NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

The Working Paper Series was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series.

U. S. Department of Education
Office of Educational Research and Improvement

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

A Bootstrap Variance Estimator for Systematic PPS Sampling

Working Paper No. 98-12 October 1998

Contact: Steven Kaufman
Surveys and Cooperative Systems Group
email: steve_kaufman@ed.gov

U. S. Department of Education
Office of Educational Research and Improvement

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

> National Center for Education Statistics
> Office of Educational Research and Improvement
> U.S. Department of Education
> 555 New Jersey Avenue, NW
> Washington, DC 20208

The NCES World Wide Web Home Page is
http://nces.ed.gov

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *A Bootstrap Variance Estimator for Systematic PPS Sampling*. Working Paper No. 98-12, by Steven Kaufman. Washington, D.C.: 1998.

October 1998

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The ***Working Paper Series*** was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Ruth R. Harris at (202) 219-1831 (e-mail: ruth_harris@ed.gov) or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5654.

Marilyn McMillen
Chief Statistician
Statistical Standards and Services Group

Samuel S. Peng
Director
Methodology, Training, and Customer Service Program

This page intentionally left blank.

A Bootstrap Variance Estimator for Systematic PPS Sampling

Prepared by:

Steven Kaufman
National Center for Education Statistics

Prepared for:

U.S. Department of Education
Office of Educational Research and Development
National Center for Education Statistics

October 1998

Notice

Table of Contents

List of Tables

Abstract

In large multipurpose surveys, it is common to select the sample systematically proportional to some measure of size (PPS) which is correlated with an important variable of interest. Assuming the frame is sorted in a useful deterministic manner, systematic sample methodologies provide an additional control on the sample allocation, beyond the control provided from the stratification. This makes it less likely to select a 'bad sample'. This should reduce the variability of the estimates as compared to a comparable nonsystematic selection procedure. The problem with systematic samples is that variance estimators are biased. This paper presents a bootstrap variance estimator, which can have less bias than standard methodologies, such as half-sample replication. The results will be demonstrated with a simulation study based on an important National Center for Education Statistics' survey–The Schools and Staffing Survey.

Key Words: Simulation, Half-Sample Replication

The author would like to acknowledge Michael Cohen for his helpful comments on the draft manuscript.

1.0 Introduction

Systematic probability proportionate to size sampling (PPS) is a common selection method for establishment surveys. One way of selecting a PPS sample, given a frame and measure of size, is to do the following within each sampling stratum (h): Partition the Primary Sampling Units (PSUs) on the frame into n_h groups, so that the sum of the measures of size in each group is equal. For the sum of the measures of size to be equal within each group, some PSUs may be split into two adjacent groups, with some positive measure of size in each group. The sampling interval is the total measure of size in each group. Within each group, the first PSU is assigned a cumulative measure of size equal to the PSU's measure of size. The second PSU's cumulative measure of size is the sum of the previous PSU's (first PSU) cumulative size plus the current PSU's (second PSU) measure of size. This process continues until each PSU in a group is assigned a cumulative measure of size. The cumulative measure of size for the last PSU in each group is equal to the sampling interval. A random distance (d_h) is chosen between 0 and the length of the sampling interval. The first PSU with a cumulative measure of size larger or equal to d_h, within the first group, is the first PSU selected in the sample. The rest of the sample is selected by making the same cumulative measure of size comparison for each group. A total of n_h PSUs are selected within each stratum. (See Wolter 1985, pp. 283-286 for more details).

The measure of size for some PSUs may be larger than the sampling interval. There are two ways of handling this. All such PSUs can be excluded from the selection process and place in sample with certainty. The stratum sample sizes can be adjusted and new sampling intervals computed. The alternative is to select the sample without modification and adjust the weighting or second stage sample sizes to accommodate PSUs selected multiple times. The former is generally considered more efficient because the number of distinct selected PSUs will equal the original sample size, while the latter will be somewhat less.

Systematic sampling procedures are efficient in terms of ease of selection and lowering sampling error. For this reason they are used extensively in large-scale surveys. Since each stratum systematic sample is selected using a single random start (d_h), the sample can be viewed as a sample of size one, where each sample consists of a single sample cluster of

n_h PSUs. Therefore, it is impossible to produce an unbiased variance estimator. However, a number of biased methodologies are used for variance estimation.

These methodologies generally take one of two forms: 1) assume the systematic sample can be approximated by a simpler sample design with a known variance estimator or 2) assume the response variable follows some super-population model and a variance estimator is produced appropriate for that model. Both these approaches allow for grouping of PSUs, so variances can be computed within groups. Wolter (1985, chapter 7) provides a good discussion of a number of systematic sample variance estimators that can be classified into one of these two forms. An example, using balanced half-sample replication (BHR) is provided below.

BHR is a widely used variance replication methodology for complex survey designs. It is designed for samples where two primary sampling units (PSUs) within each stratum are selected with replacement. With BHR, choosing one PSU within each stratum generates a half-sample. A number of half-samples are generated by alternating which PSU, within stratum, go into the half-samples. The BHR variance is the simple variance of the half-sample estimates. Through a balancing process of the half-samples, the BHR variance estimate, for linear estimates, equals the direct sample variance estimate.

BHR can be adapted to designs where more than two PSUs are selected in a stratum by consecutively pairing selected PSUs, after placing them in the original order of selection; and assuming each pair is a stratum for variance estimation (variance stratum). If without-replacement sampling is used then a finite population adjustment can be applied. See Wolter (1985, pp. 110-152) for a more complete description of BHR.

In order to use BHR with systematic PPS sampling, it must be assumed that a PPS selection can be approximated by the deep stratification induced by the pairing described above. This assumption is reasonable, considering that the first sort variable, ignoring the lack of independence between breaks in the variable, can be considered an implicit stratification. However, BHR also assumes that the variance estimate is proportional to the inverse of the sample size. In section 2.0, it will be demonstrated, through a simulation study, that systematic sampling variances are not necessarily inversely proportional to the sample size. Two possible reasons for this are: 1) there is a finite population adjustment effect and 2) the variance decreases by some function other than $1/n_h$. From the simulation study, the amount of deviation from the $1/n_h$ model is a function of the response variable, as well as, the sample size. Hence, it is unlikely that the error can be corrected using a simple finite population correction. Therefore, BHR is not expected to perform well in this situation.

One methodology that does not necessarily assume the variance is inversely proportional to the sample size is the bootstrap (Efron, 1982). Given an estimate $\hat{\Theta}(X_1, X_2, ..., X_n)$, with $X_1, X_2, ..., X_n$ i.i.d. and distributed according to a distribution function $F(X_1)$, the bootstrap variance estimate for $\hat{\Theta}$ is based on the observation that $V(\hat{\Theta}) = \sigma^2(\hat{\Theta}, n, F)$ (i.e., the variance of $\hat{\Theta}$ is a function of $\hat{\Theta}$, the sample size n, and F). Given $\hat{\Theta}$ and n, $V(\hat{\Theta}) = \sigma^2(F)$. By analogy, the bootstrap variance, $V_*(\hat{\Theta})$, equals $\sigma^2(\hat{F})$, where $\hat{F}$ is the empirical probability distribution of X. If $\sigma^2(\cdot)$ is unknown then $V_*(\hat{\Theta})$ can be estimated by $1/B\sum_{i=1}^{B}(\Theta_i^* - \overline{\Theta}^*)^2$, for $\Theta_i^*, i = 1$ to B. $\hat{\Theta}_i^*$ is computed like $\hat{\Theta}$ is computed, except $X_1^*, X_2^*, ..., X_n^*$ is used instead of $X_1, X_2, ..., X_n$, with the X_j^*'s being independent realizations from $\hat{F}(x)$. This process is repeated B times to obtain B $\hat{\Theta}_i^*$'s. $X_1^*, X_2^*, ..., X_n^*$ is called a bootstrap sample with a bootstrap sample size (n^*) n.

The bootstrap procedure can be applied to stratified simple random sampling by applying the above bootstrap procedure within each sampling stratum. In this case, $V_*(\overline{X})$ equals $\sum_{h=1}^{L} 1/n_h[(n_h - 1)/n_h]s_h^2$, s_h^2 being the usual stratum population variance. Comparing this to the usual variance estimator, this methodology has two problems: 1) the finite population adjustment is missing and 2) there is a scaling bias term $[(n_h - 1)/n_h]$. If the sample rates are high then the missing finite population correction adjustment can be significant. If n_h is small, which is quite common in finite population sampling, then $[(n_h - 1)/n_h]$ can be large. With the basic BHR sample design, where n_h equals 2, the bias is -50%. Many adjustments have been suggested to correct these deficiencies.

In the BHR type sample design, where n_h equals 2 for all h and the estimate of interest is linear, setting $n_h^* = 1$ corrects for the $[(n_h - 1)/n_h]$ bias. Since BHR type designs are with-replacement designs, there is no finite population adjustment. The bootstrap variance estimator, therefore, becomes unbiased and consistent (see Efron 1982). This bootstrap variance estimator is similar to the BHR variance estimator. The difference is that PSU's selected for the BHR replicates are specified to produce the exact sample variance estimate, while PSUs are randomly selected for the bootstrap replicates and may not equal the exact sample variance.

When n_h is greater than 2, other more complicated adjustments have been proposed. The simplest of these is to, adjust n_h^* to eliminate the finite population correction and $[(n_h - 1)/n_h]$ bias problems (see McCarthy and Snowden 1985). However, n_h^* is not unique. Rao and Wu (1988) propose a rescaling bootstrap, which determines n_h^* to match the third order moments between $\hat{\Theta} - E\hat{\Theta}$ and $\hat{\Theta}^* - E\hat{\Theta}^*$ after rescaling $\hat{V}^*(\hat{\Theta}^*)$ to be unbiased for any n_h^*. Sitter (1992a) proposed a bootstrap designed for the case where PSUs are selected without replacement. In this bootstrap, $X_1^*, X_2^*, ..., X_{n_h^*}^*$ are generated from a series of without replacement samples from $X_1, X_2, ..., X_n$. By selecting $X_1^*, X_2^*, ..., X_{n_h^*}^*$ without replacement, the original sample design is followed more closely. All these methodologies directly select $X_1^*, X_2^*, ..., X_{n_h^*}^*$ from the original $X_1, X_2, ..., X_n$ and adjust n_h^* to correct for any biases and inconsistencies introduced in the bootstrap process.

