AREA FRAME ESTIMATORS IN AGRICULTURAL SURVEYS:

SAMPLING VERSUS NONSAMPLING ERRORS

By

Barry L. Ford, Jack Nealon and Robert D. Tortora

1986

Area Frame Estimators in Agricultural Surveys: Sampling Versus Nonsampling Errors

Barry L. Ford, Jack Nealon, and Robert D. Tortora*

Abstract

This article describes and compares three area frame estimators used for agricultural surveys and shows the importance of sampling and nonsampling errors to both economists and statisticians. Comparisons among the estimators are based on their applicability, sampling efficiency, and susceptibility to nonsampling errors. The weighted segment estimator has the widest applicability and highest precision, but it also has important nonsampling errors. This article also discusses research conducted to evaluate and reduce the nonsampling errors.

Keywords

Area frame, closed segment estimator, open segment estimator, weighted segment estimator, nonsampling errors

Introduction

Probability surveys that make agricultural estimates often use an area frame. The area frame includes all the land within a specified geographical area, such as the continental United States, and is used to define the sample for a survey. Area frames have been used by the Statistical Reporting Service (SRS) since the early sixties. The June Enumerative Survey (JES), an annual survey which measures planted acreage of crops and numbers of livestock, is an example of a major survey which depends almost entirely on an area frame. Although the use of list frames for probability surveys has increased greatly in recent years, area frames are still needed to measure the incompleteness of the lists.

The Farm Costs and Returns Survey, measuring farm economic values, is an example of a survey that uses an area frame to measure the incompleteness of the list frames. In this case, the lists include only farms with gross sales greater than $100,000; thus, the sample from the area frame to measure the incompleteness of the lists is an essential part of the survey. The sample from the area frame contributes about 50 percent of the agricultural economic estimates. Area frame estimators are a crucial part of almost all probability surveys on which agricultural estimates are based.

SRS area frames are stratified in each State according to several factors, intensity of agriculture on the land being a primary factor. SRS selects a sample from each stratum. Each element of the sample is a continuous parcel of land called a segment. SRS draws the boundaries of each segment on an aerial photograph; enumerators use these photographs when collecting data.

After data collection, SRS uses three estimators: the closed, open, and weighted segment estimators. All three require that the enumerator establish what farms are related to each segment. (For SRS purposes, a farm is defined to be all land under one operating arrangement with gross farm sales of at least $1,000 a year.) The enumerator finds out what portion of the segment is under the operation of each farm. This portion is called a tract, and the enumerator draws the boundaries of each tract on the aerial photograph, accounting for all land in the segment.

*The authors are mathematical statisticians with the Statistical Reporting Service, U.S. Department of Agriculture. Ford is the Head of the Crops, Prices, and Labor Surveys Group, Estimates Division; Nealon is the Head of the Sampling Frame Development Section, Statistical Research Division; and Tortora is Director of the Survey Division.

When an enumerator interviews a farmer, the closed segment approach requires that the enumerator obtain data for only that part of the farm within the tract. For example, the enumerator might ask about the total number of hogs on the land in the tract.

The open segment and weighted segment approaches require that the enumerator obtain data on the entire farm. For example, the enumerator would ask about the total number of hogs on all land in the farm. However, the open segment approach uses these data only when the headquarters of the farm is within the segment boundaries. (Thus, the headquarters is used to identify each farm uniquely with one segment.)

Using the weighted segment approach, the enumerator obtains farm data for each tract, but these farm data are weighted; the current weight used by SRS is the ratio of tract acres to farm acres.

The formulas for the three estimators and their standard errors appear in the appendix at the end of this article.

Suppose the following situation occurs for a specific farm: tract acres = 10, farm acres = 100, hogs on the tract = 20, and hogs on the farm = 40. The closed segment value of number of hogs would be 20; the weighted segment value would be $40 \times (10/100) = 4$; and the open segment value would be 40 (if the headquarters is in the segment) or 0 (if the headquarters is not in the segment).

