NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report May 1999

Indirect State-Level Estimation for the Private School Survey

U.S. Department of Education
Office of Educational Research and Improvement NCES 1999-351

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report **May 1999**

Indirect State-Level Estimation for the Private School Survey

Beverly D. Causey
Leroy Bailey
U.S. Bureau of the Census

Steven Kaufman
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement **NCES 1999-351**

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
Paul Planchon
Associate 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 20208-5574

May 1999

Suggested Citation
U.S. Department of Education. National Center for Education Statistics. *Indirect State-Level Estimation for the Private School Survey*. NCES 1999-351. By B. D. Causey, L. Bailey, and S. Kaufman. Washington, DC: 1999

Contact:
Steven Kaufman
(202) 219-1337
steve_kaufman@ed.gov

Contents

Page

Acknowledgments iv

1. Introduction 1

2. Current Methodology - Direct Estimation 2

3. Proposed Indirect Estimation 3
   3.1 Derivation of an Overall Adjustment 3
   3.2 Derivation of Subgroup Adjustments 4

4. Adjustments to Regional Totals 10

5. Final Indirect Estimate 11

6. Empirical Results 13

7. Conclusions and Recommendations 15

References 16

Appendix A: Principal Estimates 17

Appendix B: Error Estimation 25

Variance Estimation for Indirect Estimation 26
The First Component 26
The Second Component 27

Acknowledgments

The authors wish to thank all of those who contributed to the preparation of this report. We want to especially express our appreciation to Todd Williams, U.S. Bureau of the Census for providing timely computing support for the comparative analysis; Tina Arbogast, U.S. Bureau of the Census for capably typing the initial draft and several revisions; Dan Kasprzyk and Michael Cohen, National Center for Education Statistics, Easley Hoy and Dennis Schwanz, U.S. Bureau of the Census for their careful review of the manuscript and comments which improved its clarity and cohesiveness. We also want to convey our gratitude to Graham Kalton of Westat, Inc. for his valuable comments and detailed suggestions relating to the general focus and structure of the report. Finally, we acknowledge the very helpful comments provided by the following technical reviewers: Charles Alexander, Graham Kalton, Michael Cohen, Debra Gerald, Thomas Snyder, and Andrew Kolstad. Thanks also to Allison Pinckney and Carol Rohr of Pinkerton Computer Consultants, Inc., for formatting the text and tables in this report in their final form.

1. Introduction

The Private School Universe Survey (PSS) is conducted by the Bureau of the Census, under the sponsorship of the National Center for Education Statistics. It is a mail survey, designed to provide data relating to all private schools in the 50 states and the District of Columbia. Private schools for this purpose are defined as institutions which provide educational services for any of grades 1 to 12, have one or more teachers, are not administered by a public agency, and are not operated in a private home. The survey is in fact a census of private schools. It is conducted biannually and attempts to achieve a complete count of private schools and accompanying counts of their students, teachers, and graduates. The private school register derived from the PSS is updated, prior to the survey mailout, during subsequent administrations of the survey by two sources: a synthesis of association, state and commercial listings of schools, which we will subsequently refer to as the list frame, and an independent listing of private schools included in a sample of geographical areas. The list frame is described in the next paragraph and the area frame (sample) is discussed in the succedent paragraph.

The initial PSS, conducted for the1989–90 school years, was based on the Quality Education Data (QED) Inc. list. This is a commercial list of private schools compiled from various sources. The QED was updated and its listings were supplemented by those of 12 private school associations to comprise the 1989–90 list frame. The number of lists used to update the register of private schools was expanded for the 1991–92 and 1993–94 PSS school years. The list frames included updates of the QED and lists from 26 associations, the 50 states and the District of Columbia, and a private vendor, the Jostens Education Data. (Jackson et al. 1994 and 1995) reported that the supplemental lists expanded the number of schools in the frame by about 4,900 for 1991–92 and 2,300 for 1993–94. Despite these efforts, the private schools' list frame remains incomplete. The most recent estimate of the undercoverage rate for private schools was about 8 percent (Jackson and Frazier 1995), that is about 8 percent of the private schools are not included on the register after the update from the list frame. The list enumeration is therefore supplemented by an area sample designed to identify and represent unlisted private schools in the PSS estimates.

A sample of primary sampling units (PSUs), consisting of a single county or a group of counties, is chosen for the area sample. Therefore our area frame consists of the list of PSUs of which the nation is comprised. The sample facilitates the identification of private schools not included in the list frame. Within each selected PSU a list of private schools is compiled from such sources as telephone books, yellow pages, local government offices, chambers of commerce and religious institutions. This list is merged with the list frame, and therefore represents an expansion of the survey frame to the extent that unlisted schools were detected.

The PSS sample design can readily support the computation of direct survey estimates of the number of private schools and their numbers of students, teachers, and graduates at the national and regional level. These direct survey estimates are obtained in the conventional manner in survey analysis, where sampled schools are weighted up to represent unsampled and nonresponding schools.

While direct estimation, as discussed in section 2, produces estimates of adequate precision for the four geographical regions, the national-level design of the area sample can result in less reliable estimates for individual states. In order to address this problem, the use of indirect estimation methods is recommended. This report describes the development and evaluation of the statistical models used to produce indirect state estimates from the PSS for the 1991–92 and 1993–94 school years.

The statistical models are based on the data obtained from the area sample PSUs. Within these PSUs, data are available for both listed and unlisted private schools. From these data, models can be developed to predict the probability that a school of a given type is included in the list frame. Then for nonsampled PSUs, the listed schools of the designated school type can be weighted up by the inverse of this probability, in order to represent the corresponding unlisted schools in those PSUs. Section 3 describes the models developed for this purpose.

A problem that arises with the use of indirect estimates for relatively small geographical areas is that when the estimates from such areas are added together, the sum will not be consistent with the direct estimate for the combined area. Consequently, the sum of the indirect estimates for the states in a region will in general not equal the direct estimate for the region. This problem is handled by a constrained estimation procedure that adjusts the indirect state estimates so that the resultant estimates for the states in a region sum to the direct regional estimate. The procedure is described in section 4, while final indirect estimates are presented in section 5.

Empirical results are provided in section 6 and appendix A, and conclusions and recommendations are presented in section 7. The estimation of the variance of an indirect estimate is more complex than that of a direct estimate. Variance estimation for the indirect estimates is discussed in appendix B.

2. Current Methodology - Direct Estimation

This section describes the PSS sample design and direct estimation procedures currently used to produce national and regional survey estimates. For the 1991–92 and 1993–94 area surveys, a stratified sample of primary sampling units (PSUs), comprised of counties or groups of counties, was drawn with probability proportional to the square root of the populations of the PSUs. The nation was divided into 16 sampling strata defined by the four principal geographic regions, whether the PSUs were in metropolitan statistical areas (MSAs), and whether they had a “high or low percentage” of private school enrollment. A total of 123 PSUs were selected for the survey; eight of the largest counties were selected with certainty. In each sample PSU up to seven different sources (yellow pages, local government offices, etc.) were used to identify private schools that were not included on the list frame. The area search identified 355 and 421 missing (unlisted) schools in the selected PSUs in 1991–92 and 1993–94, respectively (see Jackson et al. 1994 for further details).

For direct estimation each unlisted school added to the list frame’s total through the area sample is weighted by the reciprocal of its PSU’s selection probability. All list frame schools are included in the PSS, and therefore receive a sampling weight of 1.0. Consequently, the overall weight adjustment for those schools reflects only a noninterview adjustment. An estimated 8 percent of the targeted private schools did not respond for the 1993–94 survey period (U.S. Department of Education 1996). The corresponding rate for 1991–92 was 2 percent. Within the sample PSU the weighted estimate of the number of unlisted schools from the area sample is added to the list frame count. This sum is aggregated over PSUs within the individual states to obtain state totals, and over states to obtain the four regional totals for the number of private schools. Estimates are obtained similarly for the number of students, teachers, and graduates.

Let L and A denote respectively all schools in the list frame and schools in sample PSUs of the area frame, and w_k be the weight for the *kth* school in a given frame. In addition, let u_k be an indicator variable for unlisted schools detected in the area sample. That is u_k will be 1 if the *kth* area sample school was unlisted and 0 if it was included in the list frame. Direct estimates may be obtained in a routine way. The total numbers of students, teachers and graduates from private schools in the nation are estimated by

$$\sum_L w_k y_k + \sum_A w_k u_k y_k, \tag{2.1}$$

where y_k denotes the numbers of students, teachers or graduates in the sampled schools from the two frames. The total number of schools in the nation is obtained by simply setting $y_k = 1$.

The above approach is readily extended to produce estimates for subgroups, such as regions or type of school, by confining the summations to schools in a specified subgroup. While this procedure can be used to provide unbiased estimates for states, the estimates produced in this manner are subject to considerable sampling error. The reason for this lack of precision is that the samples of PSUs for the area frame were not stratified geographically by state but only by region. As a result, the number of PSUs sampled in a state is random. The percentage of sample PSUs in a given region, from a particular state, can differ considerably from the percentage of the total population of the region ascribed to the state. If the number of PSUs sampled in the state is larger than expected, the state estimates will be too large, and if smaller than expected, they will be too small. Examples reflecting this problem will be provided in section 6.