To further mimic the actual sample design (see Gross 1980; Chao and Lo 1985), a bootstrap-frame can be generated to select the bootstrap samples. If $N_h = k_h n_h$ (N_h being number of PSUs in the frame for stratum h), the bootstrap-frame is generated by replicating each of the n_h sampled PSUs k_h times. Bootstrap samples are then selected using the original sampling methodology (i.e., simple random sampling without replacement). The bootstrap variance estimates based on these bootstrap samples no longer have a finite population correction bias, but have a scaling bias term $k_h(n_h - 1)/(k_h n_h - 1)$ (see Shao and Tu 1995, p. 250). By adjusting k_h and n_h^* according to Sitter (1992b), this bias can be eliminated.

All the bootstrap procedures described above assume simple random sampling (with or without replacement) within each stratum. Sitter (1992b) proposed a bootstrap variance estimator for the Rao-Hartley-Cochran sampling scheme (Cochran 1977, pp. 266-267). This is a 'PPS type' sampling scheme because it is similar to a PPS systematic selection where the stratum frame is placed in a random order before sample selection. The Rao-Hartley-Cochran sampling scheme independently selects one PSU, within each of n_h groups, proportional to some measure of size. The groups are generated by randomly assigning a specified number of PSUs into each of the n_h groups. Since the sum of the measures of size within a group are not equal the procedure is not strictly PPS either. The bootstrap methodology is similar to the bootstrap-frame procedures described above. First, a bootstrap frame is generated. Next, n_h^* is chosen to eliminate any biases. The bootstrap-PSUs are then randomly placed into one of n_h^* groups. Next, one bootstrap-PSU is independently selected within each of the n_h^* groups, generating a bootstrap sample. Before selecting each bootstrap sample, the bootstrap frame is re-randomized. A number of bootstrap samples are generated, as well as the appropriate bootstrap estimate. The Monte Carlo variance estimator of the bootstrap estimates is an estimate of the bootstrap variance estimator.

The bootstrap procedures described above assume the variance is inversely proportional to n_h (i.e., n_h^* and /or k_h are chosen knowing exactly how the true variance is related to n_h). If the true variance has a known relationship to n_h, different than proportional to $1/n_h$, then n_h^*, most likely, can be adjusted to compensated for the biases in the bootstrap variance. If the relationship is unknown, as is the case with systematic sampling (PPS or equal probability), then a simulation study can be done to compute an n_h^* that is approximately unbiased. This is the approach used in this paper to produce a bootstrap variance estimate for stratified PPS systematic sampling. A bootstrap-frame will be used to reduce any bias due to sampling without replacement. n_h^* will be computed by means of a simulation study, comparing the bootstrap variance for a specified n_h^* with an estimate of the true variance, to reduce any additional biases. A super-population model will be introduced to determine how to randomize the bootstrap-frame before selecting the bootstrap samples. Given the super-population model, the bootstrap variance estimator will be shown to be consistent.

2.0 Using the BHR Model with Systematic Sampling

The BHR model assumes systematic sampling can be approximated by a deep stratification introduced by pairing consecutive sampling PSUs. For this to work, the stratum variances must be proportional to $1/n_h$, since BHR makes this assumption. (When all BHR assumptions are true, this follows from $V_{BHR}(X) = V((X_1 + X_2)/2) = 1/2V(X_1)$, where subscript 1 and 2 represents the estimate based on the first and second PSUs respectively selected in each stratum.) If this assumption is not true then the BHR model is unlikely to produce accurate results. There are two reasons for systematic sampling to violate this assumption. The first reason deals with any implicit finite population correction in the variance to reflect sampling without replacement. If the sampling rates are high, this could be a significant contributor to the violation of the $1/n_h$ assumption. The second reason is the correlation between PSUs within a systematic sample. As n_h increase or decreases, these correlations may change dramatically because of the original sort ordering. Unlike the finite population adjustment, this effect can be noticeable even when the sampling rates are small.

To investigate the $1/n_h$ assumption, a simulation study is done, using the National Center for Education Statistics (NCES) elementary/secondary private school frame. Four thousand systematic samples are selected with sample sizes of $p_i n_h$, where $p_i = 1, 0.75, 0.5 \text{ and } 0.25$. By computing the simple variance of the 4,000 simulation estimate, an estimate of the true variance is computed. This is done for estimates of total students, teachers and schools. If the variance is proportional to $1/n_h$, then the ratio $\hat{V}(X_l^k)p_k / \hat{V}(X_l^j)p_j - 1$ should be close to 0; where l represents the estimate type (total students, teachers or schools) and k, j represents the sample size (n_h, .75n_h, .5n_h or .25n_h). When the ratio is less than 0, the systematic sample variance decreases faster than the $1/n_h$ assumption would imply; and when it is greater than 0, the systematic sample variance decreases slower than the $1/n_h$ assumption would imply. A negative ratio means that BHR should overestimate the variance, while a positive ratio means that BHR should underestimate the variance. This relationship is not necessarily true, since $\hat{V}(X_l^k)$ includes an unknown implicit finite population correction whose impact on the variance as the sample changes is unknown.

The results in table 1 demonstrate that sometimes the ratio is close to 0. Other times, it is a great deal different than 0. The systematic sampling variance does not necessarily decrease faster than the $1/n_h$ assumption would imply; sometimes its decrease is slower. This is an indication that BHR will not necessarily produce an overestimate of the variance, which is a common assumption among sampling statisticians. When there is a large difference from 0, the magnitude is dependent on the variable. This seems to imply, since the sampling rates are not high, (especially with the .5/.25 comparison), that the violation of the $1/n_h$ assumption is due to the initial sort ordering (i.e., the within sample correlation).

It should be noted that the table 1 results exaggerate the true impact of the $1/n_h$ assumption. Using the $1/n_h$ assumption, the ratio, used in the table, adjusts the variance with the smaller sample size to approximate the variance with the larger sample size. This approximation uses the smaller sample estimate's unknown finite population correction. Since the true finite population correction is likely larger than the one used in the approximation, the absolute value of the true impact of the $1/n_h$ assumption should be expected to be smaller than what table 1 indicates.

The important conclusion from this example is that variance estimates, based on designs using systematic sampling, will not necessarily be proportional to $1/n_h$, as n_h increases or decreases. When this occurs, an important BHR assumption is violated, and the BHR variance estimator should not be expected to perform well when the magnitude of the violation is large.

The statements concerning the proportionality of the variance estimate are qualified with 'as n_h increases or decreases'. The importance of this qualification can be seen with equal probability systematic sampling. Here, the variance can be expressed proportional to $1/n_h$ (e.g., $V(\bar{y}_h) = [(N_h - n_h)/N_h][S_{wst}^2/n_h][1+(n_h-1)\rho_{wst}]$, see (Cochran 1977, p. 209). If S_{wst}^2 and ρ_{wst} are constant for an arbitrary n_h then $V(\bar{y}_h)$ would be approximately proportional to $1/n_h$, as n_h increases or decreases. However, both S_{wst}^2 and ρ_{wst} are within systematic sample population estimates.

This implies that as n_h changes, the systematic samples change; hence S^2_{wst} and ρ_{wst} also change by some unknown function of n_h. Therefore, even though $V(\bar{y}_h)$ is proportional to $1/n_h$ for fixed n_h, as n_h increases or decreases, the variance may not be proportional or even closely proportional to $1/n_h$. In section 3.5, a similar result will be presented for unequal probability systematic sampling.

In terms of BHR where $V_{BHR}(X) = 1/2V(X_1)$, the S^2_{wst} and ρ_{wst} used in $V(X_1)$ may be different than the S^2_{wst} and ρ_{wst} from $V(X)$. So, $V_{BHR}(X)$ may over or under estimate the true variance ($V(X)$) based on the relationship between the two sets of S^2_{wst} and ρ_{wst}.

Table 1.–Measurement of degree the true systematic sampling variance is proportional to $1/n_h$ with respect to different sample sizes[1]

Stratum (h)	n_h/N_h (%)	Number of Teachers			Number of Students			Number of Schools		
		R100/50[2] (%)	R50/25[2] (%)	R75/25[2] (%)	R100/50[2] (%)	R50/25[2] (%)	R75/25[2] (%)	R100/50[2] (%)	R50/25[2] (%)	R75/25[2] (%)
01911	2.0	-31.2	19.5	55.2	14.1	-13.5	0.3	-28.0	-8.5	100.8
01912	2.8	-27.0	24.5	8.2	-2.5	6.6	6.0	-14.3	12.5	22.3
01913	6.1	-14.3	-11.9	3.8	20.9	19.2	-3.6	-3.0	3.2	10.2
01914	3.4	-23.8	-26.6	2.5	-5.3	8.8	2.8	-19.3	-3.7	-7.4
01921	19.6	23.5	25.3	18.0	56.5	141.9	110.1	19.6	34.7	10.8
01922	25.7	-32.7	14.9	132.1	-34.0	18.5	15.2	-16.4	35.4	27.1
01923	13.7	-4.3	-51.7	-31.2	6.0	12.0	175.5	-8.4	10.1	-10.1
01924	12.2	-46.3	38.2	-45.2	-37.6	-26.5	-7.9	-37.6	28.9	-13.9
01931	4.5	23.7	-7.5	-21.3	18.0	-4.5	-17.7	4.6	-6.3	-11.0
01932	4.9	2.4	1.1	-29.6	-25.8	4.0	-37.5	3.9	-9.4	-0.9
01933	6.2	18.5	43.5	51.7	-21.4	-24.0	-52.0	14.4	47.9	100.6
01934	4.3	-20.4	-9.6	-34.4	-26.3	-16.9	-18.9	-7.0	1.5	-3.8

[1] Negative numbers represent how much more efficient the variances are than the $1/n_h$ assumption.
Positive numbers represent how much less efficient the variances are than the $1/n_h$ assumption.