Comparing the Estimators

Economists and statisticians who wish to obtain data or understand the nature of a particular estimate should be familiar with the distinct advantages and disadvantages of the closed, open, and weighted approaches. This section compares each estimator with respect to its applicability, the size of sampling errors, and its susceptibility to nonsampling errors.

Applicability

One of the most common uses of closed segment estimates is to estimate crop acreages and livestock inventories. An enumerator accounts for all land in each tract by type of crop or use and for all livestock in the tract. The main disadvantage of the closed segment estimator arises when the farmer can report only values for the farm rather than for a tract which is a subset of the farm. For example, "How many tractors do you own?" can only be answered on a farm basis. Thus, the closed segment estimator is not applicable for many agricultural items. Economic items and crop production are two major examples which farmers find difficult or impossible to report on a tract basis.

The open and weighted segment estimators, by contrast, do not have this limitation. They can be used to estimate all agricultural characteristics. This broad applicability is a major advantage for both estimators.

Sampling Efficiency

Sampling efficiency refers to the precision and cost of the estimators. The precision of an estimate can be measured by: (1) the variance of the estimate, (2) the standard error which is the square root of the variance, or (3) the coefficient of variation (CV) which is the standard error divided by the estimate. An estimate becomes less precise as any of these measures increases.

Given the same number of segments to make each estimate, weighted segment estimates are usually more precise than closed segment estimates, and closed segment estimates are usually more precise than open segment estimates *(5, 6, 7)*.[1]

Table 1 shows the variances of the three types of estimates from a data set collected in 1982. By comparing the estimated variances, the reader can see that the weighted segment estimates generally have smaller variances than the closed segment estimates, and the closed segment estimates have smaller variances than the open segment estimates. The reduced applicability of the closed segment estimator is clear by the fact that one cannot estimate the number of farms by using only tract information. Thus, there are no closed segment estimates for estimating the number of farms.

[1]Italicized numbers in parentheses refer to items in the References at the end of this article.

Table 1—Estimated variances of open, closed, and weighted segment estimates for five farm characteristics, 1982

State	Segments	Estimate	Cattle and calves	Hogs and pigs	Corn acres	Soybean acres	Number of farms
	Number	*Type*	10^{10}				10^{7}
Georgia	436	Open	3.55	1.59	3.66	17.92	1.13
		Closed	2.70	2.88	.49	4.76	NCL
		Weighted	1.22	1.51	.60	3.91	.88
Indiana	324	Open	2.42	18.67	50.12	28.44	1.19
		Closed	2.35	19.74	4.78	4.14	NCL
		Weighted	.89	7.85	2.92	2.15	1.61
Missouri	450	Open	24.11	41.21	10.20	73.36	3.13
		Closed	12.07	29.59	3.75	9.43	NCL
		Weighted	5.20	8.10	2.22	6.88	2.60
North Carolina	391	Open	1.20	1.54	8.09	11.67	2.73
		Closed	.75	6.25	1.82	2.30	NCL
		Weighted	.52	2.35	1.46	1.62	2.12
Ohio	324	Open	6.19	7.81	31.38	20.13	2.16
		Closed	3.02	6.17	3.73	3.36	NCL
		Weighted	1.94	3.68	2.76	1.75	1.37
Total	1,925	Open	37.46	70.81	103.45	151.51	11.14
		Closed	20.89	64.62	14.58	23.98	NCL
		Weighted	9.76	23.48	9.97	16.31	8.58

NCL = Not calculable.

Why would these relationships among the variances occur? Open segment estimates use farm values. Thus, the amount of variability in the value of an agricultural item across the entire set of segments can be great. For example, many segments may have no farm headquarters and get a zero value for a certain agricultural variable; another segment may have one or two headquarters of large farms and get an extremely large value for the agricultural variable. The open segment estimate spreads the agricultural data unevenly throughout the segments.

