DOCUMENT RESUME

ED 283 874 TM 870 413

AUTHOR Olejnik, Stephen; Algina, James
TITLE Bootstrap Estimation and Testing for Variance Equality.
PUB DATE Apr 87
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Estimation (Mathematics); Measurement Techniques; *Monte Carlo Methods; *Population Distribution; Sample Size; *Sampling; *Statistical Distributions; *Statistical Significance; Statistical Studies
IDENTIFIERS *Bootstrap Methods; Chi Square Test; F Ratio; Population Validity; Statistical Analysis System; Type I Errors; *Variance (Statistical)

ABSTRACT

The purpose of this study was to develop a single procedure for comparing population variances which could be used for distribution forms. Bootstrap methodology was used to estimate the variability of the sample variance statistic when the population distribution was normal, platykurtic and leptokurtic. The data for the study were generated and analyzed using the Statistical Analysis System computing package. The bootstrap estimates of variability underestimated the theoretical variance value, and the mean square error of the estimated variance was small for both the normal and platykurtic distributions, but large for the leptokurtic distribution. The F-ratio and chi-square test statistics were computed for comparing the variability of two populations using the bootstrap estimates of variance. Observed Type I error rates within two standard errors of the nominal significance level were obtained only when the population distribution was platykurtic. The study concluded that in the case of sample variance the bootstrap methodology could provide some indication of sample accuracy but the degree of accuracy would depend on the distribution form. (JAZ)

Bootstrap Estimation and Testing for Variance Equality

Stephen Olejnik

University of Georgia

James Algina

University of Florida

Paper presented at the meetings of the American Educational Research Association, Washington D.C. April, 1987

ABSTRACT

Bootstrap methodology was used to estimate the variability of the sample variance statistic when the population distribution was normal, platykurtic and leptokurtic. The bootstrap estimates of variability underestimated the theoretical variance value and the mean square error of the estimated variance was small for both the normal and platykurtic distributions but large for the leptokurtic distribution. The F-ratio and chi-square test statistics were computed for comparing the variability of two populations using the bootstrap estimates of variance. Observered Type I error rates within two standard errors of the nominal significance level were obtained only when the population distribution was platykurtic.

Bootstrap Estimation and Testing for Variance Equality

Developing statistical procedures for comparing two or more population variances has been an interest among statisticians for many years. The first procedures developed (e.g. Bartlett, 1937; Cochran, 1941; Hartley, 1950) were based on the likelihood ratio test. These procedures all assume that the sampled population distributions are normal. When the distributions are non-normal the actual Type I error rate can be greater than or less than the nominal significance level depending on whether the distributions are platykurtic or leptokurtic respectively (Box, 1953). Since in practice the population distribution form is generally unknown and often non-normal, these approaches are of limited value. Procedures were then developed which attempted to modify the likelihood ratio tests by adjusting for the kurtosis of the distribution (e.g Box, 1953; Scheffe, 1959; Layard, 1973). These procedures estimate the population kurtosis using sample data. But kurtosis estimated on a sample is not a very stable statistic and these tests typically have actual Type I error rates which deviate considerably from the nominal significance level. A third approach taken for testing variance equality has been the use of nonparametric rank tests of scale (eg. Mood, 1954; Siegel-Tukey, 1960; Klotz, 1962). The nonparametric tests however are extremely sensitive to differences in the location parameter (Moses,

1963). As differences in the population medians increase the procedures become increasingly conservative and statistical power decreases. Adjusting for differences in sample medians or sample means has been suggested but these procedures do not provide valid tests of scale equality when the distributions sampled are asymmetric (Conover, Johnson and Johnson, 1981; Olejnik and Algina, 1987).

Still another category of tests for variance equality are procedures which transform the original data to a measure which reflects variability rather than location and compares the mean transformed scores using analysis of variance (eg. Miller, 1968; Brown and Forsythe, 1974; O'Brien, 1978). Empirical studies of these approaches have had mixed results. Conover, Johnson and Johnson (1981) compared 59 procedures (excluding O'Brien's) from the four categories of approaches and could only recommend the Brown-Forsythe for non-normal distributions. Additional studies have shown both the Brown-Forsthe and the O'Brien procedures as valid tests with reasonable statistical power (O'Brien, 1978; Olejnik and Algina, 1987). Neither procedure is uniformly superior to the other for all distribution forms however. O'Brien's procedure is more powerful with normal and platykurtic distributions and the Brown-Forstyhe approach is more sensitive to variance differences when the distributions are leptokurtic. Although both of these procedures provide valid tests for variance equality it would be convenient if a single

procedure could be developed that could be used for all distribution forms with equal statistical power.