[2] Ra/b is the comparison for a % of original sample size to b % of original sample size.

3.0 Bootstrap Variance Model

To address the situation when the systematic variance in not proportional to $1/n_h$, a bootstrap variance estimator is proposed in this paper, which is less dependent on the $1/n_h$ assumption than the BHR estimator. This section first describes the necessary super-population model; next, a consistency theorem for the bootstrap estimator is presented; by example, the super-population model, used in the proposed bootstrap procedure, is demonstrated; next, the mechanics of the bootstrap procedure is presented; and finally, the consistency of the bootstrap procedure is established. We begin by describing the super-population model.

3.1 The General Super-Population Model

Let s_{ih}, the i^{th} possible sample from stratum h, be a systematic PPS sample such that $s_{ih} = \bigcup_{c=1}^{C_h} s_{ihc}$, where the s_{ihc}'s are a partitioning of the sample s_{ih} and C_h is the number of partitions within stratum h.

Each sampled PSU $j \in s_{ih}$ is assigned a known weight (w_j) based on the selection probability:

$$w_j = T_h / (n_h e_{hj}),$$

where $T_h = \sum_{j=1}^{N_h} e_{hj}$, (1)

e_{hj} is the measure of size for PSU j in stratum h, (2)

n_h is the number of sampled PSUs in stratum h,

and N_h is the number of PSUs in stratum h.

Let $p_h = n_h / n$, (3)

where $n = \sum_h n_h$.

We will assume that the p_h 's are constant as n increases.

For a random variable X_j evaluated for each PSU j, assume the joint distribution of $Y_j = (T_h /(p_h e_j))X_j$ given s_{ihc} is $\prod_{j=1}^{n_{ihc}} F_{hc}(Y_j)$, where $F_{hc}(Y_j)$ is a distribution function, and n_{ihc} is the number of sampled PSUs in partition c within stratum h.

Assume the distribution of $\mathbf{Y}_i^{hc}$ is independent of $\mathbf{Y}_i^{hc'}, c \neq c'$, given s_{ih}, where $\mathbf{Y}_i^{hc} = (Y_1,...,Y_{n_{ihc}})'$ on s_{ihc} and $\mathbf{Y}_i^{hc'} = (Y_1,...,Y_{n_{ihc'}})'$ on $s_{ihc'}$.

The joint distribution of $\mathbf{Y}_h = \bigcup_{c\in s_{ih}} \mathbf{Y}_i^{hc}$, given s_{ih}, is $\prod_{c=1}^{C_h} (F_{hc})^{n_{ihc}}$.

The distribution of $\mathbf{Y}_h$ given s_{ih} can be summarized in the following way: Within a partition of s_{ih} the Y_j 's are i.i.d. The Y_j 's $\in s_{ihc}$ and the Y_j 's $\in s_{ihc'}$, $c \neq c'$, are independent, but not identically distributed.

Theorem 1

Given the above assumptions and definitions, the bootstrap variance estimator of $\hat{X} = \sum_h \sum_{j\in h} w_{hj} x_j$ given the super-population model is consistent, as $n \to \infty$, provided $\hat{F}_{hc}(y) \to F_{hc}(y)$ and $\mu^*_{yhc} \to \mu_{yhc}$, as $n \to \infty$.

The proof is provided in the appendix.

This result can be generalized by noting that implementing the bootstrap methodology does not require knowledge of the conditional distributions, F_{hc}. The only thing required is the knowledge that the PSUs within s_{ich} are i.i.d. and PSUs between s_{ich} and $s_{ic'h}$, $c \neq c'$, are independent. This implies PSUs within s_{ich} must be randomized together as a group.

The following theorem states this generalization.

Theorem 2

The required assumptions are:

1) a systematic sample (s_{ih}) has a known partition (i.e., $s_{ih} = \bigcup_{c=1}^{C_{ih}} s_{ihc}$);

2) $\hat{X} = \sum_h \sum_{j \in h} w_{hj} x_j = 1/n \sum_h \sum_{j \in h} y_j$ is the estimate of interest;

3) as n increases the sample allocation between stratum remains constant (i.e., the p_h 's are constant as n increases);

4) for PSUs in s_{ihc}, the y_j 's are conditionally i.i.d. given s_{ihc} and are generated from an otherwise unspecified distribution function $F_{hc}(y) \in \Im_{2,1}$. $\Im_{2,1}$ being defined in the appendix and $E(y_j) = \mu_{yhc}$;

and

5) $\mathbf{Y}_i^{hc}$ is conditionally independent of $\mathbf{Y}_i^{hc'}, c \neq c'$, given s_{ih} .

It then follows that the bootstrap variance estimator of $\hat{X}$ given s_{ih} generated from the bootstrap estimates $\hat{X}_b^* = 1/n \sum_h \sum_{j \in h} y_j^*$, where the y_j^*'s are generated from $\hat{F}_{hc}(y)$, is consistent, as $n \to \infty$, provided $\hat{F}_{hc}(y) \to F_{hc}(y)$ and $\mu_{yhc}^* \to \mu_{yhc}$, as $n \to \infty$. μ_{yhc}^* is the bootstrap expectation of y within a partition.

3.2 Bootstrap Model Example

In practice, the statistician never knows the required partitioning ($s_{ih} = \bigcup_{c=1}^{C_h} s_{ihc}$). However, the statistician usually orders the frame before sample selection. With this ordering, the statistician is implicitly assuming that nearby PSUs are similar, at least in terms of the most important response variables. This implicit assumption can be used to develop a partitioning that approximately meets the required assumptions. This approach is similar to the BHR approach, so that the starting point of the bootstrap and the BHR are the same. Any differences in results reflect the divergence of assumptions at this point.

An example is provided below. However, it is first necessary to redefine the meaning of the term sampling interval. In the introduction, the sampling interval was the total measure of size within a sampling group. From now on, a sampling interval will refer to the PSUs within a sampling group. Consecutive sampling intervals are consecutive sampling groups.

Example

For a fixed even numbered sample size (n_h), the elements of the partition (s_{ich}) can be determined by consecutively pairing sampling intervals, after the frame has been placed in its original sort ordering. All samples have the same partitioning (i.e., the partitioning is only a function of stratum, -- s_{ch} , $c = 1$ to C_h) and each s_{ich} (s_{ch}) has exactly two PSUs. The marginal distribution of $\mathbf{Y}_i^{hc}$ is $(F_{hc})^2$. If y is distributed as $n(\mu_c, \sigma_c)$ with distribution function $\Phi_{\mu_c,\sigma_c}(y)$ then the distribution of $\mathbf{Y}_i^{hc}$ is $(\Phi_{\mu_c,\sigma_c}(y))^2$. In terms of consistency, it is assumed that the partitioning remains fixed as the sample size increases and more PSUs are selected within a partition.

This "type" of partitioning is used in the bootstrap procedures proposed in this paper. Implementing the partitioning in the bootstrap processes is similar to the BHR model, in that, for both methodologies, all the replicate variability is introduced through the variability within consecutive pairs of selected PSUs. The BHR replicates reflect variability by alternating PSUs within pairs in and out of the replicates, while the bootstrap replicates reflect variability by randomizing bootstrap-PSUs generated from consecutively paired PSUs (see section 3.3 for a description of the bootstrap-PSU process). The bias in the bootstrap procedure gets smaller as the number of PSUs per partition (s_{ch}) gets larger (see theorem 2). Since the proposed partitioning has only two PSUs per partition, the bias may be large; hence, setting the bootstrap sample size, n_h^*, equal to n_h should underestimates the bootstrap variance (see the bootstrap discussion in the introduction). Using $n_h^* = n_h / 2$, as in BHR, may have less bias. The bootstrap, unlike BHR, does not have to use $n_h^* = n_h / 2$. Any n_h^* between 1 and n_h can be used and may have less bias than either $n_h / 2$ or n_h. In this sense, the bootstrap model is more flexible than the BHR model.

The actual bootstrap sample size must be computed through a series of trial and error simulations, comparing and estimate of the true variance with the bootstrap variance for a specific bootstrap sample size. The bootstrap sample size that minimizes the bias in the bootstrap variance is used in the final implementation. The trial and error process is necessary because there is no direct formula that expresses the systematic variance as a function of n_h, as n_h increase or decreases, such as being proportional to $1/n_h$.

Determining n_h^* through a simulation provides a robust variance estimate because $V^*(\hat{X}_h)$, by construction, will be almost unbiased, even if the model assumptions are false. The disadvantage of the simulation is that it can only be implemented with frame variables. However, if n_h^* is relatively flat for non-frame variables, the bootstrap replicate weights should be applicable for those variables, too.

An additional observation about this partitioning is:

If the partitioning methodology described above correctly models the distribution of X; the n_h's are even and increase by multiples of C_h then the $\underset{2}{E}(\hat{X}_i \mid s_h\text{'s}) = K$, a constant; where $\underset{2}{E}$ refers to the expectation with respect to the super-population model. Therefore, $V(\hat{X}_i) = \underset{1}{E}\,\underset{2}{V}(\hat{X}_i \mid s_h\text{'s}) + \underset{1}{V}\,\underset{2}{E}(\hat{X}_i \mid s_h\text{'s}) = \underset{1}{E}\,\underset{2}{V}(\hat{X}_i \mid s_h\text{'s})$, where 1 refers to the selection of the s_h's. In this situation, since the bootstrap variance estimator is consistent for $\underset{2}{V}(\hat{X}_i | s_h\text{'s})$, the bootstrap variance is consistent to an estimator that is unbiased for the unconditional variance.