A closed segment estimate spreads the data more evenly through the segments, and, thus, decreases the variance. For the combined five-State estimates in table 1, this result is especially true for corn and soybean acres, items related directly to land. However, for numbers of livestock, the closed segment estimates are not much better because livestock tend to be in herds or groups that will either be in the tract or not, often forcing the tract value to be either zero or a large value. A weighted segment estimate apportions the farm livestock according to the percentage of the farm acres in the tract. This effect causes the livestock data to be more evenly spread throughout the segments than for either the open or closed segment approaches.

Efficiency also involves the costs of data collection. Because each farmer operating a tract in the segment must be interviewed, the data collection costs for the closed and weighted segment approaches are approximately equal. The cost of collecting only tract data may be slightly less if contacting the farmer is difficult. In this case, the enumerator can usually observe the tract values because the tract boundaries can be established as the enumerator works the rest of the segment. In contrast, the weighted segment estimates require farm values which are much more difficult to observe, and the enumerator would have to put more time into contacting the farmer. However, this cost difference would be slight compared with overall survey costs. Thus, the reader should consider the closed and weighted segment estimates as having the same costs.

An open segment estimate does not require interviewing all farmers with tracts in the segment, but only requires interviewing farmers whose headquarters are located in the segment. Depending on factors such as the dispersion of the sample and the size of farms, the costs of obtaining data for open segment estimates can be 5-20 percent less than for closed or weighted segment estimates. Thus, the open segment approach has a cost advantage, but less precision. To make the open segment estimates as precise as weighted segment estimates would require more segments to be enumerated. Depending on what agricultural item is estimated, the extra cost of these segments may or may not exceed the advantage of a lower cost per segment for the open segment estimates.

Nonsampling Errors

Whereas precision relates to the sample size and the variability of the data, nonsampling errors refer to the effects of biases in the data or estimators. For estimates from large-scale surveys, and agricultural surveys are no exception, nonsampling errors may be of more concern than precision because such errors are harder to control.

Closed Segment Estimates

An accurate measure of the total number of acres within each segment is available prior to interviewing because each segment is delineated on aerial photographs. This measurement provides a control on the total land within the segment accounted for by the enumerator, and it increases the accuracy of the closed segment estimates for crop acreages.

Two types of nonsampling errors occur for closed segment estimates. The first type occurs because there are small areas of waste within the fields. Because of the scale of a photograph, accounting for areas of waste less than 1 acre is particularly difficult. These small waste areas are usually not visible on the photograph and generally cannot be observed from the actual location of the interview. The second type of nonsampling error occurs in some Western States where cattle and sheep roam freely through open gates and cross tract or segment boundaries. In these situations, the operator may not know the exact location of the livestock at the time of the interview and, thus, may not be able to report exactly how many livestock are on the tract at that time. Neither of these types of nonsampling errors is considered serious because research studies have never shown that a consistent positive or negative bias has resulted.

A major advantage of closed segment estimates is their ability to decrease any bias caused by nonresponse. When the farmer refuses to supply information or is inaccessible, the enumerator can usually observe the crops and livestock in the tract. The acreages associated with the crops can be measured from photographs. Often the livestock can be counted; even when they cannot, an indication of their presence allows SRS to do a better job of adjusting for nonresponse (*4*).

Open Segment Estimates

Four important nonsampling errors are associated with open segment estimates. The first is caused by the incorrect application of the "headquarters" rule. Although application of this rule is straightforward for farms run by individuals, the identification of headquarters can become difficult for partnerships, corporations, and managed farms where enumerators need to ask a series of questions to eliminate potential duplicate reporting. SRS has never been able to measure fully the effects of this nonsampling error on the estimates, but it has developed better questionnaires and has stressed the problem during enumerator training to minimize the effects.

The second nonsampling error is the underestimating of the farm population. This error probably arises because the headquarters of farm operators may be inadvertently missed in more densely populated areas. For most farmers, the headquarters is the home, and if the home is in town or in a subdivision, an enumerator may have difficulty identifying that home as the residence of a farmer, especially if the farmer does not consider farming as the primary occupation. Other factors may contribute to this underestimation.

