DOCUMENT RESUME

ED 306 273 TM 013 107

AUTHOR	Luecht, Richard M.; Smith, Phillip L.
TITLE	The Effects of Bootstrapping Strategies on the Estimation of Variance Components.
PUB DATE	Mar 89
NOTE	29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Analysis of Variance; Educational Research; *Error of Measurement; *Estimation (Mathematics); Factor Analysis; *Monte Carlo Methods; Nonparametric Statistics; *Sampling
IDENTIFIERS	Bootstrap Hypothesis; *Confidence Intervals (Statistics); *Parametric Analysis

ABSTRACT

Two bootstrapping or resampling strategies were investigated to determine their applicability to estimating standard errors and ensuing confidence intervals on variance components in two-factor random analysis of variance models. In light of prior negative findings regarding the application of bootstrapping to this particular problem, a recommendation of an "optimal" approach to resampling was sought. The study used Monte Carlo simulations to test the variance component estimation accuracy under simultaneous resampling of all effect factors in a random model versus resampling a single factor. Results indicate that single-factor was a preferable method of resampling and produced reasonable estimates of both standard errors and confidence intervals (parametric and non-parametric). Suggestions for the appropriate application of the technique to educational measurement are discussed. (Author/TJH)

The Effects of Bootstrapping Strategies on the Estimation of Variance Components

Richard M. Luecht
Phillip L. Smith

University of Wisconsin-Milwaukee

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1989.

Abstract

Two bootstrapping or resampling strategies were investigated, as applicable to estimating standard errors and ensuing confidence intervals on variance components in two-factor random ANOVA models. In light of prior negative findings regarding the application of bootstrapping to this particular problem, a recommendation of an "optimal" approach to resampling was sought. The study used Monte Carlo simulations to test the variance component estimation accuracy under simultaneous resampling of all effect factors in a random model verses resampling a single factor. The results indicated that single-factor was a preferrable method of resampling and produced reasonable estimates of both standard errors and confidence intervals (parametric and non-parametric). Additional suggestions for appropriate application of the technique are discussed.

Introduction

The application of variance component estimates stemming from analysis of variance (ANOVA) and other quadratic forms of linear models (Searle, 1971) to educational research settings has continued to increase in recent years. These variance component estimators provide important information about measurement parameters of interest (e.g. generalizability theory, Cronbach, Gleser, Rajaratnam and Nanda, 1972, Brennan, 1983) as well as experimental effect sizes and intraclass correlations. However, in either measurement or experimental applications, researchers usually require some confidence about the accuracy of the estimators obtained.

Unfortunately, many of the attempts to define the distributions of variance components have involved tedious and complex algorithms and there remains to some extent a lack of consensus about the most appropriate distributional form to use (Searle, 1971, Smith, 1982). Nonetheless, the issue of the <u>accuracy</u> of variance component estimators can be dealt with, despite any controversy over distributional form, if normality and orthogonality of the data are assumed. That is, distributional <u>properties</u> of variance component estimators can be sought. The most common such property is the sampling variance of the variance components (Searle, 1971, Smith, 1978 and Brennan, 1983), which can be directly extended to estimating confidence intervals. However, as Smith (1982) demonstrated, even under the assumptions of normality and orthogonality, estimation of the sampling variances of variance components and thus any ensuing confidence intervals may, at best, be marginal.

For this latter reason, more recent attempts to look at the distributional properties of variance components have considered the use of empirical confidence intervals (Brennan, Harris and Hanson, 1987, Smith, Luecht and Anderson, 1988). Under this approach, multiple samples are drawn and the desired confidence interval is determined directly from the percentiles (e.g. 0.05 and 0.95) of the distribution of samples. However, in practice, the acquisition of multiple samples may not be feasible. Accordingly, researchers have needed to look at alternatives for establishing empirical confidence intervals on variance components.

One method of estimating empirical confidence intervals from single samples of data has been termed "bootstrapping". The general bootstrap technique described by Efron (1979, 1982) is a resampling approach to estimating confidence intervals upon statistical parameters of interest. The technique involves rebuilding multiple data sets from a single sample data set. That is, an initial data set is resampled, with replacement, until a new data set is constructed, matching in size the original data set. The statistical parameter estimates of interest are computed and another data set is then drawn from the sample, again with replacement, and analyzed. This process of resampling continues until a large number of data sets have been constructed and analyzed. Empirical confidence intervals can then be directly estimated as percentile point equivalents in the derived distribution of resampled or reconstructed data sets.