3.3 Bootstrap Implementation

The object of this section is to produce a set of bootstrap replicate weights, similar to BHR or Jackknife replicate weights. To do this, one must make an important distinction between finite population sampling and i.i.d. sampling. With i.i.d. sampling the variable of interest, X, is consider random. Therefore, it's logical to repeat the bootstrap sampling independently for every random variable. In finite population sampling, the response variable, X, is considered known for all PSUs in the frame. What is random is the sample selection variable, S_j, which specifies which PSUs are in the sample (i.e., S_j =1 means PSU j is in sample, while S_j =0 means PSU j is not in sample). Since there is only one random variable, only one set of bootstrap samples need be generated. This can be seen in Sitter's Rao-Hartley-Cochran bootstrap variance estimator. In it, the bootstrap sampling is done independent of the variable of interest. Therefore, once the bootstrap samples are selected, they are appropriate for any variable of interest. Likewise, for random sampling schemes described in the introduction. A PSU can be selected in the bootstrap selection process instead of a response variable without changing the procedure. Once selected, the bootstrap samples are appropriate for any response variable. Therefore, one set of selections can be used for all response variables. This does assume that the bootstrap sample size is not a function of the response variable.

Therefore, given a bootstrap sample, an appropriate bootstrap replicate weight for PSU j for any response variable X, is the sum of the bootstrap weights from selected bootstrap-PSUs, which have been generated from PSU j. For the bootstrap procedure proposed here, it may be very difficult and time consuming to produce a set of bootstrap replicated weights, but that process only has to be done once. Given availability of high-speed computers, their cheap run times, and the usual long time period from sample selection and the production of final weights, producing a single set of bootstrap replicate weights is certainly practical.

One potential difficulty with a single set of replicate weights for the proposed procedure is the process of choosing n_h^* to be more unbiased. With systematic sampling, this process is likely dependent on the response variable. Hence, each variable may have a slightly different set of replicate weights. It is assumed here that one set of replicate weights that works well for the frame variables will also work well for other correlated variables.

Since $\hat{F}_{hc}(y)$ is based on only 2 PSUs, it is not likely to be close to $F_{hc}(y)$ and therefore, $V^*(\Theta)$ may not be close to $V(\Theta)$. Hence, the bootstrap variance estimator requires the computation of a bootstrap sample size, n_h^*, to make the variance estimator more unbiased. Since there is no exact expression that relates the true systematic sampling variance with n_h, as n_h increases or decreases, determining an appropriate n_h^* will be accomplished by a simulation study. To perform the simulation study frame variables are used, so estimates can be computed for any selected sample. The statistician always has three estimates available for this purpose. One is the measure of size or some function of the measure of size. The second is the estimate of the total number of PSUs (sum of the sample weights). The third is the average measure of size per PSU or the average per PSU of some function of the measure of size. If the measure of size is used in the simulation, it will be necessary to use a different year's data to produce estimates; otherwise, the variances will be zero.

To determine the appropriate n_h^*'s, the simulations must first be applied to individual stratum estimates Θ_h . Therefore, the simulation process for estimating the bootstrap variance, $V^*(\Theta_h)$ for an estimator Θ_h , works as follows:

3.3.1 Bootstrap Procedures

1. Select a sample (s_i) from the original frame, using the methodology of the original sample design.
2. For the initial bootstrap sample size values, n_h^*, use either $n_h/2$ or n_h. $n_h/2$ would be appropriate if it is believed that the BHR deep stratification model is appropriate. If n_h is large then setting $n_h^* = n_h$ may be appropriate, if one believes that the deep stratification model is an oversimplification. After the initial simulation,

n_h^* will likely require adjustment for at least some of the strata. If such is the case, it will be required to repeat the simulation with the new n_h^*'s.

3. Generate a bootstrap frame based on the selected sample. The idea behind the bootstrap frame is to use the sample weights (w_j) from the selected PSUs (j) in s_i to estimate the PSU frame distribution. The bootstrap frame is generated in the following manner:

   For each selected PSU j, w_j bootstrap PSUs (bj) are generated by replicating the j^{th} PSU w_j times. If w_j has a noninteger component then a full bootstrap-PSU is generated with a reduced selection probability. The bj^{th} bootstrap-PSU has the following measure of size (m_{bj}):

$$m_{bj} = I_{bj} \cdot 1/w_j, \qquad (4)$$

   where:

$$I_{bj} = \begin{cases} 1, \text{ if } bj \text{ is an integer component of } w_j \\ C_i \text{ , if } bj \text{ is a noninteger component of } w_j \\ \qquad C_j \text{ being the noninteger component} \end{cases} \qquad (5)$$

$$w_j : \text{is the full-sample weight for PSU } j$$

4. Randomize the bootstrap frame according to super-population model specification. This is accomplished by placing the bj bootstrap-PSUs generated from PSU j within stratum h and sample s_i in their original order of selection. Next, bootstrap-PSUs generated from the first PSU are paired with the next set of bootstrap-PSUs generated from the second PSU. The third set of bootstrap-PSUs is paired with the fourth set. This process continues until all bootstrap-PSUs are paired. If there are an odd number of PSUs then the last set of groupings of bootstrap-PSUs contains the bootstrap-PSUs generated from the last three PSUs in stratum h. This is repeated for every stratum in s_i. Now, the bootstrap-PSUs are randomized within their respective pair.

5. Select B bootstrap samples from the bootstrap frame, re-randomizing the bootstrap frame before each selection. The bootstrap frame, bootstrap frame ordering, measure of size (m_{bj}), and bootstrap sample size (n_h^*) have been

specified. Using these quantities select the bootstrap samples using the same procedures used to select the original systematic PPS sample. The one exception to this is that a bootstrap-PSU generated from noncertainy PSUs that become certainty in the bootstrap selection should not be eliminated from the selection process and taken in sample with probability 1. Their selection probability should remain unchanged and if the bootstrap-PSU is selected multiple times that should be reflected in the bootstrap weight (see 6 below).

6. Compute bootstrap estimates Θ^*_{ibh} for each of the B bootstrap samples in an analogous manner as is done to compute the full sample estimate Θ_{ih} from s_i. This is accomplished by computing a bootstrap weight, w^*_j, and then computing Θ^*_{ibh} the same way Θ_{ih} is computed, except using w^*_j instead of w_j.

   The bootstrap-PSU weight, w^*_j, is:

$$w^*_j = \sum_{bj \in S^B_j} w^p_{bj}$$

S^B_j : is the set of all bj generated from j that are selected in the B^{th} bootstrap sample.

and

$$w^p_{bj} = I_{bj} \cdot M_{bj} / p_{bj}$$

where:

I_{bj} : is as previously defined

M_{bj} : is the number of times the bj^{th} bootstrap- PSU is selected, (6)

p_{bj} : is the bootstrap selection probability for the bj^{th} bootstrap-PSU.

$$p_{bj} = m_{bj} / SI_h,$$

where:

m_{bj} : is previously defined

SI_h : $\sum_{bj \in s_h} m_{bj} \Big/ n^*_h$.

7. Compute the simple variance of the Θ^*_{ibh} from $b = 1$ to B, as the Monte Carlo estimate of $V^*(\Theta_{ih})$ from s_i.

$$V^*(\Theta_{ih}) = 1/(B-1)\sum_{b=1}^{B}(\Theta^*_{ibh} - \overline{\Theta}^*_{ih})^2, \text{ where } \overline{\Theta}^*_{ih} = 1/B\sum_{b=1}^{B}\Theta^*_{ibh}.$$

8. Repeat steps 1-7, for a large number of samples, s_i, say T times.

9. Compute the simple variance of Θ_{ih} from $i = 1$ to T, $\hat{V}(\Theta_h)$, as a measure of the true variance; and compute the average bootstrap variance $\overline{V}^*(\Theta_h)$, averaged over the T, $V^*(\Theta_{ih})$ estimates.

$$\hat{V}(\Theta_h) = 1/(T-1)\sum_{i=1}^{T}(\Theta_{ih} - \overline{\Theta}_h)^2, \text{ where } \overline{\Theta}_h = 1/T\sum_{i=1}^{T}\Theta_{ih}.$$

$$\overline{V}^*(\Theta_{ih}) = 1/T\sum_{i=1}^{T}V^*(\Theta_{ih}).$$

10. Compare $\overline{V}^*(\Theta_h)$ with $\hat{V}(\Theta_h)$ and adjust n^*_h to reduce the bias between $\overline{V}^*(\Theta_h)$ and $\hat{V}(\Theta_h)$. If $\overline{V}^*(\Theta_h)$ is smaller than $\hat{V}(\Theta_h)$ then n^*_h should be reduced. If $\overline{V}^*(\Theta_h)$ is larger than $\hat{V}(\Theta_h)$ then n^*_h should be increased. Since $\overline{V}^*(\Theta_h)$ and $\hat{V}(\Theta_h)$ may not be proportional to $1/n^*_h$, it may be difficult to predict how much n^*_h should be increased or decreased. If that is the case then trial and error may be necessary.

11. Repeat steps 1-10, until the bias between $\overline{V}^*(\Theta_h)$ and $\hat{V}(\Theta_h)$ has been reduced to a satisfactory level.

12. Using the n^*_h from step 11, repeat steps 3-6 for the actual collected sample, generating a set of bootstrap replicate weights, w^*_j that can be used to compute variances of other, more complex statistics that are not necessarily computed within h.

3.4 Consistency of the Proposed Bootstrap Estimator

To apply theorem 1or 2 for the consistency of the bootstrap estimator, it must be established that $\hat{F}_{hc}(y) \to F_{hc}(y)$ and $\mu^*_{hcY} \to \mu_{hcY}$, as $n \to \infty$.

To do this, observe that for an arbitrary domain D:

$$E_*\left(\sum_{bj\in D} w^p_{bj}X_{bj}\right) = \sum_{bj\in D} I_{bj}X_{bj} = \sum_{j\in D} w_j X_j, \text{ since } M_{bi} \text{ has expectation } p_{bi}.$$

where:

E_* is expectation over the bootstrap samples,

X_{bj}, X_j is an arbitrary response variable defined for the bootstrap-PSUs and PSUs, respectively.