A new approach that might be suggested to test for variance equality is an application of the bootstrap methodology. This resampling procedure estimates the standard error of a statistic of interest by empirically developing the sampling distribution of the statistic through Monte Carlo simulation procedures. Beginning with a sample of size n units randomly selected from a population, multiple bootstrap estimates of the statistic of interest are calculated. These estimates are based on subsamples of n units created by resampling with replacement from the original n units. The standard deviation of the distribution of bootstrap estimates provides a measure of accuracy of the statistic. Although it is not clear how many bootstrap estimates are needed to provide a good estimate of accuracy, Efron and Tibshirani (1986) has suggested that as few as 25 subsamples may be sufficient for some situations and 200 subsamples should be sufficient for most estimators. The approach has been used with success in several contexts (Lunneborg and Tousignant, 1985; Efron and Tibshirani, 1986). Given that the standard error of the sample variance can be estimated accurately it could then be used to calculate an F or a chi-square test statistic. Since the standard error is determined empirically from the sample data, the approach could be used for all population distribution forms.

The present study had two objectives: First to apply the bootstrap methodology to estimate the accuracy of the sample variance and to compare this estimate with the theoretical measure of accuracy derived by Box (1953). Second, use the empirical estimate of the sample variance standard error to compute a test statistic and develop the sampling distribution of that test statistic to determine the Type I error rate under a true null condition and the statistical power when population variances differ. Finally the Type I error rate and the statistical power of this approach is compared to similar results obtained using O'Brien's test.

Bootstrap estimate of accuracy

Method. The adequacy of the bootstrap methodology to estimate the variance of the unbiased sample variance was studied under several conditions. Three factors were manipulated: sample size (n), distribution form and the number of bootstrap subsamples (B). Four sample sizes were considered from three distributional forms. Samples of size 9, 15, 25 and 30 were selected from distributions which were normal (0,0), platykurtic (0,-1) or leptokurtic (0,3.75). Since it is not clear how many bootstrap samples are needed, three levels of this factor were considered: 50, 100 and 200. Thus this part of the study considered 36 conditions in a completely crossed 4x3x3 factorial design. Each of these conditions were replicated 200 times to determine: the average bootstrap estimate of variance ($\bar{S}^2$), the average estimate of variability for the sample variance ($\overline{Var}(S^2)$),

and the average mean square error of the bootstrap estimate of variability (MSE). The mean square error was calculated as the average squared difference between each replication estimate of variability and the theoretical measure of variance provided by Box (1953). The variance of the unbiased sample estimate of the population variance is calculated as the following:

$$Var(S^2) = \sigma^4 \left[\frac{2}{n-1} + \frac{k}{n} \right]$$

where σ^4 is the squared population variance;

n is the sample size;

and k is the kurtosis of the population distribution.

Data Generation. The data for the study were generated and analyzed using the Statistical Analysis System (SAS, 1984) computing package. Observations were generated having a mean of 10 and variance equal to 1 using RANNOR, the normal random number generating function in SAS. The non-normal distributions, were generated by transforming the normal random variables using a polynomial power procedure suggested by Fleishman (1978): $W = [(dx+c)x+b]x+a$. Where x is the normally distributed random variable and a, b, c, and d are constants which modify the skew and kurtosis of the distribution leaving the mean and variance unchanged.