The adoption of a ratio adjustment procedure would seemingly provide a plausible approach to a reduction in the sampling error associated with direct state-level estimates. However, currently there is no apparent data source that could provide adequate counts to serve as the “adjustment standard” for the ratio procedure. The Census of Service Industries would be a likely source; however, the counts of private schools derived from the list are incomplete. For example, schools such as those within churches may not be identified on the educational services establishment list. In addition, enrollment data would not be available, even for those schools included on the list.

We have developed a model-based procedure for state estimation in an effort to improve upon estimates derived from direct estimation. The methodology for these indirect state estimates is described in the next section.

3. Proposed Indirect Estimation

An indirect or synthetic estimator is generally defined as a non-traditional estimator which “borrows strength” from a domain and/or time period, other than those of interest, in deriving desired predictions or estimates. The indirect estimator depends on values of the study variable, introduced through a model, from other domains and/or time periods. A discussion of the use of indirect estimators in federal programs can be found in Statistical Policy Working Paper 21 (1993), prepared by the Federal Committee on Statistical Methodology.

With indirect estimation, as with direct estimation, the PSS sample is treated as being comprised of schools from both the list and area frames. However, the indirect procedure uses the area frame sample to identify schools not included in the list frame, and to establish a basis for data adjustment in nonsampled PSUs to account for the missing schools. The unweighted counts from these unlisted (missed) schools are added to the list frame counts, providing a complete count in sampled PSUs. For nonsampled PSUs, noncoverage adjustment factors, derived from the area sample are applied to the list frame sample to compensate for the unlisted schools.

3.1 Derivation of an Overall Adjustment

The application of the suggested indirect approach requires the specification of a model for noncoverage. The simplest of such models assumes that the unlisted schools are missing completely at random (MCAR). Under this model, the probability that a school is missed or unlisted is the same for every school. This probability may be estimated from the PSS by

$$\hat{p}_T = \frac{\sum_A w_k u_k}{\sum_L w_k + \sum_A w_k u_k}, \quad (3.1)$$

that is, by the ratio of the estimated number of unlisted private schools in the country to the estimated total number of private schools in the country (both listed and unlisted). The undercoverage weighting adjustment that is applied to each listed school in the nonsampled PSUs is $1/\hat{q}_T$, where $\hat{q}_T = 1 - \hat{p}_T$. This undercoverage adjustment is multiplied by each school's nonresponse adjustment factor to give its final weight. Equivalently, for each listed school not in sampled PSUs, there would, on average be p_T/q_T schools added to the count.

3.2 Derivation of Subgroup Adjustments

The MCAR assumption is a stringent one that is unlikely to hold in practice. Coverage can be very different for different domains of the PSS population. This point is illustrated in tables 3.1–3.3, where for the 1991–92 and 1993–94 survey periods, variation is noted among school type and survey item in the percentage contribution of the area frame adds to the given totals. For example, in tables 3.1a and 3.1b we can readily see that the smallest overall percentage of private schools obtained through the area frame search is for Catholic schools, while the largest is for the Non-sectarian schools. Moreover, in table 3.3b we observe that for 1993–94 the corresponding percentages within the Non-sectarian group range from 2.75 to 14.71. Consequently, it seemed desirable to consider the application of undercoverage adjustments for several subgroups of the private school population (where the MCAR assumption may be more plausible) before computing state estimates. Moreover, Jackson et. al. (1995) provide evidence of a significant relationship between school size, as measured by student enrollment, and the probability of the school's inclusion in the original list frame. This led to the fitting of logistic regression models to the 1991–92 and 1993–94 PSS data in the nine domains or subgroups defined by school type. The domains are represented in tables 3.1–3.3.

Table 3.1a—Number of private schools by private school typology and type of frame: 1991–92

Private school type	Total	List frame	Area frame*	Area frame percent total
Total	**25,998**	**23,927**	**2,071**	**7.97**
Catholic	**8,889**	**8,581**	**308**	**3.46**
Parochial	5,485	5,347	138	2.52
Diocesan	2,502	2,399	103	4.12
Private order	901	834	67	7.44
Other religious	**11,760**	**10,718**	**1,042**	**8.86**
Conservative Christian	4,291	3,943	349	8.13
Affiliated	3,950	3,653	297	7.52
Unaffiliated	3,519	3,123	396	11.25
Non-sectarian	**5,349**	**4,628**	**721**	**13.48**
Regular	2,376	2,089	287	12.08
Special emphasis	1,810	1,558	252	13.92
Special education	1,163	980	182	15.65

* Estimated adds or unlisted units based on area sample.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92.

Table 3.1b—Number of private schools by private school typology and type of frame: 1993–94

Private school type	Total	List frame	Area frame*	Area frame percent total
Total	**26,093**	**24,067**	**2,026**	**7.70**
Catholic	**8,331**	**8,261**	**69**	**0.83**
Parochial	5,127	5,099	28	0.54
Diocesan	2,371	2,350	20	0.86
Private order	833	812	21	2.57
Other religious	**12,222**	**10,935**	**1,286**	**10.52**
Conservative Christian	4,530	4,101	429	9.47
Affiliated	3,640	3,403	238	6.53
Unaffiliated	4,051	3,432	620	15.30
Non-sectarian	**5,541**	**4,871**	**671**	**12.10**
Regular	2,198	2,038	160	7.29
Special emphasis	2,106	1,675	431	20.45
Special education	1,237	1,157	80	6.44

* Estimated adds or unlisted units based on area sample.
NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

Table 3.2a—Private school enrollment by private school typology and type of frame: 1991–92

Private school type	Total	List frame	Area frame*	Area frame percent total
Total	**4,745,989**	**4,545,984**	**200,005**	**4.21**
Catholic	**2,523,151**	**2,444,955**	**78,196**	**3.10**
Parochial	1,430,904	1,401,323	29,582	2.07
Diocesan	757,270	727,276	29,994	5.21
Private order	334,977	316,357	18,620	5.56
Other religious	**1,531,486**	**1,447,858**	**83,629**	**5.46**
Conservative Christian	569,203	534,265	34,938	6.14
Affiliated	593,609	573,691	19,918	3.36
Unaffiliated	368,374	339,901	28,773	7.80
Non-sectarian	**691,352**	**653,171**	**38,181**	**5.52**
Regular	466,859	449,266	17,593	3.77
Special emphasis	152,678	141,930	10,748	7.04
Special education	71,815	61,975	9,839	13.70

* Estimated adds or unlisted units based on area sample.
NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92.

Table 3.2b—Private school enrollment by private school typology and type of frame: 1993–94

Private school type	Total	List frame	Area frame*	Area frame percent total
Total	**4,836,442**	**4,705,585**	**130,857**	**2.71**
Catholic	**2,448,101**	**2,474,392**	**13,709**	**0.55**
Parochial	1,409,828	1,403,684	6,144	0.44
Diocesan	751,175	745,658	5,518	0.73
Private order	327,097	325,050	2,047	0.63
Other religious	**1,629,581**	**1,552,700**	**76,880**	**4.72**
Conservative Christian	610,578	588,798	21,780	3.57
Affiliated	593,647	569,103	24,544	4.13
Unaffiliated	425,356	394,800	30,556	7.18
Non-sectarian	**718,761**	**678,493**	**40,268**	**5.60**
Regular	481,423	466,421	15,001	3.12
Special emphasis	163,251	141,037	22,214	13.61
Special education	74,087	71,035	3,052	4.12

* Estimated adds or unlisted units based on area sample.
NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

Table 3.3a—Number of private school teachers by private school typology and type of frame: 1991–92

Private school type	Total	List frame	Area frame*	Area frame percent total
Total	**339,257**	**322,612**	**16,646**	**4.91**
Catholic	**143,214**	**138,499**	**4,715**	**3.29**
Parochial	75,839	74,248	1,592	2.10
Diocesan	42,239	40,705	1,534	3.63
Private order	25,136	23,546	1,590	6.33
Other religious	**117,389**	**110,635**	**6,754**	**5.75**
Conservative Christian	42,176	39,380	2,796	6.63
Affiliated	46,511	44,817	1,694	3.64
Unaffiliated	28,702	26,438	2,263	7.88
Non-sectarian	**78,655**	**73,478**	**5,177**	**6.58**
Regular	48,538	46,494	2,044	4.21
Special emphasis	16,552	15,262	1,290	9.51
Special education	13,564	11,721	1,843	13.59

* Estimated adds or unlisted units based on area sample.
NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92.