And

$$\hat{X}_D^* = \sum_{bj\in D} w_{bj}^p X_{bj}^* = 1/n^* \sum_{bj\in D} Y_{bjX}^* = \overline{Y}_{DX}^* \Rightarrow E_*(\hat{X}_D^*) = E_*(\overline{Y}_{DX}^*)$$

where:

$$Y_{bjX}^* = (I_{bj} M_{bj} T_h^*)/(p_h^* m_{bj}) X_{bj}^*$$
$$T_h^* = \sum_{bj\in h} m_{bj}$$
$$p_h^* = n_h^* / n^*$$

I_{bj}, M_{bj} and m_{bj} are defined in (5), (4) and (6), respectively from section 3.3.1 step 3 or step 6.

If D_1 is the domain defined as the observations in a stratum partition, s_{ich} then

$$E_*(\overline{Y}_{D_1X}^*) = E_*(\sum_{bj\in D_1} w_{bj}^p X_{bj}^*) = \sum_{j\in D_1} w_j X_j = 1/n \sum_{j\in D_1} Y_{jX} \to \mu_{hcY} \text{ , as } n \to \infty \text{ .}$$

Where:

$$Y_{jX} = (T_h / (p_h e_{hj})) X_j$$

T_h, p_h and e_{hj} are defined in (1), (3) and (2), respectively from section 3.1.

Define for each j in s_{ich}, $Z_{jch}(y_0) = \begin{cases} 1 \text{ if } y_j \le y_0 \\ 0 \text{ otherwise} \end{cases}$,

and for each bj in s_{ich}, $Z_{bjch}^*(y_0) = \begin{cases} 1 \text{ if } y_{bj}^* \le y_0 \\ 0 \text{ otherwise} \end{cases}$

then

$$
\begin{aligned}
\hat{F}_{hc}(y_0) &= (1/\sum_{bj\in s_{ich}} w_{bj}^p) E_*(\bar{Y}^*_{D_1 Z^*_{bjch}(y_0)}) = (1/\sum_{bj\in s_{ich}} w_{bj}^p) E_*(\sum_{bj\in D_1} w_{bj}^p Z^*_{bjch}(y_0)) \\
&= (1/\sum_{j\in D_1} w_j Z_{jch}(y_{\max}))\sum_{j\in D_1} w_j Z_{jch}(y_0) = (n/\sum_{j\in D_1} Y_{jZ_{jch}(y_{\max})})(1/n\sum_{j\in D_1} Y_{jZ_{jch}(y_0)}) \\
&= (1/\sum_{j\in D_1} Y_{jZ_{jch}(y_{\max})})(\sum_{j\in D_1} Y_{jZ_{jch}(y_0)}) \to F_{hc}(y_0), \text{as } n_{ch} \to \infty
\end{aligned}
$$

where: $y_{\max}$ is the maximum value of the y 's in s_{ich} .

The convergence above follows from $\sum_{j\in D_1} Y_{jZ_{jch}(y_{\max})}$ converging to the population total in D_1 , and $\sum_{j\in D_1} Y_{jZ_{jch}(y_0)}$ converging to population total with $y \le y_0$ in D_1 .

3.5 The Variance of $\hat{X}_h^*$ ($V^*(\hat{X}_h^*)$) from the Bootstrap Frame

In this section, an approximate expression for $V^*(\hat{X}_h^*)$ is derived to help explain: 1) the type of finite population implied by the proposed bootstrap variance estimator; and 2) how the original sample size affects the variance estimator. To do this, the following assumptions are made: 1) all weights are whole integers; 2) within a stratum partition the weights are equal and 3) $n_h^* = n_h$. Given these assumptions, the proposed bootstrap is closely related to the Rao, Hartley Cochran sampling method (Cochran 1977, pp. 266-267). The Rao, Hartley, Cochran sampling method randomly places the frame PSUs into n_h groups. One PSU is independently selected within each group proportional to a measure of size z_i . In the bootstrap procedures described above, within each partition, referenced by g , bootstrap-PSUs are randomly placed into 2 groups, referenced by n_g . Given the above assumptions, the bootstrap procedures can be viewed as similar to as the Rao, Hartley, Cochran procedure, except that the sampling is done systematically instead of independently. By additionally assuming there are enough bootstrap-PSUs in g , so that the correlation of PSUs within g is small enough to be assumed equal to zero; and that PSUs in different g 's have correlation $\rho_{n_h gg'}$, it becomes possible to derive an expression for $V^*(\hat{X}_h^*)$ **.** The n_h in $\rho_{n_h gg'}$ represents the fact that the correlations are average (across all possible orderings) within sample correlations; as such, they will change based on the response variable, the original ordering and the original sample size, n_h .

$$V^*(\hat{X}_h^*) = \sum_g \sum_{g'} \operatorname{cov}(X_g^*, X_{g'}^*),$$

where g and g' represent the original frame partitioning within h,

X_g^* and $X_{g'}^*$ are the estimates within g and g', respectively

$$= \sum_g \operatorname{var}^*(X_g^*) + \sum_{g \neq g'}\sum \operatorname{cov}^*(X_g^*, X_{g'}^*),$$

$$= \sum_g \operatorname{var}^*(X_g^*) + \sum_{g \neq g'}\sum \rho_{n_h gg'} \sqrt{\operatorname{var}^*(X_g^*)\operatorname{var}^*(X_{g'}^*)}$$

$$= \sum_g \{[n_g^* / (N_g^*(N_g^* - 1)] \sum_{i \in g} N_l^*(N_l^* - 1)\} V_z^*(\hat{X}_g^*) +$$

$$\sum\nolimits_{g \neq g'} \sum \rho_{n_h gg'} \sqrt{[n_g^* \sum_{l \in g} N_l^*(N_l^* - 1)/(N_g^*(N_g^* - 1))] V_z^*(\hat{X}_g^*) [n_{g'}^* \sum_{l' \in g'} N_{l'}^*(N_{l'}^* - 1)/(N_{g'}^*(N_{g'}^* - 1))] V_z^*(\hat{X}_{g'}^*)} \qquad (7)$$

where n_g^* is the bootstrap sample size in g,

N_g^* is the number of bootstrap - PSUs in g,

N_l^* is the number of bootstrap - PSUs in the l^{th} bootstrap sampling interval,

$V_z^*(\hat{X}_g^*) = 1/n_g^* \sum_{bj \in g} z_{bj} \left(x_{bj}^* / z_{bj} - X_g^*\right)^2,$

z_{bj} is the $m_{bj} / \sum_{bj \in g} m_{bj} = 1/N_g^*,$

m_{bj} is the measure of size for unit bj,

$X_g^* = \sum_{bj \in g} x_{bj}^* = \hat{X}_g.$

The first term in (7) is the variance of the Rao, Hartley, Cochran estimator. If the $\rho_{n_h gg'}$'s equal zero then the systematic bootstrap variance is the Rao, Hartley, Cochran estimator. Since the N_l^*'s are equal (i.e., $N_l^* = N_g^* / n_g^*$) then the $[n_g^* / (N_g^*(N_g^* - 1))] \sum_{l \in g} N_l^*(N_l^* - 1)$ term above equals $1 - (n_g^* - 1)/(N_g^* - 1)$, a close approximation to the simple random sample finite population correction. Therefore, the proposed bootstrap variance estimator includes a finite population correction that reflects the extra variance reduction due to sampling without replacement. Since the PSUs in a sampling interval change as the sample sizes and/or ordering change, one would not expect the $\rho_{n_h gg'}$'s, in the second term, to be constant, as n_h increases or decreases. They are unknown non-constant functions of n_h. Therefore, one does not expect $V^*(\hat{X}_h^*)$ to be proportional to $1/n_h$. The $1-(n_g^*-1)/(N_g^*-1)$ finite population corrections, also makes this unlikely. Since the bootstrap variance is consistent for the original systematic variance, one would not expect the systematic variance, $V(\hat{X}_h)$, to be proportional to $1/n_h$, as n_h increases or decreases. This is also demonstrated in the simulation of section 2.0. Likewise, since the bootstrap selection is systematic, one would not expect $V^*(\hat{X}_h^*)$ to be proportional to $1/n_h^*$, as n_h^* increases or decreases and n_h fixed.

4.0 Simulation

To demonstrate the advantages of the bootstrap variance estimator, a simulation study is presented comparing BHR and the bootstrap variance estimator. Two thousand simulations are generated using frame variables. The frame is the National Center for Education Statistics' (NCES) Private School Survey (PSS). The PSS is NCES's school frame for private elementary and secondary schools. Three totals (number of schools, number of teachers, and number of students), two averages (average students and average teachers per school), and one ratio (ratio of number of students to number of teachers) are estimated in the simulation. In tables 3-8, estimates are computed by each stratification variable (affiliation, region and school level), as well as one of the sort variables (Urbanicity). The School and Staffing Survey (SASS) sample design is used to select the simulation samples. Relative error, relative mean square error, and coverage rates are generated to evaluate the bootstrap and BHR variance estimator performance.

4.1 Comparison Statistics

In this section, the statistics used to compare the bootstrap and BHR variances are described.

4.1.1 Relative Error

Rel. Error $= (\overline{V}_e(\Theta)^{1/2} / V_t(\Theta)^{1/2} - 1) \cdot 100$

Where: $\overline{V}_e(\Theta)$ is the average of the variance estimates ($V_e(\Theta_s)$) from either the bootstrap or BHR procedure,

(i.e., $\overline{V}_e(\Theta)$ is $1/2000 \sum_{s=1}^{2,000} V_e(\Theta_s)$, Θ_s is the s^{th} simulation estimate of Θ)

$$V_t(\Theta) = 1/1999 \sum_{s=1}^{2,000} (\Theta_s - \overline{\Theta})^2, \ \overline{\Theta} \text{ is } 1/2000 \sum_{s=1}^{2,000} \Theta_s$$

4.1.2 Relative Mean Square Error

Rel. MSE $= \{[VV_e(\Theta) + (\overline{V}_e(\Theta) - V_t(\Theta))^2]^{1/2} / V_t(\Theta)\} \cdot 100$,

Where: $VV_e(\Theta) = 1/1999 \sum_{s=1}^{2,000} (V_e(\Theta_s) - \overline{V}_e(\Theta))^2$.