Procedure. For each condition studied a random sample of n observations were generated. Bootstrap subsamples were then created by resampling with replacement from the

original sample of n observations with each bootstap subsample consisting of n scores. The unbiased estimate of the sample variance of each bootstrap subsample was then calculated ($S^2_{b1} = \Sigma(X_j - X_.)^2/(n-1)$) creating a sampling distribution of the sample variance based on B estimates of the sample variance. Finally the mean and variance of the sampling distribution was computed along with the estimated squared error (the difference between the estimated sample variance and the theoretical value for the variance of the sample variance). This procedure was replicated 200 times and the average mean variance ($\bar{S}^2$), the average variance across the bootstrap samples ($\overline{var}(S^2)$) and the mean squared error ($\overline{MSE}$) were computed. In pilot testing the program it was noted that over the 200 replications the average mean variance consistently underestimated the population variance which was equal to 1. The underestimation was greatest when the sample size was small. Before generating the sampling distributions of the sample variances, the program was modified to include two additional methods of calculating the variance of the bootstrap subsamples. Method 2 divided the sum of squared deviations around the subsample mean by n-2 rather than n-1, ($S^2_{b2} = \Sigma(X_i - X_.)^2/(n-2)$). Method 3 calculated the sum of squared deviations of the bootstrap subsample observations around the mean of the original n observations from which the bootstrap subsamples were generated. The sum of squares were then divided by n-1,

$(S^2_{b3} = \Sigma (X_i - M_.)^2)/(n-1)$). In pilot testing the revised program the number of replications per condition was increased to 500. The results for 200 and 500 replication were not noticably different so the sampling distributions were generated using 200 replications.

Results. Table 1 presents the results using the three methods of calculating the bootstrap subsample variance, for sample sizes of 9, 15, 25 and 30 using 50, 100 or 200 bootstrap subsamples when the population distribution sampled was normal. Tables II and III report similar results when the sampled distributions were platykurtic and leptokurtic respectively. For all four sample sizes and all

Insert tables I, II and III here

three methods of calculating the subsample variance the results were similar when the number of bootstrap subsamples were 50, 100 or 200. Using method 1 for calculating the unbiased estimate of sample variance, the average variance ($\overline{S^2}$) consistently underestimated the population variance across all sample sizes and across all sizes of bootstrap subsamples. The average variance ($\overline{var}(S^2)$) of the sample variances also underestimated the theoretical value of the sample variance. For the platykurtic distribution the difference was not great except for a sample size of 9. The greatest difference was found with the leptokurtic

distribution and a moderate difference was observed for the normal distribution.

Using method 2 which used n-2 rather than n-1 as the denominator for the calculation of the bootstrap subsample variance, the results were much closer to the theoretical values. The lowest mean variance estimate equalled .944 and the largest mean variance estimate equalled 1.075. The average variance deviated most from the theoretical value when the original sample had only 9 observations. As sample size increased the average variance approached the theoretical value for the sample variance.

Method 3 which used the original sample mean in calculating the subsample variance also had results similar to the theoretical values for the sample variance and the mean variance. The smallest mean variance was calculated as .965 and the largest mean variance equalled 1.094. Only with a sample of size 9 did the mean variance deviate greatly from the theoretical value for the variance.

Examining the mean square errors for all three methods indicate considerable variability between the estimated variance of the sample variance and the theoretical value. The errors were greatest for the leptokurtic distribution and smallest for the platykurtic distribution. For the normal distribution the errors appeared small when the sample size was at least 15. These results indicate that estimating the standard error of the sample variance may be adequate for hypothesis testing only when sample sizes are

at least 15 and the population distribution is normal or platykurtic. With a leptokurtic distribution the average variance for all three methods seriously underestimated the theoretical value of variance. More importantly the mean square error was large for the three methods of calculating subsample variance for all sample sizes. How important these differences are for hypothesis testing for variance equality is not clear. However one might expect that using the bootstrap methodology to estimate the standard error for a sample variance would result in a liberal test for variance equality if the sampled distribution was leptokurtic. With a platykurtic population the test may be valid for even small sample sizes. When the population distribution is normal the bootstrap procedure might be valid when sample sizes are at least 15. The second part of the study investigated the accuracy of these predictions.

Testing for variance equality

Method. Two classical procedures for comparing population parameters using sample statistics were considered. The sample variances were compared using an F-ratio and a chi-square test statistic. Although both procedures could be used to compare the variances of several populations, the present study was limited to a comparison of two populations. An F-ratio can be computed as the ratio of the squared difference between two sample statistics and the sum of the variances of the two statistics. In the case

of the sample variance the F-ratio is computed as the following:

$$F = \frac{(S_1^2 - S_2^2)^2}{Var(S_1^2) + Var(S_2^2)}$$

The test statistic is referred to an F distribution with 1 and n_1+n_2-2 degrees of freedom.