Table 3.3b—Number of private school teachers by private school typology and type of frame: 1993–94

Private school type	Total	List frame	Area frame*	Area frame percent total
Total	**378,109**	**364,150**	**13,959**	**3.69**
Catholic	**149,789**	**148,905**	**973**	**0.65**
Parochial	79,736	79,293	443	0.56
Diocesan	44,997	44,633	364	0.81
Private order	25,145	24,980	166	0.66
Other religious	**141,993**	**133,997**	**7,996**	**5.63**
Conservative Christian	51,289	48,750	2,539	4.95
Affiliated	52,237	50,126	2,111	4.04
Unaffiliated	38,467	35,121	3,346	870
Non-sectarian	**86,237**	**81,248**	**4,989**	**5.79**
Regular	51,748	50,326	1,422	2.75
Special emphasis	20,794	17,736	1,422	14.71
Special education	13,695	13,186	509	3.72

* Estimated adds or unlisted units based on area sample.
NOTE: Details may not add to totals due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

3.2.1 Logistic Regression Model

For the *jth* school type, (j = 1,2,...,9) let p_{jk} be the probability that the *kth* school is unlisted, and u_{jk} and w_{jk} be defined analogously to the corresponding terms of section 2. In addition, let x_{jk} be the size of the enrollment of the *kth* school in the cell, and let $c_{jk} = 1/(1-p_{jk})$ denote the related undercoverage adjustment factor. The model given by

$$p_{jk} = 1/[1 + \exp(\alpha_j + \beta_j x_{jk})] \quad \mathbf{(3.2)}$$

relates the "undercoverage proportion" (or the probability that a given school is not listed) to the regressor variable (school size). It can be estimated for area sample schools. Note that p_{jk}, as defined above, properly assumes a value between 0 and 1, and with positive β_j it decreases with increases in x_{jk}. Applying the logit transformation we obtain the following:

$$-\ln\left(\frac{p_{jk}}{1-p_{jk}}\right) = \alpha_j + \beta_j x_{jk} \ . \quad \mathbf{(3.3)}$$

From which we get

$$\hat{p}_{jk} = \frac{1}{[1 + \exp(a_j + b_j x_{jk})]} \ , \quad \mathbf{(3.4)}$$

where a_j, and b_j, are the solutions for α_j and β_j .

Now for school type *j* the adjustment for school *k*, denoted by c_{jk}, is $1 + r_{jk}$, where

$$r_{jk} = \frac{p_{jk}}{1 - p_{jk}} = \frac{1}{\exp(a_j + b_j x_{jk})} \quad . \tag{3.5}$$

This defines the “odds” ratio r_{jk} for the probability of a school being missed in the original list. Therefore for the area frame the fitted number of missed schools can be expressed as

$$\hat{M} = \sum_{!} f_{jk}(1 - u_{jk}) r_{jk} \quad ,$$

where f_{jk} is the noninterview adjustment factor for the *kth* school of the *jth* adjustment cell.

Moreover, the desired solution for the regression coefficients require the satisfaction of the following relationship.

$$\sum_{A} f_{jk}(1 - u_{jk}) r_{jk} = \sum_{A} u_{jk} f_{jk} \tag{3.6}$$

Analogously the fitted number of students missed is denoted by

$$\hat{S} = \sum_{A} f_{jk}(1 - u_{jk}) r_{jk} x_{jk} \quad .$$

and we require that

$$\sum_{A} f_{jk}(1 - u_{jk}) r_{jk} x_{jk} = \sum_{A} u_{jk} f_{jk} x_{jk} \tag{3.7}$$

Equations (3.6) and (3.7) are essentially equivalent to the familiar estimating equations for the iterative least squares algorithm. Estimates derived from these equations were used for schools and students, respectively. With slight modifications in the procedure, it was adapted to teachers and graduates. For teachers within adjustment cell *j* the estimates for the coefficient for the regressor $\$_j$ were retained, and the constant term was refitted to meet the following constraint,

$$\sum_{!} f_{jk}(1 - u_{jk}) r_{jk} x'_{jk} = \sum_{!} f_{jk} u_{jk} x'_{jk} \tag{3.8}$$

where x'_{jk} represents the number of teachers. This constraint assures that the estimated teacher undercount will equal the actual teacher undercount where it is known (i.e., in the sampled area frame PSUs).

For many of the PSS schools there was a relatively small number of current graduates. Therefore, for this variable the adjustment cells were not treated separately; rather, an across-the-board amount was added to each a_j so that the constraint of (3.8) is satisfied in the summation over domains.

The undercoverage adjustments were determined and applied to the listed schools and students in the nonsample PSUs.

Estimates of the regression coefficients of the model were obtained from the SAS iterative reweighted least squares logistic procedure, and are shown below in table 3.4.

Table 3.4—Estimated regression coefficients: SAS logistic regression

	1991–92		1993–94	
Private school type	Constant (a)	School size (b)	Constant (a)	School size (b)
Catholic				
Parochial	2.7734	0.3153	4.2113	0.6150
Diocesan	3.4359	0.0372	4.5449	0.0747
Private order	2.4235	0.0524	1.3205	1.5800
Other religious				
Conservative Christian	2.1196	0.2379	0.9259	1.8500
Affiliated	1.7499	0.7703	2.9301	0.2370
Unaffiliated	1.7407	0.1660	0.9521	1.1120
Non-sectarian				
Regular	0.7373	1.1306	1.5981	0.5850
Special emphasis	1.0747	1.1832	0.8958	0.6420
Special education	1.4580	0.2249	1.8426	2.4300

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92 and 1993–94.

The indirect estimation alternative for the total number of private schools at the state level, denoted by $\hat{t}_s$, can be represented by the following:

$$\hat{t}_s = \sum_{(L \setminus A)_s}\sum f_{jk}(1 + r_{jk}) + \sum_{A_s}\sum f_{jk} = \sum_{(L \setminus A)_s}\sum f_{jk}c_{jk} + \sum_{A_s}\sum f_{jk}, \tag{3.9}$$

where the sums are over the nine adjustment cells and the PSUs within state *s*. The list frame for state *s* minus the listed schools of the area sample is denoted by $(L \setminus A)_s$, while the area sample for the state is indicated by A_s.

3.2.2 Model Assessment

In addition to providing some discrimination relative to the functional relationship between enrollment and undercoverage, the nine school type cells selected for this study represent one of the groupings currently used by NCES in the presentation and analysis of PSS data. In addition, with only a few hundred unlisted schools identified, it would have been very difficult to estimate the regression parameters precisely for stratification requiring too many domains beyond the selected nine. An additional factor which led to the decision to provide estimates for the nine school types was the comparison of measures of the weighted absolute deviation defined below.

$$d = \sum_j\sum_k w_{jk}\left|u_{jk} - \hat{p}_{jk}\right| \tag{3.10}$$

This is the weighted sum of the absolute value of the so-called error term of the model. Intuitively one would expect this measure to compare favorably for alternative groupings of the sample data for estimation purposes. There was no appreciable difference in the value of *d* based on the nine school types used and on an alternative using the 16 sampling strata of the direct estimate, which are defined by region, metropolitan status, and relative size of enrollment.

Relative to the adequacy of the model within the subgroups both parameters were considered significant for all nine school types; however, the Hosmer-Lemeshow goodness of fit statistics included in the SAS analysis, yielded mixed results. For six of the nine school types there was a reasonably good fit. However, for the Conservative Christian, the unaffiliated groups of the other religious category, and the Non-sectarian Special emphasis group, the p-values suggested a lack of fit of the model.

Table 3.5—Hosmer-Lemeshow goodness of fit statistics for logistic regression fit of 1993-94 PSS data

Private school type	GOF statistic[1]	DF[2]	p-value
Catholic - Parochial	7.2988	8	0.5048
Catholic - Diocesan	4.9310	7	0.6684
Catholic - Private	2.4074	6	0.8787
Other religious - Conservative Christian	14.5630	8	0.0682
Other religious - Affiliated	8.1621	8	0.4178
Other religious - Unaffiliated	22.921	8	0.0035
Non-sectarian - Regular	9.2882	8	0.3186
Non-sectarian - Special emphasis	18.065	8	0.0207
Non-sectarian - Special education	9.2525	8	0.3214

[1]Hosmer-Lemeshow Goodness of Fit Statistic.
[2]Degrees of freedom.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

4. Adjustments to Regional Totals

In an effort to achieve greater precision and consistency, the regional totals based on the indirect estimation method were adjusted to those based on direct estimation. This was achieved in the following manner:

The area sample PSUs included both the actual list frame counts and the unweighted "adds" based on an area search for schools that were missed in the list frame.

The regional sum from the area sample PSUs for the indirect method was subtracted from the regional total based on the direct methodology.

The regional scaling factor was defined as the ratio of the difference formed in the previous step and the estimate of the regional total for the nonsampled PSUs based on the indirect methodology.

For each region, the nonsampled PSU component of the indirect estimate was adjusted across the board so that the regional sum equaled the regional estimate for the nonsampled PSUs, based on the current procedure. Thus for the indirect method, the sum over all PSUs within a region equaled that of the direct method.

Analytically we describe the regional adjustment procedure as follows. For school *k* and PSU *i* in a given region, let w_{ik} be the associated sampling weight, f_{ik} the nonresponse adjustment factor, and u_{ik} the indicator variable for an unlisted school. Then the regional total for the direct estimation method is

$$\sum\sum_{A} w_{ik} f_{ik} u_{ik} + \sum\sum_{L} f_{ik} \quad ,$$

where the summation for the first term is over the area sample and that of the second term is over the list frame.

The sample PSU count, based on the indirect methodology, $_{A}$ f_{ik}, is subtracted from the regional total based on the direct estimate, in order to get the direct method's regional total to reflect the nonsampled PSUs only. The nonsampled PSU counts, obtained by the indirect method, are scaled to the regional total for nonsample PSUs for the direct procedure, that is to

$$\sum\sum_{A}(w_{ik} - 1)f_{ik}u_{ik} + \sum\sum_{L\backslash A} f_{ik} \quad .$$

Table 4.1 shown below gives the 1993–94 regional totals based on direct estimation, which includes area sample estimates of the unlisted schools. The scaling factors associated with the proposed procedure are given in table 4.2. A discussion of the error associated with the scaled estimates is given in appendix B.