The bootstrap technique has been successfully implemented for many applications (Chatterjee, 1984, Lunneborg and Tousignant, 1985, Lunneborg, 1985, Iventosch, 1987), however, the use of bootstrapping for obtaining confidence intervals on variance components has only met with marginal success. Brennan, Harris and Hanson (1987), looked at the bootstrap technique for developing confidence intervals in measurement situations and concluded that the method was somewhat ineffectual. It should be noted, however, that Brennan et al. used a single replication (data set) which may have limited their findings.

Smith, Luecht and Anderson (1988) extended the work of Brennan et al. (1987) by investigating bootstrapping under three orthogonal designs (a two factor crossed design, a three factor crossed design and a three factor nested design). Using Monte Carlo simulations involving many replications of data sets across a variety of design sizes, Smith et al. were able perform a series of large-scale tests of the bootstrap methodology with respect to confidence interval estimation and estimation of the sampling variances of the variance components in their designs.

In general, the results obtained by Smith et al. (1988) were somewhat less than favorable. The point estimates of the variance components (mean and median values) tended to overestimate the theoretical values (used to generate the data sets) of the variance components for main effects and underestimate any residual terms in their linear models. Since the data used by Smith et al. was controlled to simulate normality, both the parametric (use of the sampling variance) and non-parametric (empirical) confidence intervals produced similar results. However, those estimated results showed under- and overestimation inconsistency with the expected values of the theoretical sampling variances.

These marginal findings would ordinarily suggest that the bootstrap method holds little promise when applied to the problem of variance component estimation. However, two key points were alluded to but not specifically investigated in both studies (Brennan et al. ,1987, Smith et al., 1988). The first point concerns the size of data set(s) being resampled under bootstrapping. As Efron (1982) suggests, the variance of a statistical parameter of interest should take the form

$$\hat{\sigma}^2_{boot} = \frac{n - 1}{n} \hat{\sigma}^2 \quad . \tag{1}$$

Clearly, the size of the design under which the resampling takes place will impact the underestimation of the total variance, (σ^2). The variance components, as independent linear parameters, for example,

$$\sigma^2_y = \sigma^2_\alpha + \sigma^2_e \tag{2}$$

can be expected to likewise be restricted by any underestimation on the total variance, σ^2_Y in Equation (2).

Second, different strategies employed during the bootstrapping can be envisioned to have differential effects on the obtained variance components. For example, Smith et al. (1988) simultaneously resampled all possible independent parameters (i.e. main effect factors--a strategy similarly used by Brennan et al., 1987). That is, Smith et al. did not look at the potential of alternative bootstrapping strategies (e.g. resampling only one factor in a design). Although Brennan et al. did consider the issue alternative bootstrapping strategies, their use of a single data set may have precluded any positive findings.

These two points therefore provide the primary objectives of the present study. Under the assumptions of normality and orthogonality of linear designs, this study (1) evaluates the effect of design size on the estimation of variance components and distributional estimators (parametric sampling variances and non-parametric confidence intervals) and (2) seeks to provide some understanding of various resampling strategies, ultimately arriving at one "recommendable" strategy for resampling to estimate variance components.

Preliminary to an empirical investigation of these objectives, an initial attempt is made to describe the potential or expected impact of bootstrapping strategies and sample sizes upon variance components estimators. The next section describes the derivations of the variance components for a two factor random analysis of variance model, without replications and comments upon that expected impact as it relates to the objectives. Following that section, methods and results are provided for a two-part Monte Carlo study, meant to formally test various bootstrapping strategies across a variety of sample sizes.

Derivation of Variance Components under Bootstrapping

The present study considers a rather basic two factor random effects analysis of variance (ANOVA) design, without replications. This particular design was chosen because of its generality to testing contexts (e.g. generalizability theory, Cronbach et al, 1972, Brennan, 1983) and many experimental settings involving repeated measures. Although more complex crossed and nested designs were considered (see Smith et al., 1988), there seemed to be no explicit reason to include them here.