4.1.3 Coverage Rates

Coverage Rate of $V_e(\Theta) = \left(1/2000 \sum_{s=1}^{2,000} R_s^e \right) \cdot 100$,

Where: $R_s^e(\Theta) = \begin{cases} 1 \text{ if } \Theta_s - 1.96 \cdot V_e(\Theta_s)^{1/2} \leq \Theta \leq \Theta_s + 1.96 \cdot V_e(\Theta_s)^{1/2} \\ 0 \text{ otherwise} \end{cases}$

4.2 SASS Sample Design

The sample frame, used in the simulation, is the list frame component of NCES's Private School Survey (PSS). The list frame is stratified by detailed School Association (19 groups), within Association by Census Region (4 levels), and within Region by school level (elementary, secondary and combined). The school sample is selected using the systematic probability proportionate to size sampling procedure, described in the introduction. The measure of size is square root of the number of teachers in the school. Before sample selection, the school frame is ordered by state, school highest grade, urbanicity, zip code, and school enrollment. To reduce the necessary time to complete 2,000 simulation only one detailed school association is simulated.

4.3 Determining n_h^* for the Bootstrap Variance

As described in section 3.3, the determination of n_h^* requires a simulation study in itself. For each stratum, a series of simulations was done for various n_h^*. The n_h^* that produced the best relative error was used in the simulation presented below. The optimum n_h^* is dependent on the estimate of interest. The optimum n_h^* for estimating numbers of teachers is usually different than the optimum n_h^* for estimating numbers of students or schools. Likewise, the optimum n_h^* for estimating averages or ratios can be different than the total optimums. Each different n_h^* imply a different set of replicate weights. Since we want only one set of replicate weights, a compromise n_h^* is determined that works reasonably well for all estimates. The results presented below use the compromise set of n_h^*. Table 2 presents the values for n_h and n_h^*. Each simulation used in the determination of n_h^* had at least 250 samples.

Table 2.–Original (n_h) and bootstrap (n_h^*) sample size by stratum

Stratum	n_h	n_h^*	Stratum	n_h	n_h^*	Stratum	n_h	n_h^*
01911	14	12	01921	10	5	01931	48	35
01912	16	11	01922	10	8	01932	46	33
01913	52	28	01923	10	10	01933	114	81
01914	34	24	01924	10	10	01934	52	40

4.4 BHR Variances

The r^{th} school half-sample replicate is formed using the usual textbook methodology (Wolter 1985) for establishment surveys with more than 2 units per stratum. This is described in the introduction. Two BHR variance estimates are presented. The first (BHR without FPC Adjustment) is the variance estimates described above. This estimate does not make any type of Finite Population Correction (FPC) adjustments. The second makes a simple FPC adjustment. The second BHR variance estimate (BHR with FPC Adjustment) adjusts the first variance estimator by $1 - P_h$, where P_h is the average of the selection probabilities for the selected units within stratum h .

4.5 Number of Replicates

An important aspect of this analysis is a comparison of the stability of the two variance estimators. To do this, each variance estimator will have the same number of replicate estimates. Since producing bootstrap replicate weights is far more time consuming than producing the BHR replicate weights, it has been decided to use a relatively small number of replicates. Thirty-two replicates have been used in the BHR variances and thirty have been used in the bootstrap variances.

4.6 Results

According to tables 3-8, in terms of extremes, the bootstrap variance estimator is better than either BHR variance estimator with respect to relative error, relative MSE, or coverage rate. The bootstrap relative errors are large in absolute value (greater than 20% or less than –20%) 4 times, while the BHR, with and without FPC adjustment, relative errors are large 17 and 12 times, respectively.

Only 13 of the bootstrap relative MSEs are larger than 50% and only one is greater than 100%. The BHR without FPC adjustment has 31 relative MSEs larger than 50% and 6 greater than 100%. The FPC adjusted BHR has 26 relative MSEs larger than 50% and 5 larger than 100%.

The bootstrap procedure has only 2 high coverage rates (coverage rate greater than 98%) and 2 low coverage rates (coverage rate less than 89%). The bootstrap has 1 coverage rate greater than 99%. The BHR without FPC adjustment has 12 high coverage rates, 1 low coverage rate and 9 larger than 99%. Even with a FPC adjustment, the BHR has 10 high coverage rates, 2 low coverage rate, and 9 coverage rates greater than 99%.

The difference between the bootstrap and BHR is largest for the Urbanicity estimates. For these estimates the BHR relative MSE can be almost 4 times larger than the bootstrap relative error (see tables 4 and 7 Urban). One difference between the Urbanicity and other estimates is the amount of sample size control in the sample design. The Urbanicity sample size is controlled through the sorting. Urbanicity is the third sort variable, so the control on sample size is small. The sample size in all other estimates is directly controlled by the stratification. One possible reason for the big Urbanicity differences is that the bootstrap mimics the sample process better than BHR.

4.7 Conclusion

This paper discussed how BHR can be used to measure the variances from surveys utilizing systematic PPS selection procedures. Two assumptions are necessary: 1) the extra stratification introduced by the variance stratum is sufficient to reflect the systematic process and 2) the variance is inversely proportional to the sample size. In table 1, it has been observed that systematic PPS sampling variances may not be inversely proportional to the sample size. Instead, a large number of times, they are a great deal more efficient and sometimes they are less efficient then the inverse sample size rule would imply. One reason for this is that the clustering induced by systematic sampling changes as the sample size change, which makes the intercluster correlation an unknown function of the sample size. In this situation, the variance may not be inversely proportional to the sample size.

To correct this problem, a bootstrap variance estimator has been introduced which does not make the inverse sample size assumption. Given an appropriate super-population model, the bootstrap procedure produces consistent variance estimates. It has also been demonstrated that the bootstrap procedure adjusts for without replacement sampling. Based on the simulation of the SASS survey design (tables 3-8), the bootstrap variance estimator performs better the BHR with respect to relative error, relative MSE and coverage rates. This is especially true with the Urbanicity estimates. This remains true even after a simple finite population adjustment is made to the BHR. One drawback of the proposed bootstrap procedure is that the determination of an appropriate bootstrap sample size can only be implemented using frame variables. However, with appropriate frame variables, the bootstrap variances are close to unbiased, even when the super-population model assumption fails.

Table 3.—% relative error, % relative mean square error and % coverage rates for the bootstrap and BHR variance estimator for students estimates by affiliation, region, level and urbanicity

Estimate	Bootstrap			BHR without FPC Adjustment			BHR with FPC Adjustment		
	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate
Other Affil.	-0.3	30.0	93.2	12.4	44.1	96.1	9.0	35.3	96.1
Northeast	-10.7	43.7	90.4	1.9	47.7	94.9	-0.7	45.4	94.8
Midwest	-1.9	46.0	93.4	11.2	45.6	97.1	8.1	40.7	97.0
South	15.7	55.8	96.2	25.4	66.4	99.8	20.6	55.0	99.7
West	-2.0	38.6	92.1	12.8	47.9	94.7	9.9	42.8	94.6
Elementary	-12.8	38.0	89.9	-6.3	28.7	93.1	-8.3	29.6	91.9
Secondary	-7.5	48.7	88.6	5.7	45.2	94.6	-6.1	34.9	94.2
Combined	13.3	51.2	96.3	26.0	71.2	96.1	21.9	61.6	96.0
Rural	16.6	54.4	97.1	32.8	89.1	97.3	28.6	78.6	97.1
Suburban	8.0	38.9	94.9	-0.9	26.5	92.4	-3.8	26.0	92.4
Urban	31.6	90.4	98.2	48.0	132.1	100.0	43.7	119.4	100.0

Table 4.—% relative error, % relative mean square error and % coverage rates for the bootstrap and BHR variance estimator for schools estimates by affiliation, region, level and urbanicity

Estimate	Bootstrap			BHR without FPC Adjustment			BHR with FPC Adjustment		
	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate
Other Affil.	-4.5	27.8	93.1	12.7	40.7	97.0	9.7	35.4	97.0
Northeast	4.3	43.6	94.6	10.3	52.3	94.9	8.0	49.0	94.9
Midwest	4.2	42.8	92.9	12.5	51.4	98.3	9.8	46.9	95.7
South	-10.9	32.7	90.7	-6.6	26.5	89.6	-10.3	29.0	89.4
West	-2.4	35.1	92.9	7.8	43.7	92.2	5.1	40.0	92.2
Elementary	1.3	34.9	93.6	16.1	57.0	95.9	14.0	52.9	95.9
Secondary	-2.9	57.0	90.5	26.3	107.1	97.2	14.8	81.9	95.9
Combined	-6.2	29.5	91.2	-1.1	28.2	92.3	-4.2	27.7	92.3
Rural	7.5	36.8	95.7	24.2	71.2	98.7	20.9	63.8	98.7
Suburban	6.5	36.6	95.0	23.1	67.5	97.4	19.9	60.6	97.4
Urban	11.5	43.2	96.1	53.7	147.6	97.5	49.5	135.0	97.5

Table 5.—% relative error, % relative mean square error and % coverage rates for the bootstrap and BHR variance estimator for teachers estimates by affiliation, region, level and urbanicity