Table 4.1—Direct estimates of regional totals for 1993–94

Characteristics	Northeast	Midwest	South	West	Total
School	6,183	7,146	7,558	5,207	26,094
Students	1,275,924	1,309,211	1,386,268	865,039	4,836,442
Teachers	94,662	81,862	105,509	56,128	338,161
Graduates	78,926	61,182	69,060	38,112	247,280

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

Table 4.2—Scaling factors for regions

Characteristics	Northeast	Midwest	South	West
Schools	0 .9789	0.9725	1.0518	0.9557
Students	0.9933	0.9813	1.0156	0.9838
Teachers	0.9880	0.9796	1.0199	0.9831
Graduates	1.0147	1.0104	1.0218	1.0240

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

While the estimated factors for schools for the South and West were 1.05 and 0.96, respectively, the other 14 factors were within three percent of 1.00. In fact, the perceived bias associated with four of the estimates was less than 1 percent. However, it is also interesting to observe that all the table entries for the South and for graduates exceeded 1.00, while all of the other entries were less than 1.00.

5. Final Indirect Estimate

The final indirect estimates for 1993–94 are given in table 5.1. The table entries, which are referred to as adjusted indirect estimates are state counts after the application of the regional adjustment procedure. The indicated errors are root mean squared errors. The bias component of the estimates were derived from estimates of bias at the regional level, based on differences between totals, using the current estimation procedure (unbiased) and the proposed indirect method. These regional bias estimates were proportionately allocated, relative to population, across the respective states.

Table 5.1.—Adjusted indirect estimates of private elementary and secondary schools, enrollment, teachers, and high school graduates: 1993–94

State	Number of schools		Enrollment		Teachers		Graduates	
	Total	Root mean square error	Total	Root mean square error	Total	Root mean square error	Total	Root mean square error
Total	**26,093**	**205**	**4,836,442**	**12,875**	**338,162**	**1,319**	**247,278**	**697**
Alabama	354	17	70,764	1,286	5,115	112	3,964	64
Alaska	72	4	6,192	265	510	23	226	10
Arizona	282	18	43,056	1,448	2,933	120	2,412	63
Arkansas	174	11	26,542	740	1,815	56	1,024	19
California	3,082	140	566,723	9,627	35,117	637	24,617	343
Colorado	368	56	50,186	4,533	3,723	273	1,591	169
Connecticut	350	11	69,360	1,133	6,319	136	6,322	88
Delaware	97	6	22,808	625	1,841	88	1,487	35
District of Columbia	86	6	16,276	661	1,595	71	1,074	38
Florida	1,306	73	239,440	4,649	17,262	411	10,008	134
Georgia	536	30	97,192	1,957	8,295	206	5,663	71
Hawaii	130	6	31,239	841	2,202	65	1,938	37
Idaho	85	3	8,473	261	592	21	347	20
Illinois	1,341	38	289,268	5,483	17,334	360	14,655	59
Indiana	667	24	93,664	2,117	6,319	160	4,140	59
Iowa	268	8	50,357	1,049	3,270	75	2,544	36
Kansas	235	16	40,972	3,092	2,655	198	1,669	41
Kentucky	332	18	60,212	1,200	4,011	101	3,013	43
Louisiana	485	26	145,946	2,571	9,356	206	7,996	70
Maine	156	7	17,774	447	1,623	49	1,959	62
Maryland	589	33	117,100	2,495	9,088	236	5,705	75
Massachusetts	625	16	126,040	1,354	11,223	178	10,227	106
Michigan	1,118	35	188,826	3,936	11,445	265	9,035	78
Minnesota	572	19	87,249	1,974	5,713	143	3,493	61
Mississippi	198	5	58,433	761	3,920	55	3,822	49
Missouri	594	14	114,456	1,877	7,534	145	5,597	67
Montana	90	4	9,433	256	717	21	361	13
Nebraska	233	8	39,734	936	2,592	67	1,899	38
Nevada	60	2	10,771	203	661	15	652	27
New Hampshire	142	6	19,138	514	1,826	61	1,777	56
New Jersey	899	23	198,051	2,231	14,443	235	11,193	110
New Mexico	181	9	20,774	565	1,652	50	921	25
New York	1,933	45	469,395	4,016	34,317	481	26,432	163
North Carolina	521	29	71,173	1,464	6,016	150	3,085	55
North Dakota	61	2	7,619	207	535	16	337	9
Ohio	950	12	242,880	1,756	14,582	110	12,369	72
Oklahoma	152	7	23,586	420	1,951	42	1,319	25
Oregon	266	14	35,063	906	2,344	66	1,696	44
Pennsylvania	1,867	49	343,075	3,694	22,031	232	18,458	197
Rhode Island	115	4	23,506	540	1,870	56	1,428	34
South Carolina	320	18	52,608	1,206	4,103	110	2,477	43
South Dakota	107	4	10,218	319	758	25	260	12
Tennessee	466	26	82,373	1,598	6,693	156	5,043	81
Texas	1,239	66	203,090	3,675	16,042	360	7,957	75
Utah	72	5	10,084	378	785	34	599	29
Vermont	97	5	9,584	278	1,009	37	1,133	39
Virginia	532	25	84,273	1,393	7,221	141	4,721	71
Washington	480	17	70,937	1,552	4,699	121	2,713	58
West Virginia	172	10	14,453	380	1,186	36	704	15
Wisconsin	1,001	34	143,966	3,318	9,125	230	5,183	75
Wyoming	40	3	2,109	92	192	11	37	5

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

6. Empirical Results

Table 6.1 shown below presents the original (Listed) counts, the direct estimates, and indirect estimates from the logistic regression model (Logistic) of the number of private schools by state. In addition, for comparison, corresponding indirect estimates were produced by adjusting list frame schools in nonsample PSUs by a undercoverage adjustment based on equation (3.1). This was done for the nine school types (Ratio1) and for quartiles of the school enrollment variable (Ratio2) within school type. The assumption associated with the use of the latter adjustment is that within a given range of the school enrollment variable, the coverage probability is fairly stable. Obviously the three indirect estimates are reasonably close for the individual states, especially the first and the third. The comparison of second and third indirect estimates (Ratio 1 and Ratio 2) permits an assessment of the effect of introducing school enrollment as an additional stratifying variable for the adjustment process.

While the indirect estimates seem quite similar, a comparison between these estimates and the direct estimates shows disparity reflecting the under representation (or over representation) of sample PSUs in the area frame search. For example, there are states such as Indiana and Wisconsin for which there were no sample PSUs in the area frame search, while other states, such as Missouri and Ohio may have been “over represented.”

Table A.1 and A.2 present direct and indirect state estimates of the principal survey items for 1991–92 and 1993–94. To facilitate comparisons, direct estimates of the number of private schools and their standard errors and the corresponding indirect estimates for 1991–92 and 1993–94 are provided in table A.3. The adjusted indirect estimates are given in table A.4. Table A.5 presents a comparison of those results with the direct estimates.

In tables A.1 and A.2 we can readily see sizable differences in the random error of the direct and indirect estimates to which we alluded in earlier discussions. In the last column of table A.3, percentage difference by state between the indirect and direct estimates of the number of private schools can be observed. The range of the differences was between -32 percent and +30 percent in 1991–92 and ±23 percent in 1993–94. However, the standard errors of the direct estimates for which there were sample PSUs were of such magnitude that the corresponding estimates were not considered statistically significant. In table A.4, we see measures of the “root mean squared error” after measures of bias, described in appendix B, were added to the variances corresponding to the error measures of table A.2b. Although there are obvious increases in the size of the error of the indirect estimates, table 5.1 shows that for most (21 of 24) of the states for which there was a sampling error for the direct estimate, the root mean squared error of the adjusted indirect estimate was still less than the sampling error of the direct estimation procedure.

Appendix B includes a derivation of the overall mean absolute error of a survey estimate for the logistic regression and direct estimation methods and the ratio of the respective estimate. The ratio of the error of the logistic regression estimate to that of the direct estimate was 0.6, which indicated that the performance of the indirect estimator was better than the performance of the direct estimates, and suggests that a thorough examination of indirect alternatives is warranted.