Under this ANOVA design the two random effects factors, A and B, have respective variance component estimators as follows (Brennan, 1983):

$$\hat{\sigma}^2_A = (MS_A - MS_{AB,e}) / b \qquad (3)$$

$$\hat{\sigma}^2_B = (MS_B - MS_{AB,e}) / a \, . \qquad (4)$$

Additionally, the residual variance confounds the interaction of the A and B factors with the error term (since no cell replicates are involved) in the form

$$\hat{\sigma}^2_{AB,e} = MS_{AB,e} \ . \tag{5}$$

Finally, the sampling variance of the estimators in Equations (3), (4) and (5) can be expressed as

$$\widehat{VAR}(\sigma^2_A) = \frac{2}{a - 1}\left[\left[\hat{\sigma}^2_A + (\hat{\sigma}^2_{AB,e}/b)\right]^2 + \frac{1}{b - 1}(\hat{\sigma}^2_{AB,e}/b)^2\right] \tag{6}$$

$$\widehat{VAR}(\sigma^2_B) = \frac{2}{b - 1}\left[\left[\hat{\sigma}^2_B + (\hat{\sigma}^2_{AB,e}/a)\right]^2 + \frac{1}{a - 1}(\hat{\sigma}^2_{AB,e}/a)^2\right] \tag{7}$$

$$\widehat{VAR}(\sigma^2_{AB,e}) = \frac{2}{(a - 1)(b - 1)}(\hat{\sigma}^2_{AB,e})^2 \tag{8}$$

as suggested by Smith (1978).

Smith (1978) came to the conclusion that fairly large numbers of levels of the involved factors were needed to establish stable confidence intervals, based upon the sampling variances (i.e. to reduce the standard errors of the variance component estimators), even under normality assumptions. Generally, for designs of this type, Smith recommended that $n_A n_B$ equal 800.

If that consideration of sample size is extended to bootstrapping, then the reduction of the total variance under bootstrapping (see Equation (1)) can be expected to further confound the variance component estimators and their sampling variances at different sample sizes. In short, the total variance should be <u>reduced</u> for smaller sample sizes.

For example, if normality of the data in this A x B design is assumed, then Equation (1) can be adapted to demonstrate the impact of both sample size and bootstrapping strategy upon the total variance, using degrees of freedom appropriate to this ANOVA model (*ab* -1), such that

$$\hat{\sigma}^2_{y(\text{boot})} = \frac{ab - 1}{ab} \hat{\sigma}^2_y \; . \tag{9}$$

In other words, as $a \rightarrow \infty$ and $b \rightarrow \infty$, the bootstrapped estimate of the total variance will approach the unbiased estimator of the theoretical total variance, which itself will approach the population value as the levels of *a* and *b* approach infinity. Correspondingly, for fairly small levels of *a* and/or *b*, there will be a reduced total variance estimate. For example, if the model of this ANOVA design

$$\sigma^2_Y = \sigma^2_A + \sigma^2_B + \sigma^2_{AB,e} \tag{10}$$

is considered, the impact of the underestimation of σ^2_Y should clearly extend to all variance components, as well.

It is at this point where the issue of bootstrapping strategies needs to be dealt with. If we assume a strategy of simultaneous bootstrapping of both the A and B factors in (10) then, as Brennen, Harris and Hanson (1987) demonstrate, the estimator of σ^2_A as a function of the average unbiased covariances among the levels of factor B, becomes

$$\hat{\sigma}^2_{A(boot)} = \frac{1}{b(b-1)} \left[\Sigma_i \Sigma_{i'} \left[\Sigma_j \frac{(Y_{ji} - Y_{.i})(Y_{ji'} - Y_{.i'})}{a - 1}\right]\right] \quad (11)$$

(for *i* not equal to *i'*) and the estimator of σ^2_B, as a function of the covariances in factor A, likewise yields

$$\hat{\sigma}^2_{B(boot)} = \frac{1}{a(a-1)} \left[\Sigma_j \Sigma_{j'} \left[\Sigma_i \frac{(Y_{ji} - Y_{j.})(Y_{j'i} - Y_{j'.})}{b - 1}\right]\right] \quad (12)$$

(for *j* not equal to *j'*). As Brennan et al. further suggest, these estimators imply an <u>overestimation</u> of the variance components for factors A and B, especially where *a* and *b* are small. If the residual is then obtained as a subtractive function of the bootstrapped total variance in (9), $\hat{\sigma}^2_{y(boot)}$, less the overestimated $\hat{\sigma}^2_{A(boot)}$ and $\hat{\sigma}^2_{B(boot)}$ components, by adaptively solving Equation (10) for $\hat{\sigma}^2_{AB,e(boot)}$, we therefore see that any overestimation of the main effect variance components will result in a proportional <u>underestimation</u> of the residual term as a function of both the restriction on the total variance and overestimation of the A and B factor components.