A chi-square test statistic can be constructed by summing the ratio of the squared difference between the sample statistic and the mean sample statistic across all comparison groups and the variance of each sample statistic. In the case of the sample variance the chi-square statistic can be calculated as the following:

$$X^2 = \sum \frac{(S_i^2 - S_.^2)^2}{Var(S_i^2)}$$

The test statistic is referred to a chi-square distribution with 1 degree of freedom.

For both statistics the variances were estimated using the bootstrap methodolgy. Both of the procedures considered assume that the sampling distribution of the sample statistic is normal. Since the sampling distribution of a sample variance is positively skewed, it was necessary to transform the sample statistic before computing the test statistic. Two transformations were considered. The sample variance was transformed by taking the log of the sample statistic. This is the approach taken in Bartlett's test. The second transformation considered was recently suggested by Hawkins and Wixley (1986) using the fourth root of the

chi-squared variable. It was hoped that an better approximation to normality could improve the Type I error rate and statistical power of the tests.

A third test for variance equality was considered in the study which did not use the bootstrap methodology. O'Brien's (1978) procedure for comparing variances was included for two reasons: first since O'Brien's test is currently one of the best procedures for comparing population variances it provides a useful index for comparing observed Type I error rates when the null hypothesis is true and second it provides a standard to evaluate the statistical power of the procedures using the bootstrap methodology when the population variances differed. O'Brien suggested that a test for variance equality could be developed by transforming the sample data using:

$$r_{ij} = \frac{(w+n_j-2)n_j(x_{ij}-x_{.j})^2 - ws_j(n_j-1)}{(n_j-1)(n_j-2)}$$

where s_j is the within group unbiased estimate of variance for sample j and w is a weighting factor. O'Brien (1981) recommends setting w=.5 for most situations. The transformed observations are then used as the outcome measure in calculating the ANOVA F-ratio.

Data generation. The data for this part of the study were generated using the Rannor random generating function in the SAS computing package. Two factors were manipulated:

sample size and distribution form. Since estimates of sample variance variability in the first part of the study were shown to be somewhat accurate when sample sizes were at least 15, three sample size combinations were included in comparisons of variances: (15,15), (25,25) and (30,30). Since neither O'Brien's test nor the bootstrap procedures are affected by sample size inequality under the null condition, only equal sample sizes were considered. Three distributional forms were studied: normal (0,0), platykurtic (0,-1) and leptokurtic (0,3.75). The distributional forms were generated using the same methodology as described in part 1 of the investigation.

Finally, since the estimated stability of the sample variances in part 1 was not greatly affected by the number of bootstrap subsamples used, the estimated variances of the sample variance was based on 100 bootstrap subsamples.

Procedure. For each condition studied a random sample of n observations were generated for each of two independent groups. The sample variances were then compared using O'Brien's test. Then for each group the variance of the sample variance was estimated separately using the bootstrap methodology. For each bootstrap subsample the subsample variance was computed using the three methods described earlier. Each of these estimates was transformed using the log transformation and the 4th root transformation before the variance of the sampling distribution was calculated. Before computing the F-ratio and chi-square statistics, the

sample variances of the original n observations were also transformed. Thus for each condition studied 13 test statistics were computed: O'Brien's, the six tests using the F-ratio (3 using the log transformation and 3 using the 4th root transformation) and six tests using the chi-square statistic. This procedure was replicated 1000 times to develop the sampling distribution of the 13 test statistics for each of the 9 conditions under investigation. The proportion of times the null hypothesis was rejected at the .01, .05 and .10 levels were recorded.

Results. The empirically determined Type I error rates for the 13 test statistics are reported in tables IV, V and VI for the .01, .05 and .10 levels of significance respectively. O'Brien's test statistic had observed Type I

Insert Tables IV, V, and VI here

---

error rates within two standard errors of the nominal significance levels for all conditions studied except the condition where the sampled distribution was normal and sample size was 15 and when the distribution sampled was leptokurtic and the sample size was 25. For these conditions the observed Type I error rate underestimated the nominal significance levels at the .10 and .05 with normal distribution and at the .05 and .01 with the leptokurtic distribution. These results are consistent with previous

research findings and support the validity of the data generation procedures used in the study.