Table 6.1—Comparison of list frame counts of the number of private schools with alternative adjusted estimates

State	Listed	Direct	Logistic	Ratio 1	Ratio 2
Total	**24,177**	**26,093**	**26,166**	**26,162**	**26,207**
Alabama	308	410	340	347	339
Alaska	66	66	75	73	75
Arizona	263	263	295	295	296
Arkansas	149	179	165	165	167
California	3,009	3,145	3,224	3,220	3,229
Colorado	279	391	310	309	311
Connecticut	339	360	358	360	358
Delaware	90	90	99	99	99
District of Columbia	80	80	86	88	87
Florida	1,123	1,262	1,242	1,246	1,245
Georgia	457	580	509	514	510
Hawaii	121	121	130	133	130
Idaho	78	78	85	85	85
Illinois	1,333	1,347	1,379	1,374	1,380
Indiana	619	619	686	677	685
Iowa	260	290	276	275	276
Kansas	206	206	219	217	218
Kentucky	296	296	317	315	318
Louisiana	439	458	462	469	463
Maine	140	140	159	157	160
Maryland	522	522	560	566	562
Massachusetts	606	648	638	639	640
Michigan	1,073	1,075	1,150	1,148	1,150
Minnesota	542	542	588	586	587
Mississippi	191	221	201	209	202
Missouri	568	719	603	605	602
Montana	82	82	93	90	94
Nebraska	223	223	240	237	239
Nevada	58	58	61	61	61
New Hampshire	130	130	145	144	145
New Jersey	878	878	918	926	920
New Mexico	166	166	188	184	188
New York	1,865	1,985	1,974	1,977	1,977
North Carolina	444	463	495	493	496
North Dakota	59	59	62	62	62
Ohio	912	1,016	957	961	958
Oklahoma	128	190	147	146	147
Oregon	250	250	278	277	280
Pennsylvania	1,739	1,846	1,901	1,881	1,907
Rhode Island	112	112	117	117	118
South Carolina	275	297	304	307	305
South Dakota	96	96	106	104	107
Tennessee	400	496	443	442	444
Texas	1,025	1,353	1,178	1,185	1,181
Utah	66	66	75	74	76
Vermont	85	85	99	97	100
Virginia	459	515	510	513	512
Washington	433	486	485	484	485
West Virginia	145	145	164	159	165
Wisconsin	954	954	1,029	1,030	1,027
Wyoming	35	35	41	39	42

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

7. Conclusions and Recommendations

An indirect estimation approach is recommended as an alternative to the current procedure for the production of state estimates of the number of private schools in the nation and the associated numbers of students, teachers, and graduates. This procedure borrows strength from the area frame estimates of coverage in deriving “acceptable” and more equitable state estimates. Unless the list frame is complete for a given state, the current estimation procedure necessarily results in biased and highly variable state estimates. However, indirect estimation methods attempt to produce a distribution of the unlisted schools (and therefore of all schools) among the states, which is “close” to the actual distribution of the target population.

Empirical results of this study suggest that undercoverage rates can be successfully modeled from the area sample and used to adjust list frame estimates for survey items. This is very evident from the review of the goodness of fit statistics for six of the selected subgroups. Moreover, relative to the total error associated with state estimates, the indirect procedure showed considerable improvement over the current direct estimation method. The overall estimate of the error of the logistic regression estimator, as measured by mean absolute error, was 40 percent lower than the error for the direct estimator.

While the indirect estimates for the study, based on simple ratio adjustments for undercoverage, compared favorably with those based on the logistic regression model, there is a clear potential for improvement in the model. For example, a geographic variable could possibly be added as a regressor variable. Moreover, school level or program emphasis could be considered as an alternative undercoverage adjustment variable.

The appropriateness of the state estimation methodology under consideration should be evaluated over several survey collection cycles. Moreover, it is suggested that an effort be exerted to identify and ensure the collection of additional data that could define other explanatory variables that might be effective in the modeling of coverage.

References

Federal Committee on Statistical Methodology, Subcommittee on Small Area Estimation (1993) *Indirect Estimators in Federal Programs*, NTIS Document Sales, PB93-209294, Springfield, VA.

Jackson, Betty J. and Frazier, Richard L. (1995) "Improving the Coverage of Private Elementary-Secondary Schools," Proceedings of the Section on Survey Research Methods, American Statistical Association, 143–148.

Jackson, Betty J., Frazier, Richard L., King, Karen E., and Schwanz, Dennis J. (1994) "Improving the Coverage of Private Elementary-Secondary Schools," Proceedings of the Section on Survey Research Methods, American Statistical Association, 833–838.

U.S. Department of Education, National Center for Education Statistics. *Private School Universe Survey, 1993–94*, NCES 96–143, by Stephen Broughman. Washington, DC: 1996.

U.S. Department of Education, National Center for Education Statistics *Private School Universe Survey, 1991–92,* NCES 94–350, by S. Broughman and E. Gerald of NCES and L. Bynum and K. Stoner of Bureau of the Census. Washington, DC: 1994.

APPENDIX A: PRINCIPAL ESTIMATES

Table A.1a—Direct estimates of private elementary and secondary schools, enrollment, teachers, and high school graduates, by state: 1991–92

	Number of schools		Enrollment		Teachers		Graduates	
State	Total	Standard error	Total	Standard error	Total	Standard error	Total	Standard error
Total	**25,998**	**224**	**4,889,545**	**26,741**	**339,267**	**1,829**	**258,095**	**1,979**
Alabama	391	87	69,441	8,390	5,022	540	3,853	311
Alaska	87	25	5,520	534	516	118	178	51
Arizona	254	—	39,460	—	2,771	—	2,039	—
Arkansas	154	—	22,792	—	1,566	—	944	—
California	3,271	133	613,068	16,643	37,861	1,165	27,702	573
Colorado	363	63	57,352	11,374	4,242	893	2,384	664
Connecticut	315	—	67,374	—	5,987	—	6,361	—
Delaware	80	—	22,803	—	1,547	—	1,347	—
District of Columbia	88	9	17,776	322	1,834	61	1,241	—
Florida	1,198	66	205,600	2,988	15,302	358	9,892	125
Georgia	503	32	96,683	4,078	7,838	307	6,070	9
Hawaii	123	—	36,306	—	2,486	—	2,771	—
Idaho	65	—	6,644	—	467	—	317	—
Illinois	1,375	26	301,374	1,158	17,880	211	15,538	26
Indiana	697	89	99,450	7,004	6,762	680	4,303	366
Iowa	269	—	51,431	—	3,408	—	2,386	—
Kansas	203	—	35,077	—	2,347	—	1,468	—
Kentucky	318	—	65,990	—	4,705	—	3,368	—
Louisiana	438	—	139,248	—	8,746	—	7,552	—
Maine	122	—	14,854	—	1,311	—	1,684	—
Maryland	516	—	113,774	—	8,846	—	6,569	—
Massachusetts	655	46	125,006	3,419	10,891	342	10,269	20
Michigan	1,027	14	187,095	710	11,176	100	9,674	—
Minnesota	604	38	93,404	2,401	6,307	284	3,815	163
Mississippi	275	6	58,757	1,377	4,149	53	3,729	313
Missouri	616	46	116,440	1,884	7,950	252	5,857	—
Montana	108	—	9,644	—	766	—	431	—
Nebraska	236	—	39,673	—	2,634	—	1,995	—
Nevada	51	—	8,482	—	486	—	308	—
New Hampshire	181	50	18,712	1,330	1,929	181	1,881	20
New Jersey	956	65	209,913	8,195	15,178	582	13,385	766
New Mexico	186	—	23,236	—	1,813	—	1,045	—
New York	2,058	29	498,668	7,158	35,615	755	28,359	1,552
North Carolina	476	28	63,255	5,224	5,466	418	3,191	407
North Dakota	63	—	7,518	—	535	—	391	—
Ohio	1,096	67	269,064	13,362	15,591	640	12,314	48
Oklahoma	244	105	34,025	9,317	2,521	612	1,480	102
Oregon	282	52	30,918	1,003	2,213	210	1,511	97
Pennsylvania	1,879	53	359,440	6,920	23,127	529	19,634	65
Rhode Island	111	—	21,242	—	1,861	—	1,485	—
South Carolina	307	42	46,086	2,013	3,609	252	2,312	89
South Dakota	106	—	10,539	—	827	—	390	—
Tennessee	474	50	82,969	2,953	6,404	244	4,901	—
Texas	951	13	170,670	472	13,320	67	7,334	13
Utah	56	—	9,836	—	817	—	537	—
Vermont	81	—	8,351	—	913	—	965	—
Virginia	525	51	80,887	1,872	7,115	173	4,536	—
Washington	429	15	66,556	2,798	4,463	190	2,734	54
West Virginia	148	—	12,908	—	1,074	—	646	—
Wisconsin	955	13	142,339	220	8,920	50	5,010	—
Wyoming	27	—	1,840	—	148	—	11	—

— Too few sample cases for a reliable estimate.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92.

Table A.1b—Direct estimates of private elementary and secondary schools, enrollment, teachers, and high school graduates: 1993–94