In fact, using this form of simultaneous bootstrapping, Smith, Luecht and Anderson (1988) actually performed a large-scale validation this precise effect, with some restriction on the limits of their design sizes (a 50 x 20 matrix was the largest design considered). The expectation might be that the overestimation in $\hat{\sigma}^2_{A(boot)}$ and $\hat{\sigma}^2_{B(boot)}$ (and corresponding underestimation in the residual term) would be further confounded if the levels in factors A and B were highly disproportionate. The present study seeks to recapitulate that disproportionate effect over a more varied range of sample sizes.

In contrast, consideration must also be given to a bootstrapping strategy which only resamples along a single factor. Under this strategy, levels of a single factor are resampled with the levels of the crossed factor automatically chosen for each selected level in the bootstrapped factor. For example, if only factor A is resampled, then for each level of A selected, all levels of factor B crossed at that level are automatically chosen. An approximation of the variance component estimators as a function of the cross factor covariances could, of course, be modeled as exemplified by Equations (11) and (12); however, a more straight-forward explanation appears warranted.

If the resampling occurs only in the levels of one factor, for example, factor A for the model in (10), then the expectations of the variance components (per Equation (1), Efron, 1982) should take the following form

$$\hat{\sigma}^2_{A(boot)} = \frac{a - 1}{a} \hat{\sigma}^2_{A} \tag{13}$$

and

$$\hat{\sigma}^2_{AB,e(boot)} = \frac{(a - 1)}{a} \hat{\sigma}^2_{AB,e} \cdot \tag{14}$$

where the estimator, $\hat{\sigma}^2_{B(boot)}$ in Equation (12), as a function of the average unbiased covariances across factor A remains unchanged. In other words, underestimation is constrained to a single factor in the design, which ideally can be controlled by increasing the number of levels, since for this example, as a approaches ∞, $\hat{\sigma}_{A(boot)}$ likewise approaches the unbiased estimator of σ^2_A. Since the covariances in factor A additionally determine $\hat{\sigma}^2_{B(boot)}$, that same increase in the levels of factor A can be expected to reduce the overestimation problem implicit in factor B. Finally, the residual variance component estimator under this approach to bootstrapping should be slightly underestimated, but only to a degree proportional to the levels in factor A (i.e. for fairly small resampling levels in A).

The remainder of this paper deals with an empirical test of the recommendations offered in this section and seeks to demonstrate single factor or "one-way" bootstrapping as a recommended strategy for estimating variance components and sampling variances, under the constraint of sample sizes appropriate for the resampling application.

Methods

Monte Carlo data sets were used to simulate the effects of bootstrapping strategies and sample sizes, under the rationale suggested in the prior section of this paper. A two-phase study was implemented.

Phases of the Investigation

In the first phase of this study, the issues of sample size effects and adequacy of bootstrapping strategies were simultaneously addressed. For this phase, multiple data sets were generated under the A x B paradigm and each data set was resampled (bootstrapped) 200 times. Data sets comprised of four different sample sizes were used: (1) $a = 20 \times b = 20$; (2) $a = 150 \times b = 20$; (3) $a = 20 \times b = 150$; and (4) $a = 150 \times b = 150$

For each data set and its resampling cycle (bootstrapping sequence), two strategies were implemented. First, each data set was bootstrapped with respect to both the A and B factors, effectively drawing a sample of each parameter matching the original levels in the design. This method of resampling both parameters (factors) of interest was suggested by Brennan et al. (1987) and the method of choice for Smith et al. (1988). At the same time, the resampled A-levels only were used under the second strategy. That is, all crossed levels of the B factor were automatically chosen whenever a particular level of the A factor was selected for the bootstrap sample. Since the sampling rates varied rotationally across factors (by 20 and 150 levels), this latter approach of holding resampling constant in the A factor seemed sufficient to address the issue of resampling a single parameter in contrast to resampling all parameters in a linear model.

For each data set and its resampling sequence of 200 bootstraps, the point estimates of the average of the variance components were saved and the sampling variance of those average estimators were computed.

In the second phase of this study, the direct estimation of confidence intervals was addressed. Here, the intent was to demonstrate the adequacy of bootstrapping only one parameter in the design (factor A), as suggested by the rationale and approximations in the prior section. As previously noted, the study by Smith et al. (1988) never tested this particular bootstrapping strategy.

476 data sets having 100 levels of the A factor and 50 levels of the B factor were generated. Using 200 bootstraps for each data set, the empirical confidence intervals (as well as the average point estimates and sampling variances) were saved for each replication.