The third nonsampling error is the underreporting of farm values. The reader should consider the effect on a farmer when interviewed by an enumerator. To obtain tract values, the enumerator shows the farmer an aerial photograph on which the precise boundaries of the farmer's tract are drawn. Then the farmer is asked to report values associated with that specific piece of land. When farm values are obtained, there is no map. Both enumerator and farmer must switch into the more nebulous concept of "the farm," that is, the land operated by this farmer. Farmers tend to forget about parcels of rented land not contiguous with the main part of their farm and about parcels of woodland or wasteland under their control, but considered "nonagricultural."

A fourth nonsampling error can occur when the farmer reports livestock data. The enumerator asks the farmer about livestock on the land operated *regardless* of the ownership of the livestock. However, a farmer tends not to report livestock on the farm that are owned by someone else. This typically happens when the farmer is feeding livestock (under contract) owned by someone else. A farmer also tends to report livestock owned by the farmer, but located on someone else's land.

Weighted Segment Estimates

This section is more detailed and quantified than the previous two sections because of the large amount of SRS research to evaluate the nonsampling errors of the weighted segment estimates. SRS believed that the research was warranted because of the advantages in applicability and precision of weighted segment estimates over closed and open segment estimates. This research began when SRS reinterviewed a subset of respondents after the 1974 JES (3). Respondents were again asked many of the JES questions about their farms and were then asked to reconcile any differences between the original JES responses and their responses during the reinterview. The evaluation, which was small in scope, involved reinterviews with only 163 JES respondents in Nebraska.

Comparing the original weight with the reconciled weight, SRS found 44 differences out of the 163 reinterviews (27 percent) caused by incorrect responses. Exactly half the differences were positive and half were negative. The effect of the reconciled weights on the weighted segment estimates of hogs and cattle caused biases of -9.6 percent and -0.2 percent, respectively.

SRS also estimated the biases associated with the components of the weighted segment estimates. The estimated biases were as follows: acres in the tract, 0.3 percent; acres in the farm, -5.6 percent; hogs on the farm, -2.5 percent; and cattle on the farm, -2.4 percent. Thus, the concept of farm acres caused the largest bias.

The results of the 1974 JES reinterview were sufficiently troubling that SRS planned a detailed reinterview with a large sample size after the 1976 December Enumerative Survey (DES). The DES reinterview involved 528 respondents in three States: Indiana, North Carolina, and Oklahoma. Table 2 shows the biases in weighted segment estimates for hogs and cattle in the three States and the significance levels from statistical tests of whether the biases were significantly different from zero.

Table 2—Estimated bias in the weighted segment estimates for hogs and cattle and the alpha level from each statistical test, 1976

State	Interviews	Hogs and pigs		Cattle and calves	
		Bias	Alpha level	Bias	Alpha level
	Number	*Percent*		*Percent*	
Indiana	149	−11.7	0.15	−4.5	0.02
North Carolina	172	−16.9	.01	−.3	.96
Oklahoma	207	2.2	.42	−3.4	.36
Total	528	−10.9	NA	−3.0	NA

NA = Not available.

When investigating biases in the components of the weighted segment estimates, SRS found that problems in the denominator of the weight—farm acres—were the most serious. For the three States, 44 percent of the farmers who were reinterviewed reported a different number of farm acres. These differences were not offsetting and resulted in the following estimated biases for the farm acres that had been collected on the DES: Indiana, -2.9 percent; North Carolina, -9.9 percent; and Oklahoma,

–4.8 percent. Thus, a farmer tended to underreport the acres in the farm, a result which was consistent with the 1974 reinterview study in Nebraska.

This DES study also obtained the reasons for differences between the original responses and the reinterview responses. Of the differences, 19 percent occurred because the farmer estimated the acreage rather than taking the time to account for the exact acres in the farm. Fifteen percent of the differences were caused by problems in reporting parcels of woodland or idleland which had no crops or livestock. Thirteen percent of the differences involved land rented by or rented out by the farmer. Thirteen percent involved farmers who simply miscounted their acreage. Eight percent could only be attributed to the fact that a different respondent participated in the reinterview than in the DES. The remaining 32 percent reflected miscellaneous reasons such as inclusion of land that was to be sold in the near future, incorrect readings of JES photographs, and farmers who did not remember the initial JES interview.