State	Number of schools		Enrollment		Teachers		Graduates	
	Total	Standard error	Total	Standard error	Total	Standard error	Total	Standard error
Total	**26,093**	**205**	**4,836,442**	**12,875**	**338,162**	**1,319**	**247,278**	**697**
Alabama	410	79	72,630	4,724	5,424	456	4,174	348
Alaska	66	—	5,884	—	476	—	213	—
Arizona	263	—	41,957	—	2,796	—	2,415	—
Arkansas	179	30	29,011	3,995	2,023	335	1,023	—
California	3,145	65	569,062	1,987	35,170	248	24,436	65
Colorado	391	68	53,732	7,798	4,115	632	1,826	283
Connecticut	360	22	70,198	1,875	6,345	125	6,291	46
Delaware	90	—	22,308	—	1,780	—	1,446	—
District of Columbia	80	—	15,854	—	1,544	—	1,054	—
Florida	1,262	83	233,743	3,789	16,842	424	9,820	54
Georgia	580	81	97,726	3,586	8,283	300	5,630	127
Hawaii	121	—	30,537	—	2,144	—	1,886	—
Idaho	78	—	8,019	—	552	—	341	—
Illinois	1,347	12	293,038	794	17,550	70	14,724	98
Indiana	619	—	91,986	—	6,139	—	4,061	—
Iowa	290	30	50,602	211	3,291	34	2,495	—
Kansas	206	—	37,045	—	2,382	—	1,668	—
Kentucky	296	—	58,058	—	3,815	—	2,949	—
Louisiana	458	19	145,512	4,036	9,286	301	7,844	—
Maine	140	—	16,999	—	1,535	—	1,914	—
Maryland	522	—	112,481	—	8,646	—	5,648	—
Massachusetts	648	29	126,744	1,362	11,329	168	10,281	—
Michigan	1,075	—	187,741	—	11,322	—	8,925	—
Minnesota	542	—	86,051	—	5,595	—	3,453	—
Mississippi	221	30	58,655	1,564	3,995	150	3,901	180
Missouri	719	69	117,466	616	7,973	85	5,839	212
Montana	82	—	9,111	—	684	—	355	—
Nebraska	223	—	39,564	—	2,575	—	1,904	—
Nevada	58	—	10,723	—	654	—	646	—
New Hampshire	130	—	18,386	—	1,742	—	1,730	—
New Jersey	878	—	195,921	—	14,281	—	11,025	—
New Mexico	166	—	20,007	—	1,569	—	1,029	—
New York	1,985	59	473,119	4,776	34,771	482	26,625	125
North Carolina	463	18	6,900	1,803	5,746	147	2,983	—
North Dakota	59	—	7,577	—	529	—	332	—
Ohio	1,016	58	246,805	3,480	14,872	306	12,398	172
Oklahoma	190	62	25,837	3,584	2,250	450	1,536	288
Oregon	250	—	34,092	—	2,254	—	1,700	—
Pennsylvania	1,846	54	342,298	4,260	21,880	235	18,532	304
Rhode Island	112	—	23,153	—	1,835	—	1,408	—
South Carolina	297	21	51,600	1,819	3,989	155	2,383	—
South Dakota	96	—	9,575	—	707	—	254	—
Tennessee	496	54	84,538	2,909	6,684	162	4,970	—
Texas	1,353	98	211,337	7,591	16,726	708	8,447	469
Utah	66	—	9,793	—	749	—	590	—
Vermont	85	—	9,107	—	945	—	1,120	—
Virginia	515	55	84,438	4,584	7,391	621	4,580	—
Washington	486	53	70,205	1,858	4,798	348	2,644	—
West Virginia	145	—	13,539	—	1,085	—	672	—
Wisconsin	954	—	141,762	—	8,927	—	5,129	—
Wyoming	35	—	1,919	—	167	—	31	—

— Too few sample cases for a reliable estimate.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

Table A.2a—Indirect estimates of private elementary and secondary schools, enrollment, teachers, and high school graduates: 1991–92

	Number of schools		Enrollment		Teachers		Graduates	
State	Total	Standard error	Total	Standard error	Total	Standard error	Total	Standard error
Total	**25,879**	*	**4,868,329**	*	**337,458**	*	**256,050**	*
Alabama	316	7	63,192	1,070	4,691	97	3,698	75
Alaska	68	3	5,368	289	433	21	136	10
Arizona	282	13	41,893	1,649	2,962	117	2,123	105
Arkansas	167	6	24,009	808	1,658	54	980	40
California	3,168	29	590,755	3,710	36,173	249	25,524	211
Colorado	301	6	42,560	716	3,177	55	1,833	402
Connecticut	334	11	69,877	1,704	6,261	162	6,530	158
Delaware	87	4	23,674	1,086	1,611	76	1,384	84
District of Columbia	79	—	16,552	—	1,660	—	1,125	—
Florida	1,200	25	208,419	3,260	15,430	247	9,935	205
Georgia	517	11	96,581	1,549	7,882	135	6,243	122
Hawaii	134	7	37,654	1,424	2,586	97	2,825	97
Idaho	72	3	7,091	293	502	21	331	21
Illinois	1,399	13	305,386	2,053	18,024	126	15,707	157
Indiana	658	13	96,285	1,563	6,385	111	4,068	92
Iowa	284	5	52,808	864	3,501	57	2,434	74
Kansas	219	5	36,748	926	2,477	62	1,528	67
Kentucky	340	7	68,449	761	4,883	63	3,487	41
Louisiana	464	8	143,547	1,188	9,058	88	7,788	61
Maine	133	6	15,559	433	1,388	45	1,722	62
Maryland	552	13	115,982	1,672	9,042	148	6,526	71
Massachusetts	640	10	124,857	1,112	10,958	124	10,421	100
Michigan	1,072	18	191,489	1,779	11,401	127	9,812	75
Minnesota	593	12	92,631	1,213	6,156	94	3,664	60
Mississippi	190	4	54,114	742	3,702	51	3,485	49
Missouri	591	12	118,927	1,544	8,008	118	6,029	67
Montana	122	4	10,409	239	832	19	453	13
Nebraska	255	6	41,484	611	2,764	45	2,074	38
Nevada	55	2	8,903	181	513	14	319	26
New Hampshire	148	5	18,469	501	1,866	58	1,920	56
New Jersey	938	14	206,827	1,805	15,110	161	13,031	103
New Mexico	207	7	24,523	530	1,923	46	1,081	24
New York	2,102	19	500,658	2,513	35,588	250	27,617	131
North Carolina	486	11	60,456	939	5,277	88	2,848	53
North Dakota	67	3	7,910	288	567	21	407	28
Ohio	964	13	248,431	2,880	14,436	176	12,192	209
Oklahoma	166	4	27,056	679	2,084	50	1,427	49
Oregon	258	7	31,849	891	2,154	59	1,488	65
Pennsylvania	1,950	38	364,431	6,118	23,584	397	20,178	468
Rhode Island	119	6	22,189	923	1,961	92	1,542	94
South Carolina	279	7	45,556	872	3,442	64	2,301	48
South Dakota	119	4	11,238	379	889	34	402	21
Tennessee	445	11	83,643	2,028	6,439	153	5,050	130
Texas	997	16	176,414	2,491	13,799	204	7,545	148
Utah	62	3	9,545	448	751	31	516	41
Vermont	93	4	8,929	294	1,002	42	997	36
Virginia	526	11	82,801	1,344	7,315	119	4,661	104
Washington	456	14	67,392	1,925	4,539	135	2,790	117
West Virginia	161	4	13,584	371	1,141	31	672	23
Wisconsin	1,013	19	149,246	2,738	9,313	164	5,189	158
Wyoming	31	2	1,979	116	160	9	12	1

* These standard errors were never computed because of technical complexity.

— Too few sample cases for a reliable estimate.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92.

Table A.2b—Indirect estimates of private elementary and secondary schools, enrollment, teachers, and high school graduates: 1993–94

	Number of schools		Enrollment		Teachers		Graduates	
State	Total	Standard error	Total	Standard error	Total	Standard error	Total	Standard error
Total	**27,146**	*	**4,848,186**	*	**338,962**	*	**243,139**	*
Alabama	340	7	69,914	939	5,035	76	3,879	62
Alaska	755	3	6,285	252	518	22	221	10
Arizona	295	14	43,765	1,290	2,983	111	2,355	55
Arkansas	165	6	26,134	603	1,779	42	1,002	19
California	3,224	34	576,047	2,968	35,718	241	24,040	104
Colorado	310	6	45,605	494	3,449	44	1,554	26
Connecticut	358	8	69,828	1,040	6,396	114	6,230	86
Delaware	99	5	22,831	614	1,841	87	1,455	35
District of Columbia	86	6	16,292	650	1,595	69	1,051	38
Florida	1,242	25	235,759	2,721	16,924	219	9,794	126
Georgia	509	11	95,698	1,217	8,133	121	5,542	66
Hawaii	130	6	31,315	852	2,208	66	1,893	36
Idaho	85	3	8,372	246	583	19	339	19
Illinois	1,379	10	294,776	935	17,694	68	14,504	45
Indiana	686	16	95,447	1,214	6,451	96	4,097	58
Iowa	276	4	51,307	490	3,338	36	2,518	36
Kansas	219	4	37,873	569	2,457	42	1,652	32
Kentucky	317	7	59,308	761	3,934	63	2,949	41
Louisiana	462	8	143,710	1,188	9,173	88	7,825	61
Maine	159	6	17,894	433	1,643	45	1,931	62
Maryland	560	13	115,300	1,672	8,910	148	5,583	71
Massachusetts	638	10	126,830	1,112	11,352	124	10,079	100
Michigan	1,150	18	192,422	1,779	11,683	127	8,942	75
Minnesota	588	12	88,854	1,213	5,825	94	3,457	60
Mississippi	201	4	58,325	742	3,937	51	3,740	49
Missouri	603	12	115,584	1,544	7,623	118	5,539	67
Montana	93	4	9,534	239	726	19	353	13
Nebraska	540	6	40,467	611	2,644	45	1,879	38
Nevada	61	2	10,871	181	668	14	637	26
New Hampshire	145	5	19,267	501	1,848	58	1,751	56
New Jersey	918	14	199,388	1,805	14,619	161	11,031	103
New Mexico	188	7	20,993	530	1,672	46	899	24
New York	1,974	19	472,563	2,513	34,735	250	26,049	131
North Carolina	495	11	70,079	939	5,898	88	3,019	53
North Dakota	62	2	7,755	161	546	13	334	9
Ohio	957	10	243,948	1,437	14,623	104	12,242	72
Oklahoma	147	4	23,335	328	1,925	31	1,291	24
Oregon	278	8	35,627	728	2,383	54	1,656	39
Pennsylvania	1,901	37	344,859	3,263	22,240	251	18,191	190
Rhode Island	117	4	23,597	536	1,883	55	1,407	34
South Carolina	304	8	51,799	870	4,022	72	2,424	41
South Dakota	106	4	10,025	261	743	20	257	12
Tennessee	443	9	81,106	934	6,562	79	4,935	78
Texas	1,178	16	199,967	1,829	15,728	161	7,787	66
Utah	75	4	10,245	348	798	32	585	29
Vermont	99	4	9,648	273	1,022	35	1,117	39
Virginia	510	11	83,416	1,070	7,125	100	4,620	70
Washington	485	17	71,172	1,560	4,700	123	2,649	59
West Virginia	164	5	14,231	301	1,163	26	689	14
Wisconsin	1,029	21	146,707	1,978	9,315	135	5,130	74
Wyoming	41	3	2,112	94	192	11	36	5