Estimate	Bootstrap			BHR without FPC Adjustment			BHR with FPC Adjustment		
	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate
Other Affil.	-6.6	28.8	92.4	-4.5	25.5	92.3	-7.5	26.6	91.0
Northeast	-4.1	34.5	92.5	-8.6	33.3	94.4	-11.0	34.5	94.3
Midwest	-6.2	32.6	90.3	15.5	49.6	98.4	12.4	43.5	97.1
South	-10.5	30.7	89.4	-11.3	27.7	83.7	-14.9	31.9	83.5
West	9.6	52.1	96.1	16.4	60.7	96.4	13.3	54.2	95.1
Elementary	3.9	37.7	93.6	18.1	63.2	97.1	15.6	58.3	97.1
Secondary	17.8	75.5	95.4	31.0	98.3	97.5	17.0	64.9	93.7
Combined	-6.5	29.7	91.9	-7.3	27.6	89.9	-10.3	29.6	89.9
Rural	20.3	60.5	97.5	29.5	79.3	97.4	25.4	69.0	97.3
Suburban	11.1	41.7	96.1	4.8	30.7	92.4	1.6	27.4	92.4
Urban	49.7	139.8	99.3	81.7	244.0	100.0	76.0	222.9	100.0

Table 6.—% relative error, % relative mean square error and % coverage rates for the bootstrap and BHR variance estimator for students per school estimates by affiliation, region, level and urbanicity

Estimate	Bootstrap			BHR without FPC Adjustment			BHR with FPC Adjustment		
	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate
Other Affil.	-2.9	28.3	92.7	11.7	41.2	97.4	8.5	35.7	96.2
Northeast	-2.6	39.7	93.0	2.6	42.8	93.6	0.1	40.5	92.4
Midwest	1.6	39.4	94.0	19.5	63.5	96.0	16.2	56.4	96.0
South	1.6	32.7	93.6	9.4	32.4	96.0	5.1	25.8	94.5
West	-2.1	37.9	92.8	9.2	39.5	96.0	6.4	35.2	95.9
Elementary	-8.7	33.5	91.0	-4.3	27.1	94.5	-6.2	27.5	94.4
Secondary	1.0	53.0	92.2	13.2	60.7	94.8	1.2	41.6	93.4
Combined	5.2	36.7	95.1	17.4	52.2	95.0	13.6	44.4	94.9
Rural	5.8	41.4	94.6	45.7	131.2	99.6	41.3	118.2	99.6
Suburban	-12.9	35.7	89.1	-8.3	32.7	92.1	-10.9	33.9	90.8
Urban	10.3	44.4	95.9	32.1	91.7	97.3	28.2	81.6	97.2

Table 7.—% relative error, % relative mean square error and % coverage rates for the bootstrap and BHR variance estimator for teachers per school estimates by affiliation, region, level and urbanicity

Estimate	Bootstrap			BHR without FPC Adjustment			BHR with FPC Adjustment		
	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate
Other Affil.	-5.8	28.4	92.4	4.0	27.3	95.9	1.0	24.6	95.8
Northeast	2.1	42.1	93.7	0.6	41.2	90.9	-1.7	39.4	90.8
Midwest	-0.7	37.7	92.5	18.2	60.7	99.7	15.0	54.0	99.7
South	-10.9	32.5	89.4	-9.6	28.0	89.4	-13.2	31.4	88.1
West	5.5	41.3	95.1	12.1	45.4	93.6	9.2	40.1	93.6
Elementary	4.6	38.7	94.0	17.7	57.2	97.1	15.3	52.4	97.1
Secondary	8.6	54.2	95.2	29.4	93.3	97.4	16.6	63.8	93.7
Combined	-6.9	29.9	91.6	-4.1	26.2	92.3	-7.2	27.1	91.0
Rural	1.1	37.4	93.2	27.9	83.6	99.6	24.2	74.6	99.6
Suburban	-10.7	34.4	89.7	-2.9	34.9	91.8	-5.5	34.3	91.8
Urban	10.6	44.9	95.5	61.6	177.4	99.8	56.7	161.4	99.8

Table 8.—% relative error, % relative mean square error and % coverage rates for the bootstrap and BHR variance estimator for students/teacher ratio estimates by affiliation, region, level and urbanicity

Estimate	Bootstrap			BHR without FPC Adjustment			BHR with FPC Adjustment		
	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate	Rel. Error	Rel. MSE	Cov. Rate
Other Affil.	-0.3	31.9	94.2	12.3	46.0	96.1	9.1	40.6	94.7
Northeast	-4.3	56.3	91.6	9.3	70.4	94.7	6.7	66.4	94.7
Midwest	-5.5	67.1	91.2	3.1	43.9	95.4	0.5	41.6	94.1
South	6.7	45.2	95.4	3.2	31.5	95.9	-0.8	28.6	93.4
West	-1.0	38.3	93.9	9.8	46.9	97.3	7.1	42.8	97.3
Elementary	-2.5	43.9	93.2	11.8	52.2	99.5	9.5	48.5	99.5
Secondary	-25.3	49.1	81.1	1.2	33.2	94.1	-9.3	32.5	91.4
Combined	9.7	46.2	95.8	16.3	53.0	95.9	12.6	45.7	95.7
Rural	7.0	59.8	95.3	22.7	84.2	99.9	18.9	75.6	99.8
Suburban	1.1	37.6	93.5	18.0	61.6	97.1	14.8	55.3	97.1
Urban	5.4	45.6	94.7	15.0	58.3	93.8	11.6	52.1	93.7

5.0 References

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist*., **9**, 1196-1217.

Chao, M. T. and Lo, S. H. (1985). A bootstrap method for finite populations, *Sankhya* A, **47**, 399-405.

Cochran, W. (1977) *Sampling Techniques*. New York: John Wiley and Sons.

Efron, Bradley (1982). The jackknife, the bootstrap and other resampling plans, *Society for Industrial and Applied Mathematics*, (SIAM) No. 38. Philadelphia, PA.

Gross, S. (1980). Median estimation in sample surveys, *Proceedings of the Section on Survey Research*, pp. 181-184. Alexandria, VA: American Statistical Association.

Kaufman, Steven (1993). A bootstrap variance estimator for the Schools and Staffing Survey, *Proceedings for the Section on Survey Methods, American Statistical Association*, pp. 675-681. Alexandria, VA: American Statistical Association.

McCarthy, P. J., Snowden, C. B. (1985). The bootstrap and finite population sampling, *Vital and Health Statistics*, Public Health Service Publication 85-1369. Washington, DC: U.S. Government Printing Office.

Rao, J. N. K., and Wu, C. F. J. (1988). Resampling inference with complex survey data, *J. Amer. Statist Assoc*., **83**, 231-241.

Shao S. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Sitter, R. R. (1992a). A resampling procedure for complex survey data, *J. Amer. Statist. Assoc*., **87**, 755-765.

Sitter, R. R. (1992b). Comparing three bootstrap methods for survey data, *Canadian J. Statist*., **20**, 135-154.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Appendix

In this section, the consistency of the proposed bootstrap estimators is proven.

Theorem

Given the superpopulation model described in section 3.1, the conditional bootstrap variance estimator of $\hat{X} = \sum_{h} \sum_{j \in h} w_{hj} x_j$, given the model and the sample, is consistent, as $n \to \infty$, provided $\hat{F}_{hc}(y) \to F_{hc}(y)$ and $\mu^*_{yhc} \to \mu_{yhc}$, as $n \to \infty$.

Before the conditional result is proven, additional terms, definitions and results must be stated.

Definitions

Mallows' distance

Let

$$\Im_{\Re^s} = \{\text{all distributions on } \Re^s\}$$
$$\Im_{r,s} = \{G \in \Im_{\Re^s} : \int \|x\|^r dG(x) < \infty\}$$

For two distributions H and G in $\Im_{r,s}$, their Mallow's distance is

$$\tilde{\rho}_r(H,G) = \inf_{\tau_{X,Y}} (E\|X - Y\|^r)^{1/r},$$

where $\tau_{X,Y}$ is the collection of all possible joint distributions of the pairs (X, Y) whose marginal distributions are H and G, respectively. For random U and V having distributions $H \in \Im_{r,s}$ and $G \in \Im_{r,s}$, respectively, define $\tilde{\rho}_r(U,V) = \tilde{\rho}_r(H,G)$.

Let $X_1,\ldots, X_n$ be observations from a stochastic model P_n and $\Re_n = \Re_n(X_1,\ldots,X_n, P_n)$ be a random variable.

The sampling distribution of $\Re_n$ is given by:

$$H_{P_n}(x) = P\{\Re_n \le x | P_n\}$$

The bootstrap sampling distribution of $\Re_n$ is given by:

$H_{\text{BOOT}}(x) = P_* \{\Re_n^* \le x | \hat{P}_n\}$ where $\hat{P}_n$ is an estimator of P_n based on $X_1,\ldots,X_n$ and P_* represents the probability with respect to $\hat{P}_n$. $\Re_n^* = \Re_n(X_1^*,\ldots,X_n^*)$, where $\{X_1^*,\ldots,X_n^*)$ is a bootstrap sample from $\hat{P}_n$.

Define $\mu_{\bar{y}_h h} = \sum_{c\in h} q_{ch}\mu_{Ych}$, where $q_{ch} = n_{ch}/n_h$; and $\mu_{\bar{y}} = \mu_{\hat{X}} = \sum_h p_h \mu_{\bar{y}_h}$. The bootstrap mean ($u^*_{\bar{y}_h h}$ and $u^*_{\hat{X}}$) are defined in an analogous manner where $\hat{P}_n$ is used instead of P_n. $\mu^*_{\bar{y}_h h} = \sum_{c\in h} q_{ch}\mu^*_{Ych} = \bar{y}_h$ and $\mu^*_{\bar{y}} = \mu^*_{\hat{X}} = \sum_h p_h \mu^*_{\bar{y}_h h} = \sum_h p_h \bar{y}_h$.

Results (all from Bickel and Freedman 1981)

Let U and V be random vectors and a be a constant. Then,

$$\tilde{\rho}_r(aU, aV) = |a|\tilde{\rho}_r(U,V) \qquad \text{A.1}$$

If $E\|U\|^2 < \infty$ and $E\|V\|^2 < \infty$, then

$$[\tilde{\rho}_2(U,V)]^2 = [\tilde{\rho}_2(U - EU, V - EV)]^2 + \|EU - EV\|^2 \qquad \text{A.2}$$

Let $\{U_j\}$ and $\{V_j\}$ be two sequences of independent random vectors whose distributions are in $\Im_{r,s}$ and $EU_j = EV_j$ for all j. Then,

$$\left[\tilde{\rho}_2\left(\sum_{j=1}^m U_j, \sum_{j=1}^m V_j\right)\right]^2 \le \sum_{j=1}^m \left(\tilde{\rho}_2(U_j, V_j)\right)^2 \qquad \text{A.3}$$

In the discussion that follows, we will always assume that the partitioning is known and that the selected sample (s_i) is given.