The authors of the study decided that the underreporting of parcels of woodland or idleland was the main reason for the negative biases in reporting farm acres (*8*). Parcels of woodland and idleland mixed into the agricultural land were more typical of farming conditions in North Carolina (a bias of -9.9 percent in farm acres) than in Indiana (a bias of -2.9 percent in farm acres). The other major reasons for differences caused both negative and positive differences in the farmers' responses, while omission of parcels of woodland and idleland always caused a negative difference.

Panel discussions with enumerators in seven States before the 1982 JES confirmed the problem of obtaining accurate farm acres on the JES:

> Interviewers consider this section to be one of the hardest to get correct answers on; intensive probing is often required. Respondents often do not know the exact acreages offhand. Many operators report only cropland, omitting other types of land such as woodland and wasteland. (*12*)

In summary, closed segment estimates would be the best to use except that their restricted applicability can be an insurmountable problem for survey designers. Thus, weighted segment estimates appear to be the best alternative if the costs are not prohibitively high when compared with the costs of the open segment estimates. SRS now uses all three estimates for its JES because no one type of estimate is clearly the best for every agricultural item estimated by the survey. However, this situation adds more burden on the respondent and greater complexity to the questionnaire.

Research on Alternative Weighted Estimators

How can weighted segment estimates be improved? SRS has recognized that the prevention of nonsampling errors, especially trying to get farmers to give exact acreages rather than best guesses, is extremely difficult. Thus, SRS has decided to concentrate its research on the investigation of alternative weighting schemes for the weighted segment estimator.

Weights Defined by Agricultural Land

SRS evaluated an alternative weighting scheme in 1980 based on total land minus woodland, wasteland, and other nonagricultural land. This weight should have been less susceptible to bias than the operational weight because it subtracted from the numerator and denominator the type of land that was the major source of bias with the operational weight.

The operational and alternative versions of the weighted segment estimator were compared in three States during the 1980 JES for three important farm characteristics: number of farms, total cattle, and total hogs. Table 3 shows the relative difference between the two types of weighted segment estimates for each of the three farm characteristics. The differences between the two estimates were quite small in most instances. Thus, the study did not show that the alternative weight was less biased than the operational weight.

Some nonsampling errors associated with the alternative weight surfaced during the study. The most serious error centered on a discrepancy when farmers reported nonagricultural land at the tract level and the farm level. At the tract level, farmers were instructed to exclude not only woods and other blocks of nonagricultural land but also waste within

Table 3—Relative difference between operational and alternative weighted segment estimates, 1980

State	Segments	Relative difference[1] Number of farms	Cattle and calves	Hogs and pigs
	Number	*Percent*		
Minnesota	343	3.5	0	−0.5
Ohio	324	3.3	4.7	−2.5
Wisconsin	310	1.0	−1.2	−.1
Total	977	2.6	.3	−.8

[1]Relative difference = 100 (alternative-operational)/operational.

an agricultural field. The author of that study concluded that when reporting for the farm, farmers did not exclude small parcels of waste within agricultural fields (*2*). Therefore, some nonagricultural land was included in the denominator of the weight, producing an downward bias in the alternative weight.

SRS tested a modified alternative weight in five States during the 1981 JES. This modified weight did not include within-field waste. The estimates were compared between the operational and alternative weighted segment estimates for the same three farm characteristics evaluated in 1980. The estimates were remarkably similar again. The alternative weighting scheme was not less susceptible to an upward bias than the operational scheme. Numerous nonsampling errors associated with the alternative weight were identified during the studies (*9, 10*). Thus, SRS decided to investigate a less complex alternative weight.

Weights Defined by Cropland

SRS collected data on cropland weights during the 1982 JES in five States. Cropland was defined as land planted or to be planted to crops during 1982, idle cropland, summer fallow, and cropland used only for grazing or pasture. Field waste was excluded from the cropland acreage for both the tract and farm to avoid the problems encountered during the 1981 study.