No clear pattern of results were obtained for the tests which used the bootstrap methodology. The results for the test statistics based on the bootstrap estimates of variability for the sample variance had mixed results depending on the distributional form, the test statistic, the type of transformation used, the sample size and the method of computing the subsample variance. In general the bootstrap test statistics had observered Type I error rates which overestimated the nominal significance level when the distributions were normal or leptokurtic. With the platykurtic distribution the observed Type I error rates were within two standard errors of the nominal significance level for the tests using the F-ratio except for method 2 with the 4th root transformation when the nominal significance level was underestimated. With the chi-square test the results were mixed. Method 1 had appropriate error rates when sample sizes were at least 25 per group for both the log and 4th root transformations. Method 3 overestimated the .10 and .05 nominal significance levels but underestimated the .01 nominal significance level.

The three methods for computing subsample variance provided similar conclusions. When the log transformation was used method 1 and 2 had identical results. With the 4th root transformation method 2 provided for an underestimation of the nominal significance level when the F-ratio was

calculated. With the chi-square test and method 2 the nominal significance level was underestimated for the normal and platykurtic disturbutions and overestimated for the leptokurtic distribution.

Finally the 4th root transformation did not improve the approximations to the reference distributions for either test statistic. Using the 4th root transformation to normalize the sampling distribution of the sample variance resulted in findings similar to those reported under the log transformation.

Statistical Power. The above results indicate that with the execption of the platykurtic distributions the procedures considered in the present study using the bootstrap methodology do not provide valid tests for variance equality. Given these result attempts to estimate statistical power would not be meaningful and no further analyses were conducted.

Summary and Discussion

Developing a single procedure for comparing population variances which could be used for all distribution forms was the major objective of the present study. It was hoped that by using bootstrap methodology the sampling variability of the sample variance could be estimated accurately and an F-test or chi-square test could be computed to compare estimates of population variances. The initial results of the study indicated that the unbiased estimator of sample variance computed on the bootstrap subsamples underestimated

the population variance. But two modifications for computing subsample variance were both successful in providing more accurate estimates of the population variance.

Estimating the variability of the sample variance was less successful. The average variability over 200 replications resulted in an underestimation of the theoretical variance when the sampled distribution was normal or leptokurtic. With the platykurtic distribution the estimated variance was similar to the theoretical value. These results were consistent across all three methods of calculating subsample variance, although the unbiased estimator deviated the most from the theoretical value. As sample size increased the bootstrap estimate of accuracy also improved. The mean square errors however were large for the leptokurtic distribution for all sample sizes indicating considerable variability from replication to replication in the estimation of the variability of the sample variance with that distribution. The mean square errors for the normal and platykurtic distributions were not as great and provided some hope that a valid test could be developed using the bootstrap methodology. By comparison, Efron (1983) had reported mean square errors for the bootstrap estimate of accruacy for a correlation coefficient of the same magnitude as those reported here for the normal distribution.

Since the bootstrap estimate of variability for the sample variance generally underestimated the theoretical

value it was predicted that the F-test and chi-square tests would have Type I error rates greater than the nominal significance level when the distributions sampled were normal or leptokurtic. The extent to which the observed Type I error rates overestimated the nominal significance level was unknown however. The predictions were correct. The extent of the overestimation depended on the distribution form, the method of transformation and the test statistic used. Observed Type I error rates were observed similar to the nominal significance level however when the distribution was platykurtic.

The results of this study indicate that in the case of the sample variance the bootstrap methodology can provide some indication of sample accruacy but the degree of accuracy depends on the distribution form. If the researcher has no idea of the population distribution form then it would be impossible to interpret a bootstrap estimate. In addition except for the platykurtic distribution the tests for statistical significance using the bootstrap estimate of variability resulted in a liberal test. Since O'Brien's test for variance equality does not overestimate the Type I error rate for any distribution it appears that the bootstrap approach does not provide a useful alternative.