__Data Generation and Analysis Algorithms__

The Monte Carlo data sets in both phases of this study were constructed to conform to a precise theoretical distribution. In each case, random normal deviates were scaled to "known" values of the variance components underlying the data. That is, each data set had an *a priori*, theoretical set of constants set at $\sigma^2_A = 0.25$, $\sigma^2_B = 0.25$ and $\sigma^2_{AB,e} = 0.50$, such that σ^2_Y equaled 1, with known contributions of the component variances, as stated.

The generation of the data sets, bootstrapping, analysis of variance and estimation of variance components and sampling variances were programmed and run on an IBM-AT compatible microcomputer with math co-processing capabilities accurate to 17 or 18 decimal places. The computational algorithms were further validated against results obtained via the Systat 4.0 MGLH module

(Systat, 1988), using smaller data sets.

Results and Conclusions

The first phase of this investigation was meant to model the expectations of σ^2_α estimators under bootstrapping by contrasting two resampling strategies across a variety of design sizes: (1) two-way resampling of both the A and B factors and (2) resampling of only the levels of factor A. In the second phase of analysis, the latter method of bootstrapping (resampling a single factor--here factor A) was re-investigated in terms of adequacy in establishing empirical (non-parametric) confidence intervals for fairly large data sets (100 x 50). A comparison of distributional parameters and confidence intervals estimated via those parameters was also incorporated into this phase of the study.

Phase I: Comparison of Bootstrap Strategies

Tables 1, 2, 3 and 4 present the results from this bootstrapping comparison phase, representing data sets of 20 x 20, 150 x 20, 20 x 150 and 150 x 150, respectively. Descriptive statistics for each variance component under single-factor bootstrapping are shown in the leftmost three columns of values for each table. The rightmost three columns of values in each table display the results from simultaneous or two-way bootstrapping.

Simultaneous (Two-Way) Bootstrapping

Tables 1-4 demonstrate a consistent overestimation of the factor A and B variance component point estimators (mean and median values across data sets), as described by Equations (11) and (12). Likewise, the anticipated underestimation of the residual terms in those tables is clearly evident. However, one anomaly surfaces which was alluded to earlier. Most noticeable in Tables 2 and 3, when there is disproportionality between the numbers of levels of the A and B factors, the overestimation problem favors the factor having more levels, and may even underestimate the smaller crossed factor.

Under that same condition of disproportionality, the estimators of the sampling variances, parametric confidence intervals and non-parametric confidence intervals also appear to fluctuate dramatically from expectation. It therefore seems apparent that simultaneous bootstrapping (two-way resampling for A x B designs) can be applied, but only under two highly restrictive constraints. First, the levels of the resampled factors must be approximately equivalent. Second, fairly large data matrices (e.g. 50 x 50 or greater) would be required to (a) reduce the overestimation of the main effects variance components and (b) bring the residual estimators up to a reasonable level.

For more complex designs (see Smith, Luecht and Anderson, 1988), the overestimation and underestimation problems would appear to create even more restrictions of mentionable concern to researchers (e.g. a 50 x 50 x 50 data matrix for a three-way crossed design is hardly practical to obtain in most experimental or even testing contexts).

For these reasons and staying with the recommendations made by Brennan et al. (1987) and Smith et al. (1988), it seems reasonable to discount simultaneous bootstrapping as a viable approach to obtaining usable confidence intervals or distributional estimators as a general application.

Single-Factor Bootstrapping

The point estimators (means and medians) in Tables 1 to 4, under single factor bootstrapping appear consistent with the expectations derived in Equations (12), (13) and (14). That is, there is a tendency to underestimate the bootstrapped factor (factor A) and overestimate the non-bootstrapped factor (factor B). Also, especially for designs having sample numbers of levels on the bootstrapped factor (see Table 1 and Table 3), the residual term, $\hat{\sigma}^2_{AB,e(boot)}$, tends toward very slight underestimation. It should be noted that this mild apparent overestimation of the residuals in Tables 2 and 4 is, technically speaking, not overestimation at all. Rather, consistent with the theory of bootstrapping (Efron, 1982), the residual estimators will approach the <u>sample</u> estimator as $a \rightarrow \infty$ and $b \rightarrow \infty$. Of course, the sample estimator, $\hat{\sigma}^2_{AB,e}$, will itself approach the theoretical population residual, as the levels of factors in the design approach infinity (Searle, 1971).

The sampling variances (theoretical, expected and estimated) also demonstrate a close correspondence. Likewise, under single-factor bootstrapping, both the non-parametric (empirical) and parametric provide similar information.