Eleven percent of the farms in the five States had no cropland. Cropland acreage for the farm was missing for another 12 percent of the operations. SRS adjusted the estimation procedures to account for the data on these operations. Thus, the cropland weight quickly ran into problems that annoyingly complicated the estimation.

The relative differences between the two types of weighted segment estimates are shown in table 4. The significance level from the paired t-tests comparing the estimates are also shown because there were many significant differences. The estimates for the number of farms and cattle inventories were significantly higher for the operational estimates in Georgia, Missouri, and North Carolina and for the total combined. These results indicate that the cropland weight is less susceptible to an upward bias than the operational weight. The three States with considerably more noncropland—Georgia, North Carolina, and Missouri—were the States where the two types of weighted segment estimates were significantly different.

In reporting on their 1982 research, Dillard and Nealon concluded that the cropland weight did not appear to be as biased as the operational weight (*5*). Two disadvantages of the cropland weights were that: (1) 23 percent of the operations had cropland that was either zero or missing, and (2) the CV's were slightly higher for the estimates using the cropland weight.

Further Research

SRS will continue attempting to reduce the nonsampling errors of the weighted segment estimator that is used operationally. The applicability of this estimator plus its high precision call for this continued research. The agency will also test new weighting schemes. The most recent research proposal is to weight by the size of the major agricultural item of each farm (*1*). If a farm has cropland, the major item will be the crop with the most acreage. If a farm has no cropland, the major item will be the type of livestock that are most numerous. Preliminary research has indicated an optimistic outlook for this estimator.

Statisticians still tend to compute sampling errors and make their decisions based only on them. In the case of SRS, the weighted segment estimator was implemented operationally almost as soon as

Table 4—Relative difference between the operational and alternative weighted segment estimates and the alpha level from each statistical test, 1982

State	Segments	Number of farms		Cattle and calves		Hogs and pigs	
	Number	*Relative difference*	*Alpha level*	*Relative difference*	*Alpha level*	*Relative difference*	*Alpha level*
Georgia	436	-7.4	0.01*	-19.4	0.01*	2.2	0.83
Indiana	324	.3	.81	-2.0	.38	.6	.76
Missouri	450	-5.3	.01*	-10.8	.01*	2.2	.67
North Carolina	391	-13.0	.01*	-29.6	.01*	1.9	.80
Ohio	324	-2.4	.07	-5.3	.09	.8	.79
Total	1,925	-5.5	.01*	-11.3	.01*	1.1	.63

*Denotes a significant difference at the 0.05 alpha level. Relative difference = 100 (cropland-operational)/operational.

SRS observed the decreased variance. However, 10 years after research began, SRS is still trying to identify and eliminate the nonsampling errors involved (*11*). The overall effect of SRS's research is to illustrate the importance and difficulty of assessing nonsampling errors when one searches for the "best" estimator. First, identifying all the nonsampling errors that come, often subtly, into play is difficult. Second, once identified, nonsampling errors are extremely difficult to measure. The very studies designed to assess the effects of the nonsampling errors will have their own nonsampling errors. Third, once measured, nonsampling errors may be difficult (or impossible) to prevent or correct. For survey designers, working with nonsampling errors can be a tortuous obstacle course.

References

(1) Bethel, James. *A New Approach to Weighted Area Frame Estimates.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1985.

(2) Bosecker, R.R. *Alternative Weighted Estimators.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1980.

(3) ___________, and W. Kelly. *Summary of Results from Nebraska Survey Concepts Study.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1975.

(4) Crank, K. *The Use of Current Partial Information to Adjust for Nonrespondents.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1979.

(5) Dillard, D., and J. Nealon. *Comparison of the Operational and Cropland Weighted Estimators.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1983.

(6) Fecso, R., J. Geuder, B. Hale, and S. Pavlasek. *Estimating Dry Bean Acreage in Michigan.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1982.