References

Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Society, A901, 160, 268-282.

Box, G.E.P. (1953). Non-normality and tests on variance. Biometrika. 40, 318-335.

Brown, M.B. & Forsythe, A.B. (1974). Robust tests for the equality of variances. Journal of the American Statistical Association. 83, 364-367.

Cochran, W.G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. Annals of Eugenics, 11, 47-52.

Conover, W.J., Johnson, M.E. & Johnson, M.M. (1981). A comparative study of tests for homogeneity of vriances, with applications to the outer continental self bidding data. Technometrics. 23, 351-361.

Efron, B. & Tibscirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. Social Science 1, 54-75.

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

Hartley, H.O. (1950). The maximum F-ratio as a short-cut test for heterogeneity of variance. Biometrika, 37, 308-312.

Hawkins D.M. & Wixley R.A.J. (1986). A note on the transformation of chi-squared variables to normality. The American Statistician, 40, 296-298.

Klotz, J. (1962). Nonparametric tests for scale. Annals of Mathematical Statistics. 33, 495-512.

Layard, M.W.J. (1973). Robust large-sample tests of homogeneity of variances. Journal of the American Statistical Association, 68, 195-198.

Lunneborg, C.E. & Tousignant, J.P. (1985). Efron's bootstrap with application to the repeated measures design. Multivariate Behavioral Research, 20, 161-178.

Martin, C.G. (1975). Comment on Levy's "An empirical comparison of the Z-variance and Box-Scheffe tests for homogeneity of variance". Psychometrika, 41,551-556.

Miller, R.G. (1968). Jacknifing variances. Annals of Mathematical Statistics, 39, 567-582.

Mood, A.M. (1954). On the asymptotic efficiency of certain nonparametric two sample tests. The Annals of Mathematical Statistics, 25, 514-522.

Moses, L.E. (1963). Rank tests of dispersion. Annals of Mathematical Statistics, 34, 973-983.

O'Brien, R.G. (1978). Robust techniques for testing heterogeneity of variance effects in factorial designs. Psychometrika. 43, 327-342.

O'Brien, R.G. (1981). A simple test for variance effects in experimental designs. Psychological Bulletin. 89, 570-574.

Olejnik, S. & Algina, J. (1987). Type I error rates and power estimates of selected parametric and nonparametric tests of scale. Journal of Educational Statistics. (in press).

Scheffe, H. (1959). The Analysis of Variance. New York: John Wiley & Sons Inc.

Siegel, S. & Tukey, J.W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. American Statistical Association Journal. 55, 429-445.

Table I

Empirical estimates of mean, variance and mean square error for three bootstrap estimates of sample variance over 200 replications when the population distribution is normal.

		Method 1			Method 2			Method 3		
n	B	$\overline{S^2}$	$\overline{Var(S^2)}$	$\overline{MSE}$	$\overline{S^2}$	$\overline{Var(S^2)}$	$\overline{MSE}$	$\overline{S^2}$	$\overline{Var(S^2)}$	$\overline{MSE}$
9	50	.886	.179	.055	.997	.190	.066	1.013	.234	.085
	100	.849	.180	.054	.956	.196	.069	.971	.235	.084
	200	.867	.176	.049	.976	.190	.061	.991	.231	.075
heory		1.000	.25		1.000	.25		1.000	.25	
5	50	.933	.123	.020	1.000	.130	.023	1.000	.142	.026
	100	.919	.109	.009	.986	.116	.010	.990	.126	.010
	200	.929	.123	.015	.995	.130	.017	1.000	.143	.020
heory		1.000	.143		1.000	.143		1.000	.143	
5	50	.984	.083	.006	1.023	.087	.008	1.026	.090	.007
	100	.925	.066	.002	.963	.068	.002	.965	.072	.003
	200	.982	.078	.004	1.023	.081	.004	1.025	.084	.005
heory		1.000	.093		1.000	.093		1.000	.093	
0	50	1.004	.071	.003	1.038	.074	.003	1.040	.077	.034
	100	.975	.061	.002	1.009	.063	.002	1.009	.065	.002
	200	.973	.062	.002	1.006	.064	.002	1.008	.066	.002
heory		1.000	.069		1.000	.069		1.000	.069	

Table II

Empirical estimates of mean, variance and mean square error for three bootstrap estimates of sample variance over 200 replications when the population distribution is platykurtic.