Assuming, as we have done from the onset, that the purpose behind bootstrapping is to estimate empirical confidence intervals, it therefore seems appropriate to conclude that single-factor bootstrapping appears to succeed where simultaneous bootstrapping could not. That is, bootstrapping only one factor in a design should provide reasonable estimation of the variance component parameters and confidence intervals around those parameters, under the constraint of having sufficient levels of the bootstrapped factor. For example, the greater overestimation of variance component for factor B in the 20 x 150 design (Table 3) is not seen in the 150 x 20 design (Table 2). That is, by increasing the number of levels on [single] bootstrapped factor, estimation bias for all factors appears to be reasonably controlled. Second, the underestimation effect on the residual term seems minimized under single-factor resampling. In contrast to the even more dramatic underestimation of residuals discovered by Smith et al. (1988) for more complex designs, it appears that single-factor bootstrapping will provide consistent and reasonable estimators, provided the resampled factor has sufficient numbers of levels.

Phase II: Analysis of the Single-Factor Bootstrap Strategy

Table 5 presents the results from the Phase II analysis of 476 data sets (using the A x B random model), with *a* equaling 100 levels and *b* equaling 50. Beyond point estimators and expected sampling variances of the estimators, an additional feature was added during this analysis phase. That feature was the calculation of the lower and upper bound percentile estimates and variance estimates for the components for each bootstrapping sequence of 200 iterations, across all data sets.

Accordingly, it becomes possible to compare four different sets of confidence intervals on the estimators: (1) the parametric 95% confidence intervals, using the actual sampling variance estimate of the components, (2) the non-parametric confidence intervals (percentile points set at 2.5% and 97.5%) for the data set point estimators, (3) the mean value of the lower and upper percentile points from each bootstrapped data set and (4) the median value of the lower and upper percentile points from each bootstrapped data set.

It is quite clear that, in general, the lower and upper bound estimators provide similar information, forming fairly symmetrical intervals around the theoretical values of the components. However, one interesting condition arises when considering the mean and median lower and upper estimators of $\hat{\sigma}^2_{B(boot)}$, across data sets. The sampling variances (see mean and median estimated sample variances in Table 5) derived for each bootstrapped sample are noticeably <u>smaller</u> than expectation; also evidenced by the restricted limits on the interval. While none of the other sampling variance estimators for $\hat{\sigma}^2_{B(boot)}$ suggest such lesser variation, the reduction is nonetheless quite distinct, and seems to occur only at the level of each bootstrapped data set. By way of explanation, it should be realized that the sampling variance $\widehat{VAR}(\hat{\sigma}^2_{B(boot)})$ is actually a variance of a covariance function (see Equation [12]). Since the sampling variance form of that covariance is unknown, but can be expected to be a restricted form (i.e. subject to the levels actually resampled in factor A) a smaller amount of variation seems logical. Although the interval on the $\hat{\sigma}^2_{B(boot)}$ estimators is "tighter", is still reasonably symmetrical and captures all pertinent point estimators, this anomaly nonetheless suggests a biased estimation problem. That is, the statistic may be useful for estimating the covariance due to bootstrapping, but should be treated cautiously as a valid estimator of the actual sampling variance of σ^2_B. Furthermore, any confidence intervals derived for factor B, may be misleading. It may be possible to overcome this sampling variance problem by independently bootstrapping only factor B in a secondary analysis; however, further research may be warranted on that count.

As a final note on the results, it should be be noticed that the point estimators (means and medians), while consistent with the earlier theoretical derivations in Equations (12), (13) and (14), in terms of over- and underestimation the Phase I results, strongly suggest the effect of estimation control which can be accomplished by increasing the numbers of levels in the resampled factor.

It therefore seems reasonable to suggest that single-factor bootstrapping presents a reasonable alternative to estimating variance components, under the constraint of adequate sampling (at least for the bootstrapped factor and residual). Of course, discovery of an "optimal" number of levels to resample remains a question for additional study. Also, the applicability of this technique to more complex designs or nonorthogonal designs, or merely non-normal data, requires further work.

References

Brennan, R. L. (1983). Elements of Generalizability Theory. Iowa City: ACT Publications.

Brennan, R. L., Harris, D. J. and Hanson, B. A. (1987). The bootstrap and other procedures for examining the variability of estimated variance components in testing contexts. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Wasnington, D.C.

Chatterjee, S. (1984). Variance estimation in factor analysis: an application of the bootstrap. British Journal of Mathematical and Statistical Psychology, 37, 252-262.