* These standard errors were never computed because of technical complexity.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

Table A.3a—Direct and Indirect estimates of private elementary and secondary schools and standard errors: 1991–92

	Direct estimation		Indirect estimation		Percentage difference
State	Estimate	Standard error	Estimate	Standard error	$\left(\frac{\text{Indirect} - \text{Direct}}{\text{Direct}} \times 100\right)$
Total	**25,998**	**224**	**25,790**	*	**-0.86**
Alabama	391	87	316	7	-19.18
Alaska	87	25	68	3	-21.84
Arizona	254	—	282	13	11.02
Arkansas	154	—	167	6	8.44
California	3,271	133	3,168	29	-1.97
Colorado	363	63	301	6	-17.08
Connecticut	315	—	344	11	6.03
Delaware	80	—	87	4	8.75
District of Columbia	88	9	79	—	-10.23
Florida	1,198	66	1,200	25	0.17
Georgia	503	32	517	11	2.78
Hawaii	123	—	134	7	8.94
Idaho	65	—	72	3	10.77
Illinois	1,375	26	1,399	13	17.45
Indiana	697	89	658	13	-5.60
Iowa	269	—	284	5	-1.73
Kansas	203	—	219	5	7.88
Kentucky	318	—	317	7	-0.31
Louisiana	438	—	462	8	5.48
Maine	122	—	159	6	30.33
Maryland	516	—	560	13	8.53
Massachusetts	655	46	638	10	-2.59
Michigan	1,027	14	1,150	18	2.24
Minnesota	604	38	588	12	-2.65
Mississippi	275	6	201	4	-26.91
Missouri	616	46	603	12	-2.11
Montana	108	—	93	4	-13.89
Nebraska	236	—	240	6	1.69
Nevada	51	—	61	2	19.61
New Hampshire	181	50	145	5	-19.89
New Jersey	956	65	915	14	-4.29
New Mexico	186	—	188	7	1.08
New York	2,058	29	1,974	19	-4.08
North Carolina	476	28	495	11	3.99
North Dakota	63	—	67	3	6.35
Ohio	1,096	67	964	13	-12.04
Oklahoma	244	105	166	4	-31.97
Oregon	282	52	258	7	-8.51
Pennsylvania	1,879	53	1,950	38	3.78
Rhode Island	111	—	119	6	7.21
South Carolina	307	42	279	7	-9.12
South Dakota	106	—	119	4	12.26
Tennessee	474	50	445	11	-6.12
Texas	951	13	997	16	4.84
Utah	56	—	62	3	10.71
Vermont	81	—	93	4	14.82
Virginia	525	51	526	11	0.19
Washington	429	15	456	14	6.29
West Virginia	148	—	161	4	8.78
Wisconsin	955	13	1,013	19	6.07
Wyoming	27	—	31	2	14.81

*This number was never computed because of technical complexity.

— Too few sample cases for a reliable estimate.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1991–92.

Table A.3b—Direct and indirect estimates of private elementary and secondary schools and standard errors: 1993–94

State	Direct estimation: Estimate	Direct estimation: Standard error	Indirect estimation: Estimate	Indirect estimation: Standard error	Percentage difference $\left(\frac{\text{Indirect} - \text{Direct}}{\text{Direct}} \times 100\right)$
Total	**26,093**	**205**	**26,166**	*	**0.28**
Alabama	410	79	340	7	-17.07
Alaska	66	—	75	3	13.64
Arizona	263	—	295	14	12.17
Arkansas	179	30	165	6	-7.82
California	3,145	65	3,224	34	2.51
Colorado	391	68	310	6	-20.72
Connecticut	360	22	358	8	-0.56
Delaware	90	—	99	5	10.00
District of Columbia	80	—	86	6	7.50
Florida	1,262	83	1,242	25	-1.58
Georgia	580	81	509	11	-12.24
Hawaii	121	—	130	6	7.44
Idaho	78	—	85	3	8.97
Illinois	1,347	12	1,379	10	2.38
Indiana	619	—	686	16	10.82
Iowa	290	30	276	4	-4.83
Kansas	206	30	219	4	5.94
Kentucky	296	—	317	7	7.09
Louisiana	458	19	462	8	0.87
Maine	140	—	159	6	13.57
Maryland	522	—	560	13	7.28
Massachusetts	648	29	638	10	-1.54
Michigan	1,075	—	1,150	18	6.98
Minnesota	542	—	588	12	8.49
Mississippi	221	30	201	4	-9.05
Missouri	719	69	603	12	-16.13
Montana	82	—	93	4	13.42
Nebraska	223	—	240	6	7.62
Nevada	58	—	61	2	5.17
New Hampshire	130	—	145	5	11.54
New Jersey	878	—	918	14	4.56
New Mexico	166	—	188	7	13.25
New York	1,985	59	1,974	19	-0.55
North Carolina	463	18	495	11	6.91
North Dakota	59	—	62	2	5.08
Ohio	1,016	58	957	10	-10.14
Oklahoma	190	62	147	4	-22.63
Oregon	250	0	278	8	11.20
Pennsylvania	1,846	54	1,901	37	2.98
Rhode Island	112	—	117	4	4.46
South Carolina	297	21	304	8	2.36
South Dakota	96	—	106	4	10.42
Tennessee	496	54	443	9	-10.69
Texas	1,353	98	1,178	16	-12.93
Utah	66	—	75	4	22.73
Vermont	85	—	99	4	16.47
Virginia	515	55	510	11	-0.97
Washington	486	53	485	17	-0.21
West Virginia	145	—	164	5	13.10
Wisconsin	954	—	1,029	21	7.86
Wyoming	35	—	41	3	17.14

*This number was never computed because of technical complexity.

— Too few sample cases for a reliable estimate.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

Table A.4—Direct and adjusted indirect estimates of private elementary and secondary schools and associated errors: 1993–94

State	Direct estimation		Adjusted indirect estimation		Percentage difference
	Estimate	Standard error	Estimate	Root mean squared error	$\left(\frac{\text{Indirect} - \text{Direct}}{\text{Direct}} \times 100\right)$
Total	**26,093**	**205**	**26,098**	*	**0.03**
Alabama	410	79	354	17	-13.66
Alaska	66	—	72	4	9.09
Arizona	263	—	282	18	7.22
Arkansas	179	30	174	11	-2.79
California	3,145	65	3,082	140	-2.00
Colorado	391	68	368	56	-5.88
Connecticut	360	22	350	11	-2.78
Delaware	90	—	97	6	7.78
District of Columbia	80	—	86	6	7.50
Florida	1,262	83	1,306	73	3.49
Georgia	580	81	536	30	-7.59
Hawaii	121	—	130	6	7.44
Idaho	78	—	85	3	8.97
Illinois	1,347	12	1,341	38	-0.45
Indiana	619	—	667	24	7.75
Iowa	290	30	268	8	-7.59
Kansas	206	30	235	16	14.08
Kentucky	296	—	332	18	12.16
Louisiana	458	19	485	26	5.90
Maine	140	—	156	7	11.43
Maryland	522	—	589	33	12.84
Massachusetts	648	29	625	16	-3.55
Michigan	1,075	—	1,118	35	4.00
Minnesota	542	—	572	19	5.54
Mississippi	221	30	198	5	-10.41
Missouri	719	69	594	14	-17.39
Montana	82	—	90	4	9.76
Nebraska	223	—	233	8	4.49
Nevada	58	—	60	2	3.45
New Hampshire	130	—	142	6	9.23
New Jersey	878	—	899	23	2.39
New Mexico	166	—	181	9	9.04
New York	1,985	59	1,933	45	-2.62
North Carolina	463	18	521	29	12.53
North Dakota	59	—	61	2	3.39
Ohio	1,016	58	950	12	-6.50
Oklahoma	190	62	152	7	-20.00
Oregon	250	—	266	14	6.40
Pennsylvania	1,846	54	1,867	49	1.14
Rhode Island	112	—	115	4	2.68
South Carolina	297	21	320	18	7.74
South Dakota	96	—	107	4	11.46
Tennessee	496	54	466	26	-6.05
Texas	1,353	98	1,239	66	-8.43
Utah	66	—	72	5	9.09
Vermont	85	—	97	5	14.12
Virginia	515	55	532	25	3.30
Washington	486	53	480	17	-1.23
West Virginia	145	—	172	10	18.62
Wisconsin	954	—	1,001	34	4.93
Wyoming	35	—	40	3	14.29

*This number was never computed because of technical complexity.

— Too few sample cases for a reliable estimate.

NOTE: Details may not add to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Private School Survey, 1993–94.