Given the above assumptions, definitions and results, we will investigate the properties of the bootstrap distribution of the total $\hat{X} = \sum_h \sum_{j \in h} w_{hj} x_j$ (i.e., $H_{BOOT}(x)$, where $\Re^* = \sqrt{n}(\hat{X}^* - \hat{X})$), as $n \to \infty$. It will be shown, using Mallow's distance ($\tilde{\rho}_2$), that the bootstrap distribution is consistent. It then follows that the bootstrap variance is consistent (Shao and Tu 1995). The proof closely follows example 3.1 in (Shao and Tu 1995), which was also used by Bickel and Freedman (1981).

There are many ways of generating a bootstrap sample from $\prod \hat{F}_h$. Let $Y^* = \bigcup Y_h^*$ be one such bootstrap, which also defines a bootstrap for X. The important thing is that $\hat{F}_{hc}(y) \to F_{hc}(y)$ and $\mu^*_{hcY} \to \mu_{hcY}$, as $n \to \infty$. Given a method of generating a bootstrap sample, $\hat{X}^* = \sum_h \sum_{j \in Y_h^*} w_{hj} x_j^*$ can be computed.

$$\begin{aligned}
\tilde{\rho}_2(H_{BOOT}, H_n) &= \tilde{\rho}_2(\sqrt{n}(\hat{X}^* - \hat{X}), \sqrt{n}(\hat{X} - \mu_{\hat{X}})) \\
&= \tilde{\rho}_2\left[\sqrt{n}\left(\sum_h p_h(\bar{y}_h^* - \bar{y}_h)\right), \sqrt{n}\left(\sum_h p_h(\bar{y}_h - \mu_{\bar{y}_h h})\right)\right] \\
&\le \sqrt{\sum_h \tilde{\rho}_2\left[\sqrt{n}\left(p_h(\bar{y}_h^* - \bar{y}_h)\right), \sqrt{n}\left(p_h(\bar{y}_h - \mu_{\bar{y}_h h})\right)\right]^2}, \qquad \text{[by A.3]}
\end{aligned}$$

$$\begin{aligned}
\tilde{\rho}_2(H_{BOOT}, H_n)^2 &\le \sum_h \tilde{\rho}_2\left[\frac{1}{\sqrt{n}}\left(n_h(\bar{y}_h^* - \bar{y}_h)\right), \frac{1}{\sqrt{n}}\left(n_h(\bar{y}_h - \mu_{\bar{y}_h h})\right)\right]^2 \\
&= \frac{1}{n}\sum_h \tilde{\rho}_2\left[\left(\sum_{j=1}^{n_h}(y_{jh}^* - \bar{y}_h)\right), \left(\sum_{j=1}^{n_h}(y_{jh} - \mu_{\bar{y}_h h})\right)\right]^2 \qquad \text{[by A.1]} \\
&\le \frac{1}{n}\sum_h \sum_{c=1}^{C_h}\sum_{j=1}^{n_c} \tilde{\rho}_2\left((y_{jhc}^* - \bar{y}_{hc}), (y_{jhc} - \mu_{\bar{y}_{hc} hc})\right)^2 \qquad \text{[by A.3]} \\
&= \frac{1}{n}\sum_h \sum_{c=1}^{C_h}\sum_{j=1}^{n_c} \tilde{\rho}_2\left((y_{1hc}^* - \bar{y}_{hc}), (y_{1hc} - \mu_{yhc})\right)^2 \qquad \text{[by exchangeability]} \\
&= \frac{1}{n}\sum_h \sum_{c=1}^{C_h} n_c \tilde{\rho}_2\left((y_{1hc}^* - \bar{y}_{hc}), (y_{1hc} - \mu_{yhc})\right)^2 \\
&= \frac{1}{n}\sum_h \sum_{c=1}^{C_h} n_c \left[\tilde{\rho}_2\left(y_{1hc}^*, y_{1hc}\right)^2 - \left\|Ey_{1hc} - E_* y_{1hc}^*\right\|^2\right] \qquad \text{[by A.2]} \\
&= \sum_h \sum_{c=1}^{C_h} q_{ch} \left[\tilde{\rho}_2\left(y_{1hc}^*, y_{1hc}\right)^2 - \left\|Ey_{1hc} - E_* y_{1hc}^*\right\|^2\right] \\
&= \sum_h \sum_{c=1}^{C_h} q_{ch} \left[\tilde{\rho}_2\left(\hat{F}_{hc}, F_{hc}\right)^2 - \left\|\mu_{hcY} - \mu_{yhc}^*\right\|^2\right] \qquad \text{[by definition]} \\
&\to \sum\sum q_{ch} o(1) \to o(1) \qquad \text{[by condition of theorem]}
\end{aligned}$$

This establishes the consistency of both $H_{BOOT}(x)$ and the bootstrap variance for $\hat{X}$ given the partitioning and the sample s_i and a realization of $\mathbf{Y}_h$ for all h on s_i. The consistency was established without knowledge of the distribution function F.

Listing of NCES Working Papers to Date

Please contact Ruth R. Harris at (202) 219-1831 (ruth_harris@ed.gov)
if you are interested in any of the following papers

Number	Title	Contact
94-01 (July)	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02 (July)	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03 (July)	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04 (July)	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05 (July)	Cost-of-Education Differentials Across the States	William Fowler
94-06 (July)	Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
94-07 (Nov.)	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01 (Jan.)	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02 (Jan.)	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03 (Jan.)	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95-04 (Jan.)	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings

Number	Title	Contact
95-06 (Jan.)	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08 (Feb.)	CCD Adjustment to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09 (Feb.)	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10 (Feb.)	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11 (Mar.)	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12 (Mar.)	Rural Education Data User’s Guide	Samuel Peng
95-13 (Mar.)	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14 (Mar.)	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15 (Apr.)	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16 (Apr.)	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17 (May)	Estimates of Expenditures for Private K-12 Schools	Stephen Broughman
95-18 (Nov.)	An Agenda for Research on Teachers and Schools: Revisiting NCES’ Schools and Staffing Survey	Dan Kasprzyk
96-01 (Jan.)	Methodological Issues in the Study of Teachers’ Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk

Number	Title	Contact
96-02 (Feb.)	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03 (Feb.)	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04 (Feb.)	Census Mapping Project/School District Data Book	Tai Phan
96-05 (Feb.)	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96-06 (Mar.)	The Schools and Staffing Survey (SASS) for 1998-99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96-07 (Mar.)	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96-08 (Apr.)	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96-09 (Apr.)	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998-99 SASS	Dan Kasprzyk
96-10 (Apr.)	1998-99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96-11 (June)	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96-12 (June)	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96-13 (June)	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96-14 (June)	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-15 (June)	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96-16 (June)	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96-17 (July)	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
96-18 (Aug.)	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
96-19 (Oct.)	Assessment and Analysis of School-Level Expenditures	William Fowler
96-20 (Oct.)	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96-21 (Oct.)	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96-22 (Oct.)	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96-23 (Oct.)	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96-24 (Oct.)	National Assessments of Teacher Quality	Dan Kasprzyk
96-25 (Oct.)	Measures of Inservice Professional Development: Suggested Items for the 1998-1999 Schools and Staffing Survey	Dan Kasprzyk
96-26 (Nov.)	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96-27 (Nov.)	Intersurvey Consistency in NCES Private School Surveys for 1993-94	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-28 (Nov.)	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
96-29 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96-30 (Dec.)	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-01 (Feb.)	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97-02 (Feb.)	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-03 (Feb.)	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97-04 (Feb.)	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-05 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-06 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-07 (Mar.)	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97-08 (Mar.)	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-09 (Apr.)	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97-10 (Apr.)	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year	Dan Kasprzyk
97-11 (Apr.)	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97-12 (Apr.)	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97-13 (Apr.)	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
97-14 (Apr.)	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97-15 (May)	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97-16 (May)	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97-17 (May)	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
97-18 (June)	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97-19 (June)	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97-20 (June)	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97-21 (June)	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
97-22 (July)	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-23 (July)	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97-24 (Aug.)	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97-25 (Aug.)	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97-26 (Oct.)	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
97-27 (Oct.)	Pilot Test of IPEDS Finance Survey	Peter Stowe
97-28 (Oct.)	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97-29 (Oct.)	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97-30 (Oct.)	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97-31 (Oct.)	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97-32 (Oct.)	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97-33 (Oct.)	Adult Literacy: An International Perspective	Marilyn Binkley
97-34 (Oct.)	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97-35 (Oct.)	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97-36 (Oct.)	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-37 (Nov.)	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97-38 (Nov.)	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97-39 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97-40 (Nov.)	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
97-41 (Dec.)	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman
97-42 (Jan. 1998)	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97-43 (Dec.)	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
97-44 (Dec.)	Development of a SASS 1993-94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98-01 (Jan.)	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98-02 (Jan.)	Response Variance in the 1993-94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98-03 (Feb.)	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98-04 (Feb.)	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
98-05 (Mar.)	SASS Documentation: 1993-94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98-06 (May)	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98-07 (May)	Decennial Census School District Project Planning Report	Tai Phan
98-08 (July)	The Redesign of the Schools and Staffing Survey for 1999-2000: A Position Paper	Dan Kasprzyk
98-09 (Aug.)	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
98-10 (Aug.)	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
98-11 (Aug.)	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96-98) Field Test Report	Aurora D'Amico
98-12 (Oct.)	A Bootstrap Variance Estimator for Systematic PPS Sampling	Steven Kaufman