Cronbach, L. J., Gleser, G. C., Nanda, H. and Rajaratnam, N. (1972). The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York: Wiley.

Efron, B. (1979). Bootstrap method: another look at the jackknife. The Annals of Statistics, 7, 1-26.

Efron, B. (1982). The Jackknife, Bootstrap and Other Resampling Plans, Philadelphia: CBMS-NSF, Regional Conference Series in Applied Mathematics.

Iventosch, L. (1987). An application of the bootstrap technique to variance estimation in factor analysis. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D. C..

Lunneborg, C. E. (1985). Estimating the correlation coefficient: the bootstrap approach. Psychological Bulletin, 98, 209-215.

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

Searle, S. R. (1971). Linear Models. New York: Wiley.

Smith, P. L. (1978). Sampling errors of variance components in small sample generalizability studies. Journal of Educational Statistics, 3, 319-346.

Smith, P. L. (1982). A confidence interval approach for variance component estimates in the context of generalizability theory. Educational and Psychological Measurement, 42, 459-466.

Smith, P. L., Luecht, R. M. and Anderson, A. J. (1988). Bootstrapping variance component estimates. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Systat 4.0. (1988). MGLH. Evanston, IL: Systat, Inc.

Table 1
A Comparison of Bootstrapping Strategies for
$a = 20$ and $b = 20$

(No. of Data Sets = 481)

	Single Factor Bootstrapping			Bootstrapping All Factors		
	A	B	AB,e	A	B	AB,e
Value of Theoretical Variance Component	0.2500	0.2500	0.5000	0.2500	0.2500	0.5000
Variance of Theoretical Component	0.0080	0.0080	0.0014	0.0080	0.0080	0.0014
Mean Estimator of Component	0.2361	0.2865	0.4795	0.2708	0.2877	0.4543
Median Estimator of Component	0.2307	0.2752	0.4972	0.2607	0.2756	0.4527
Actual Variance of Component Estimators	0.0070	0.0091	0.0014	0.0082	0.0090	0.0013
Mean of Expected Sampling Variances	0.0079	0.0111	0.0013	0.0093	0.0111	0.0012
Median of Expected Sampling Variances	0.0068	0.0094	0.0013	0.0085	0.0094	0.0013
Parametric Lower Bound Estimator (95%)	0.0721	0.0995	0.4062	0.0933	0.1018	0.3836
Parametric Upper Bound Estimator (95%)	0.4001	0.4735	0.5528	0.4483	0.4736	0.5250
Non-parametric Lower Bound Estimator (2.5%)	0.0952	0.1341	0.4024	0.1217	0.1363	0.3819
Non-parametric Upper Bound Estimator (97.5%)	0.4192	0.4925	0.5513	0.4577	0.4925	0.5294

Table 2
A Comparison of Bootstrapping Strategies for
$a = 150$ and $b = 20$

(No. of Data Sets = 416)

	Single Factor Bootstrapping			Bootstrapping All Factors		
	A	B	AB,e	A	B	AB,e
Value of Theoretical Variance Components	0.2500	0.2500	0.5000	0.2500	0.2500	0.5000
Variance of Theoretical Component	0.0005	0.0068	0.0002	0.0005	0.0068	0.0002
Mean Estimator of Component	0.2514	0.2455	0.5029	0.3526	0.2429	0.4777
Median Estimator of Component	0.2487	0.2410	0.5042	0.2760	0.2279	0.4791
Actual Variance of Component Estimators	0.0012	0.0084	0.0007	0.1402	0.0078	0.0006
Mean of Expected Sampling Variances	0.0010	0.0079	0.0002	0.0038	0.0072	0.0002
Median of Expected Sampling Variances	0.0010	0.0063	0.0002	0.0012	0.0005	0.0001
Parametric Lower Bound Estimator (95%)	0.1835	0.0658	0.4510	-0.3810	0.0698	0.4297
Parametric Upper Bound Estimator (95%)	0.3192	0.4251	0.5547	1.0865	0.4160	0.5257
Non-parametric Lower Bound Estimator (2.5%)	0.1930	0.1205	0.4771	0.2195	0.1140	0.4523
Non-parametric Upper Bound Estimator (97.5%)	0.3164	0.4461	0.5329	1.7959	0.4530	0.5064

Table 3
A Comparison of Bootstrapping Strategies for
$a = 20$ and $b = 150$

(No. of Data Sets = 553)