		Method 1			Method 2			Method 3		
n	B	$\overline{S}^2$	$\overline{Var(S^2)}$	$\overline{MSE}$	$\overline{S}^2$	$\overline{Var(S^2)}$	$\overline{MSE}$	$\overline{S}^2$	$\overline{Var(S^2)}$	$\overline{MSE}$
9	50	.902	.121	.006	1.016	.121	.008	1.031	.159	.011
	100	.912	.126	.007	1.026	.125	.008	1.040	.164	.012
	200	.900	.129	.007	1.012	.132	.008	1.029	.169	.012
Theory		1.000	.139		1.000	.139		1.000	.139	
15	50	.953	.069	.001	1.021	.068	.001	1.026	.079	.012
	100	.928	.070	.001	.995	.071	.001	.999	.081	.001
	200	.917	.070	.001	.983	.070	.001	.988	.081	.001
Theory		1.000	.073		1.000	.073		1.000	.073	
25	50	.962	.040	.000	1.000	.040	.000	1.000	.044	.000
	100	.969	.042	.000	1.010	.042	.000	1.010	.046	.000
	200	.960	.041	.000	.999	.041	.000	1.000	.044	.000
Theory		1.000	.043		1.000	.043		1.000	.043	
30	50	.967	.034	.000	.999	.034	.000	1.001	.036	.000
	100	.948	.032	.000	.979	.033	.000	.982	.035	.000
	200	.859	.034	.000	.993	.034	.000	.994	.037	.000
Theory		1.000	.036		1.000	.036		1.000	.036	

Table III

Empirical estimates of mean, variance and mean square error for three bootstrap estimates of sample variance over 200 replications when the population distribution is leptokurtic.

		Method 1			Method 2			Method 3		
n	B	$\bar{S}^2$	$\overline{Var}(S^2)$	$\overline{MSE}$	$\bar{S}^2$	$\overline{Var}(S^2)$	$\overline{MSE}$	$\bar{S}^2$	$\overline{Var}(S^2)$	$\overline{MSE}$
9	50	.925	.364	.967	1.040	.405	1.153	1.051	.475	1.531
	100	.957	.415	.748	1.075	.476	1.096	1.094	.542	1.183
	200	.880	.296	.592	.944	.338	.765	.960	.386	.854
Theory		1.000	.667		1.000	.667		1.000	.667	
15	50	.951	.325	1.276	1.027	.366	1.762	1.030	.377	1.710
	100	.997	.307	.487	1.069	.337	.601	1.074	.356	.647
	200	.924	.261	.257	.980	.288	.313	.995	.303	.330
Theory		1.000	.393		1.000	.393		1.000	.393	
25	50	.966	.183	.122	1.006	.199	.146	1.009	.200	.143
	100	.939	.184	.134	.978	.197	.193	.979	.200	.192
	200	.926	.177	.185	.965	.189	.220	.966	.192	.217
Theory		1.000	.233		1.000	.233		1.000	.233	
30	50	.972	.191	.159	1.006	.199	.173	1.006	.205	.183
	100	.960	.152	.076	.993	.163	.088	.995	.163	.087
	200	.988	.183	.177	1.022	.195	.221	1.023	.196	.204
Theory		1.000	.194		1.000	.194		1.000	.194	

Table IV

Observed Type I error rates for 13 tests of variance equality with sample sizes of 15, 25, and 30 when the nominal significance level is .01.