APPENDIX B: ERROR ESTIMATION

Variance Estimation for Indirect Estimation

For the indirect estimates of state totals based on the logistic model, there are two components of variance. The first corresponds to sampling error in the estimation of parameters, the second to variability around the expected value of the unknown number of unlisted schools in nonsample PSUs. As it turns out, the second component is significantly larger than the first. The use of the bootstrap procedure or another replication method should be considered for future variance estimation for indirect PSS estimates. However, for the 1991–92 and 1993–94 survey data, the adapted estimation methodology is described below.

The First Component

The first component of variability arises from sampling error in the estimation of the coefficients of α_j and β_j that correspond to each of the nine selected school types. From Section 3.2.1 we saw that the proposed estimator for the total number of schools at the state level was

$$\hat{t}_s = \sum\sum_{(L\setminus A)_s} f_{jk}(l + r_{jk}) + \sum\sum_{A_s} f_{jk} , \qquad (1.1)$$

where $r_{jk} = \dfrac{1}{\exp(a_j + b_j x_{jk})}$

The accurate variance estimation related to the use of the estimator given in (1.1) can be cumbersome. Therefore, it was approximated by a closely related model, namely, one which is simply based on an adjustment factor for each school type irrespective of school size. As mentioned above, the first component is small relative to the second; moreover, the role of the 9 adjustment cells is somewhat greater than that of school size. For each adjustment cell a simple ratio estimator was derived: the ratio of the sum of added schools in the sample PSUs, to the corresponding sum of original list schools. For each listed school in the nonsample PSUs this ratio estimator provided an estimate of the corresponding number of unlisted schools. Therefore the estimated number of added schools for a state would have a sampling error which is approximated by the estimated error associated with the adjustment cell ratio estimator.

To estimate the variance the first-order Taylor approximation was employed. For adjustment cell *j* let $\hat{Y}_j$ be the sample based estimator of the total number of unlisted schools and $\hat{X}_j$ be likewise for original list schools. Then set $\hat{R}_j = \dfrac{\hat{Y}_j}{\hat{X}_j}$. For purposes of variance estimation $\hat{R}_j$ was "linearized", that is the behavior of $\hat{R}_j$ was approximated by

$$(\hat{Y}_j - R_j\hat{X}_j)/X_j \qquad (1.2)$$

where R_j is the true ratio and X_j is the total number of original list schools. Meanwhile for a particular state let C_j be the (known) number of nonsample PSU, original list schools in the *jth* cell. The estimated number of unlisted schools is therefore $\sum_j C_j\hat{R}_j$. The approximate variance of this estimate is based on substitution of (1.2) for $\hat{R}_j$.

Thus the variance was found for the sum over sample PSUs of the weighted quantities

$$z_i = \sum_j c_j (y_{jk} - R_j x_{jk}) / X_j \qquad (1.3)$$

with i denoting sample PSU i, y_{jk} and x_{jk} the unlisted and original list schools in PSU i, and where R_j and X_j are viewed as fixed (and, in practice, estimated from the sample). Depending on the sample design, the variance of the estimated number of added schools for a given state, is estimated by the variance of the sum of the z_i in (1.3).

The description above was related to the variance estimation for schools. The procedure can be appropriately adapted for students, teachers, and graduates.

The principal objective again is the estimation of variance for the estimated number of added schools that are obtained by using size of school along with adjustment cell information. Observing that variances often tend to be proportional to magnitudes of expected values, it is assumed that variance based on adjustment cell and school enrollment can be approximated by the variance based on the adjustment cell alone, multiplied by the ratio of the estimated number of adds based on adjustment cell and size to that based on adjustment cell alone.

For each sampling stratum the variance of the sum of quantities z_i in (1.3) was estimated using a variance approximation formula provided by Hartley, Rao and Kiefer (1962). For a particular sampling stratum let n be the number of sample PSUs and w_i^{-1} be the inclusion probability for the *ith* PSU. For the variance of $\sum_{i=1}^{n} z_i$ the following estimator was used.

$$\frac{1}{n-1} \sum_{i<j} [1 - (w_i^{-1} + w_j^{-1}) + \frac{\hat{U}}{n}](w_i z_i - w_j z_j) \quad , \qquad (1.4)$$

where z_i and z_j are unweighted values for PSUs i and j and $\hat{U}$ equals the sum $\sum \frac{1}{w_i}$ over sample PSUs. The Hartley, Rao and Kiefer formula uses U equal to the sum of $(\frac{1}{w_i})^2$ over all PSUs. Here $\hat{U}$ is substituted , which is a convenient sample based Horvitz-Thompson estimator. The bracketed factor in (1.4) is a form of finite population correction.

The Second Component

The second component of variance arises from the variability around the expected value of the actual unknown number of schools, students, teachers, and graduates, that would be added in the nonsample PSUs for each state. Such variability exists irrespective of the sample size and how precisely the parameters are estimated. However, as the sample size gets larger, there are fewer nonsample PSUs to be concerned about.

The first and second components of variance can be viewed as stochastically independent. Therefore the total variance can be estimated as the sum of the two components.

The second component can be estimated as follows. For each list frame school there is an associated expected number of unlisted schools. This expected number is always closer, usually much closer, to 0 than 1. Since the probability is very small that there is a single unlisted school that corresponds to the list frame, the probability that there are two or more such schools is negligible. This conceptual unlisted school would be found if a 100 percent area frame survey were conducted. For the unknown random number of unlisted schools, always 0 or 1, the variance is pq with $q=1-p$.

Suppose also that within a county there is a fixed correlation, the same for all counties, between the outcomes for two schools, and correlation is 0 between the outcomes for two schools in different counties. That is, suppose that for list frame school *i* we let z_i denote the number of added schools, 0 or 1, with expected value p_i and variance $p_i q_i$. For school *k* we do likewise. If schools *i* and *k* are in different counties, the covariance between z_i and z_k is 0. If they are in the same county, then

$$Cov(z_i, z_k) = R(p_i q_i p_k q_k)^{1/2} \quad , \tag{1.5}$$

where *R* is the correlation coefficient. Thus if the value of *R* is obtained in (1.5), the second component of variance can be estimated from the relationship

$$var(\sum z_k) = \sum Var(z_k) + \sum_{i<k} Cov(z_i,\ z_k) \tag{1.6}$$

R is estimated using data from the sample PSUs. For 1991 and 1993 the estimated values of *R* were 0.0767 and 0.1069 respectively. In obtaining these values the following further minor adjustments were made. The squared differences between actual and fitted numbers of added schools were reduced by the fact that the area frame schools were used to estimate and fit 18 parameters, 2 for each of 9 adjustment cell. The sum of the actual squared differences were multiplied by *n/(n-18)* as a way of compensating. A measure of the “effective” number of schools is

$$g = \frac{(\sum w_k)^2}{\sum w_k^2} \tag{1.7}$$

where w_k is the weight assigned to school. This measure may be motivated as follow. Suppose $x_1,...,x_n$ is a series of mutually independent, identically distributed random variables, each with variance *V*. Let $\bar{x}$ be the weighted mean and

$$(\sum w_k x_k)\ \sum w_k \quad . \tag{1.8}$$

Then, Var ($\bar{x}$) is not *V/n*, but *V/g*. The value of *g* was 2234.52 for 1991 and 1841.53 for 1993, both large relative to 18, the number of fitted parameters.

The two components of variance were combined to form a total variance for counts of schools, students, teachers and graduates.

Absolute Errors

The expected absolute errors were compared for the current and the proposed method. To facilitate the discussion let *l* and *k* denote PSU and school within PSU, respectively. In addition, let

y_l be the sum of noninterview adjustments factors for the original list schools;
A_l be the sum for unlisted schools for the PSU; and
w_l be the sampling weight.

The probability of inclusion into the sample will be denoted by $h_l = 1/w_l$. Now the "actual" number of schools in county l may be expressed as $T_l = y_l + A_l$.

For the direct estimation procedure, PSU l $y_l + w_l A_l$ would be the estimated number of schools with probability h_l and y_l would be the estimate with probability $1\text{-}h_l$. The expected absolute error is

$$h_l(w_l - 1)A_l + (1 - h_l)A_l = 2(1 - h_l)A_l \qquad \textbf{(1.9)}$$

For the proposed indirect method, T_l would be the estimate used with probability h_l. With probability $1\text{-}h_l$ T_l would be the estimate, where

$$T_l = y_l + \sum_k r_{lk} y_{lk} \qquad \textbf{(1.10)}$$

Here r_{lk} is the adjustment factor for unlisted schools that are associated with list frame school k in (nonsample) PSU l. Consistent with previous results, this factor is:

$$r_{lk} = 1 \exp(a_j + b_j x_{lk}), \qquad \textbf{(1.11)}$$

with a_j and b_l corresponding to the schools's poststratum "j" and x_{lk} the school's enrollment. For a_j and b_l the sample-based estimators were used.

To get the absolute error of the proposed method, the probability that the school's county is non-sample is multiplied by the absolute error. That is

$$(1 - h_l)|T_l - T_l| \;. \qquad \textbf{(1.12)}$$

Therefore the desired comparison is between (1.9) and (1.12) summed over all PSUs. It was noted that $w_l(1\text{-}h_l) = w_l\text{-}1$ and the applicable comparison was made between $2\sum (w_l - 1)A_l$ and $\sum (w_l - 1)|T_l - T_l|$, with summation over sample PSUs. The ratio of the second term (indirect method) to the first (direct method) was 0.6032. On the basis of this, the proposed indirect estimation was considered decidedly preferable to the current procedure.