	Single Factor Bootstrapping			Bootstrapping All Factors		
	A	B	AB,e	A	B	AB,e
Value of Theoretical Variance Component	0.2500	0.2500	0.5000	0.2500	0.2500	0.5000
Variance of Theoretical Component	0.0068	0.0005	0.0002	0.0068	0.0005	0.0002
Mean Estimator of Component	0.2491	0.2807	0.4807	0.2452	0.2788	0.4775
Median Estimator of Component	0.2343	0.2806	0.4813	0.2375	0.2783	0.4776
Actual Variance of Component Estimators	0.0068	0.0011	0.0001	0.0068	0.0011	0.0001
Mean of Expected Sampling Variances	0.0070	0.0013	0.0002	0.0072	0.0012	0.0002
Median of Expected Sampling Variances	0.0059	0.0012	0.0002	0.0061	0.0012	0.0002
Parametric Lower Bound Estimator (95%)	0.0874	0.2157	0.4530	0.0836	0.2138	0.4579
Parametric Upper Bound Estimator (95%)	0.4107	0.3457	0.5084	0.4068	0.3481	0.4971
Non-parametric Lower Bound Estimator (2.5%)	0.1102	0.2160	0.4575	0.1129	0.2144	0.4550
Non-parametric Upper Bound Estimator (97.5%)	0.4091	0.3461	0.5040	0.4143	0.3461	0.5002

Table 4
A Comparison of Bootstrapping Strategies for
$a = 150$ and $b = 150$

(No. of Data Sets = 117)

	Single Factor Bootstrapping			Bootstrapping All Factors		
	A	B	AB,e	A	B	AB,e
Value of Theoretical Variance Component	0.2500	0.2500	0.5000	0.2500	0.2500	0.5000
Variance of Theoretical Component	0.0009	0.0009	0.0000*	0.0009	0.0009	0.0000*
Mean Estimator of Component	0.2481	0.2568	0.5030	0.2524	0.2545	0.4987
Median Estimator of Component	0.2467	0.2560	0.5041	0.2504	0.2548	0.4995
Actual Variance of Component Estimators	0.0012	0.0010	0.0001	0.0012	0.0010	0.0001
Mean of Expected Sampling Variances	0.0009	0.0009	0.0000*	0.0009	0.0009	0.0000*
Median of Expected Sampling Variances	0.0008	0.0009	0.0000*	0.0009	0.0009	0.0000*
Parametric Lower Bound Estimator (95%)	0.1893	0.1980	0.4932	0.1936	0.1957	0.4889
Parametric Upper Bound Estimator (95%)	0.3069	0.3156	0.5128	0.3112	0.3133	0.5085
Non-parametric Lower Bound Estimator (2.5%)	0.1882	0.2087	0.4922	0.1923	0.2055	0.4888
Non-parametric Upper Bound Estimator (97.5%)	0.3060	0.3250	0.5116	0.3080	0.3184	0.5074

----
* <0.00005

Table 5
A Comparison of Confidence Interval Estimators
under Single-Factor Bootstrapping for
$a = 100$ and $b = 50$

(No. of Data Sets = 476)

	A	B	AB,e
Theoretical Components	0.25000	0.25000	0.50000
Theoretical Sampling Variances	0.00137	0.00265	0.00010
Mean Estimators	0.25267	0.26257	0.50043
Median Estimators	0.24870	0.25560	0.50100
Variance of Estimators	0.00200	0.00376	0.00027
Skewness of Estimators	0.46947	0.54686	-0.49107
Kurtosis of Estimators	0.15032	0.07724	3.03976
Mean of Exp. Sampling Variances	0.00143	0.00348	0.00010
Median of Exp. Sampling Variances	0.00135	0.00278	0.00010
Mean of Est. Sample Variances	0.00187	0.00015	0.00016
Median of Est. Sample Variances	0.00157	0.00013	0.00014
Parametric Lower Bound (95%)	0.16502	0.14239	0.46822
Parametric Upper Bound (95%)	0.34032	0.38275	0.53264
Lower Bound of Estimators (2.5%)	0.17510	0.16440	0.46250
Upper Bound of Estimators (97.5%)	0.35550	0.40150	0.53150
Mean Boot Lower Bound (2.5%)	0.17645	0.24006	0.47576
Mean Boot Upper Bound (97.5%)	0.33687	0.28577	0.52370
Median Boot Lower Bound (97.5%)	0.17150	0.23370	0.47740
Median Boot Upper Bound (97.5%)	0.33300	0.27820	0.52300