			Normal			Platykurtic			Leptokurtic		
Procedure			15	25	30	15	25	30	15	25	30
O'Brien			.006	.006	.008	.010	.009	.010	.006	.002	.009
F-ratio	Log	M_1[a]	.017	.020	.019	.013	.007	.007	.017	.025	.032
		M_2	.017	.020	.019	.013	.007	.007	.017	.025	.032
		M_3	.026	.023	.020	.016	.010	.010	.026	.027	.035
	4th root	M_1	.018	.021	.019	.014	.010	.010	.018	.022	.029
		M_2	.000	.001	.000	.000	.000	.000	.000	.000	.000
		M_3	.029	.025	.022	.019	.014	.014	.029	.025	.030
X^2	Log	M_1	.032	.034	.026	.020	.013	.014	.073	.051	.052
		M_2	.032	.034	.026	.020	.013	.014	.073	.051	.052
		M_3	.052	.037	.033	.034	.021	.019	.093	.059	.057
	4th root	M_1	.047	.036	.029	.023	.014	.014	.101	.081	.070
		M_2	.015	.008	.005	.007	.002	.002	.034	.035	.026
		M_3	.016	.009	.006	.008	.002	.003	.035	.036	.027

[a] M_1 bootstrap subsample variance calculated using the unbiased estimator of subsample variance,

M_2 bootstrap subsample variance calculated by dividing the sum of squares by n-2

M_3 bootstrap subsample variance calculated by subtracting sample mean rather than the subsample mean in computing the sum of squares.

Table V

Observed Type I error rates for 13 tests of variance equality with sample sizes of 15, 25 and 30 when the nominal significance level is .05.

Procedure			Normal			Platykurtic			Leptokurtic		
			15	25	30	15	25	30	15	25	30
O'Brien			.033	.051	.051	.056	.048	.042	.041	.028	.038
F-ratio	Log	M_1[a]	.092	.070	.073	.044	.043	.037	.075	.084	.086
		M_2	.092	.070	.073	.044	.043	.037	.075	.084	.086
		M_3	.124	.078	.083	.069	.054	.044	.095	.093	.095
	4th root	M_1	.100	.076	.082	.059	.054	.044	.080	.088	.085
		M_2	.010	.003	.001	.002	.001	.000	.003	.009	.008
		M_3	.114	.081	.085	.075	.059	.048	.100	.093	.095
X^2	Log	M_1	.097	.087	.085	.065	.059	.053	.153	.127	.127
		M_2	.097	.087	.085	.065	.059	.053	.153	.127	.127
		M_3	.121	.095	.098	.098	.072	.064	.179	.138	.136
	4th root	M_1	.115	.094	.094	.076	.064	.056	.186	.152	.157
		M_2	.041	.036	.027	.023	.013	.016	.090	.066	.065
		M_3	.132	.102	.104	.101	.075	.064	.205	.163	.168

[a] M_1 bootstrap subsample variance calculated using the unbiased estimator of subsample variance,
M_2 bootstrap subsample variance calculated by dividing the sum of squares by n-2,
M_3 bootstrap subsample variance calculated by subtracting sample mean rather than subsample mean in computing the sum of squares.

Table VI

Observed Type I error rates for 13 tests of variance equality with sample sizes of 15, 25 and 30 when the nominal significance level is .10.

Procedure			Normal			Platykurtic			Leptokurtic		
			15	25	30	15	25	30	15	25	30
O'Brien			.076	.102	.110	.101	.097	.095	.089	.081	.082
F-ratio	Log	M_1[a]	.160	.122	.121	.086	.088	.083	.122	.149	.172
		M_2	.160	.122	.121	.086	.088	.083	.122	.149	.172
		M_3	.190	.138	.131	.117	.100	.097	.144	.162	.180
	4th root	M_1	.169	.131	.133	.106	.101	093	.136	.148	.169
		M_2	.027	.011	.004	.003	.002	.002	.007	.020	.034
		M_3	.199	.148	.136	.130	.107	.103	.154	.156	.177
X^2	Log	M_1	.155	.143	.140	.109	.105	.099	.218	.195	.215
		M_2	.151	.143	.140	.109	.104	.099	.218	.195	.215
		M_3	.176	.166	.148	.152	.118	.111	.249	.210	.225
	4th root	M_1	.172	.164	.149	.126	.111	.106	.257	.221	.237
		M_2	.075	.070	.072	.049	.033	.034	.128	.103	.116
		M_3	.200	.173	.156	.155	.124	.113	.204	.226	.244

[a] M_1 bootstrap subsample variance calculated using the unbiased estimator of subsample variance,

M_2 bootstrap subsample variance calculated by dividing the sum of squares by n-2,

M_3 bootstrap subsample variance calculate by subtracting sample mean rather than subsample mean in computing the sum of squares.