D	.734 (.874)	.826 (.919)	.860 (.953)	.888 (.979)	.904 (.987)
RAND-DD	***	***	***	***	***
RAND-R	.833 (.880)	.903 (.928)	.934 (.962)	.954 (.980)	.970 (.990)
RAND-RD	***	***	***	***	***
*RAND-D1	.312 (.270)	.367 (.304)	.387 (.321)	.402 (.330)	.411 (.335)
*RAND-D	.370 (.363)	.429 (.411)	.455 (.433)	.471 (.451)	.486 (.463)
*RAND-DD	.381 (.373)	.439 (.420)	.469 (.442)	.484 (.458)	.497 (.471)
*RAND-R	.758 (.843)	.877 (.914)	.920 (.946)	.942 (.964)	.955 (.978)
*RAND-RD	.807 (.852)	.890 (.920)	.929 (.952)	.948 (.965)	.960 (.980)

Table 8: Empirical Rejection Rates – Configuration 2 for $I = 10$, $n = 10$ (20 in parentheses), $\alpha = 0.05$, SKW(3) Distribution. * Denotes Bootstrap Shuffling. *** Denotes Test is Inappropriate for SKW(3).

Test	$m = 6.75, r = 9$	$m = 12, r = 16$	$m = 18.75, r = 25$	$m = 27, r = 36$	$m = 36.75, r = 49$
Levene	.098 (.427)	.158 (.591)	.208 (.676)	.248 (.728)	.278 (.761)
ANOMV-LEV	.241 (.604)	.356 (.760)	.444 (.830)	.507 (.866)	.553 (.889)
ANOMV-JK	***	***	***	***	***
ANOMV-TR	.252 (.939)	.550 (.993)	.753 (.999)	.860 (.999)	.924 (1.00)
F-K	.155 (.807)	.270 (.948)	.368 (.985)	.459 (.997)	.535 (.999)
RAND-D1	.152 (.113)	.171 (.125)	.182 (.132)	.194 (.136)	.200 (.138)
RAND-D	.722 (.870)	.848 (.931)	.894 (.963)	.919 (.978)	.930 (.986)
RAND-DD	***	***	***	***	***
RAND-R	.837 (.877)	.908 (.933)	.944 (.967)	.955 (.977)	.962 (.987)
RAND-RD	***	***	***	***	***
*RAND-D1	.118 (.090)	.144 (.103)	.158 (.108)	.164 (.110)	.167 (.112)
*RAND-D	.207 (.264)	.262 (.308)	.287 (.330)	.300 (.345)	.315 (.352)
*RAND-DD	.240 (.278)	.288 (.322)	.319 (.348)	.333 (.369)	.344 (.375)
*RAND-R	.726 (.843)	.879 (.912)	.916 (.955)	.946 (.973)	.961 (.983)
*RAND-RD	.828 (.853)	.908 (.919)	.940 (.958)	.961 (.974)	.974 (.985)

Table 9: Empirical Rejection Rates – Configuration 2 for $m = 27, r = 36, I = 5, n = 10$ (20 in parentheses), $\alpha = 0.05$. * Denotes Bootstrap Shuffling. *** Denotes Test is Inappropriate for Distribution.

Test	N(0,1)	chi-sq(1)	exp(1)	Gamma(4/9,1)	50:50 Mixture
Levene	.703 (1.00)	.192 (.504)	.318 (.845)	.319 (.840)	.789 (1.00)
ANOMV-LEV	.823 (1.00)	.178 (.526)	.307 (.914)	.316 (.914)	.910 (1.00)
ANOMV-JK	.993 (1.00)	***	***	***	.998 (1.00)
ANOMV-TR	.901 (1.00)	***	.802 (.999)	.796 (.999)	.921 (1.00)
F-K	.808 (1.00)	***	.752 (.996)	.750 (.998)	.831 (1.00)
RAND-D1	.206 (.890)	.069 (.122)	.065 (.122)	.060 (.094)	.885 (1.00)
RAND-D	.912 (1.00)	.139 (.630)	.186 (.869)	.171 (.839)	.993 (1.00)
RAND-DD	.997 (1.00)	***	***	***	1.00 (1.00)
RAND-R	.997 (1.00)	.465 (.887)	.826 (.997)	.840 (.995)	1.00 (1.00)
RAND-RD	.997 (1.00)	***	***	***	1.00 (1.00)
*RAND-D1	.213 (.858)	.036 (.051)	.040 (.066)	.045 (.075)	.882 (1.00)
*RAND-D	.535 (.997)	.014 (.036)	.017 (.118)	.021 (.107)	.990 (1.00)
*RAND-DD	.590 (.997)	.090 (.183)	.166 (.454)	.177 (.446)	1.00 (1.00)
*RAND-R	.998 (1.00)	.355 (.854)	.749 (.988)	.760 (.990)	1.00 (1.00)
*RAND-RD	.999 (1.00)	***	***	***	1.00 (1.00)

Table 10: Empirical Type I Error Rates – Unequal Means, $I = 5, n = 10$ (20 in parentheses), $\alpha = 0.05$, Mean Configuration = (1,1.2,1.4,1.6,1.8), Permutation Shuffling

Test	Distribution		
	N(0,1)	KUR(6)	SKW(3)
RAND-D1	.040 (.040)	.050 (.050)	.045 (.045)
RAND-D	.032 (.027)	.037 (.035)	.043 (.041)
RAND-DD	.082 (.059)	.112 (.069)	.161 (.121)
RAND-R	.059 (.043)	.086 (.070)	.050 (.054)
RAND-RD	.082 (.059)	.112 (.069)	.161 (.121)

Table 11: Empirical Rejection Rates – Unequal Means, $I = 5, n = 10$ (20 in parentheses), $\alpha = 0.05$, Mean Configuration = (1,1.2,1.4,1.6,1.8), KUR(6) Distribution, Permutation Shuffling

Test	Configuration		
	1, r=9	1, r=16	2, m=18.75 r=25
RAND-D1	.580 (.895)	.777 (.990)	.118 (.202)
RAND-D	.508 (.872)	.704 (.979)	.609 (.972)
RAND-DD	.681 (.911)	.859 (.994)	.891 (.995)
RAND-R	.604 (.901)	.806 (.989)	.823 (.994)
RAND-RD	.681 (.911)	.859 (.994)	.891 (.995)