(7) Hendricks, W.A., D.T. Searls, and D.G. Horvitz. "A Comparison of Three Rules for Associating Farms and Farmland with Sample Area Segments in Agricultural Surveys," *Estimation of Areas in Agricultural Statistics.* Rome: Food and Agriculture Organization of the United Nations, 1965.

(8) Hill, G., and M. Farrar. *Impact of Nonsampling Errors on Weighted Tract Survey Indications.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1977.

(9) Nealon, J. *An Evaluation of Alternative Weights for a Weighted Estimator.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1981.

(10) ___________. *Response Errors in the Weighted Estimator.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1982.

(11) ___________. *Review of the Multiple and Area Frame Estimators.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1984.

(12) Weidenhamer, M. *Views on the June Enumerative Survey: A Qualitative Analysis of Discussions with Enumerators and Supervisory Enumerators.* SRS Staff Report. U.S. Dept. of Agr., Stat. Rep. Serv., 1983.

Appendix: Formulas for the Area Frame Estimators

Each area frame estimator can be described by the following notation. For some characteristic, Y, of the farm population, the sample estimate of the total for the closed segment estimator is:

$$Y_c^* = \sum_{i=1}^{s} \sum_{j=1}^{p_i} e_{ij} \sum_{k=1}^{n_{ij}} y_{ijk} \qquad (1)$$

where:

s = the number of land-use strata;

p_i = the number of substrata in the i^{th} stratum;

n_{ij} = the number of segments sampled in the j^{th} substratum in the i^{th} stratum;

e_{ij} = the expansion factor or inverse of the probability of selection for each segment in the j^{th} substratum in the i^{th} stratum;

$$y_{ijk} = \begin{cases} \sum_{m=1}^{f_{ijk}} t_{ijkm} & \text{if } f_{ijk} > 0 \\ 0 & \text{if } f_{ijk} = 0 \end{cases} \qquad (2)$$

f_{ijk} = the number of tracts in the k^{th} segment, j^{th} substratum, and i^{th} stratum, and

t_{ijkm} = the tract value of the characteristic, Y, for the m^{th} tract in the k^{th} segment, j^{th} substratum, and i^{th} stratum.

For the open segment estimator, the sample estimate would be the same form as Y_c^*:

$$Y_o^* = \sum_{i=1}^{s} \sum_{j=1}^{p_i} e_{ij} \sum_{k=1}^{n_{ij}} y_{ijk} \qquad (3)$$

except here:

$$y_{ijk} = \begin{cases} \sum_{m=1}^{f_{ijk}} b_{ijkm}\, y_{ijkm} & \text{if } f_{ijk} > 0 \\ 0 & \text{if } f_{ijk} = 0 \end{cases} \qquad (4)$$

b_{ijkm} = 1 if the farm headquarters is within the segment
0 if the farm headquarters is not within the segment

y_{ijkm} = the value of the entire farm for the m^{th} tract in the k^{th} segment, j^{th} substratum, and i^{th} stratum

The weighted segment estimator would also be of the same form:

$$Y_w^* = \sum_{i=1}^{s} \sum_{j=1}^{p_i} e_{ij} \sum_{k=1}^{n_{ij}} y_{ijk} \qquad (5)$$

but with the distinction that:

$$y_{ijk} = \begin{cases} \sum_{m=1}^{f_{ijk}} a_{ijkm}\, y_{ijkm} & \text{if } f_{ijk} > 0 \\ 0 & \text{if } f_{ijk} = 0 \end{cases} \qquad (6)$$

a_{ijkm} = the weight for the m^{th} tract in the k^{th} segment, j^{th} substratum, and i^{th} stratum.

Operationally, SRS has used the weight:

$$a_{ijkm} = \frac{\text{tract acres for the } m^{th} \text{ tract}}{\text{entire farm acres for the } m^{th} \text{ tract}} \qquad (7)$$

since the early seventies.

For all three estimators, the formula for the estimated variance can be written as:

$$\text{Var}(Y^*) = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \frac{(1 - \frac{1}{e_{ij}})}{(1 - \frac{1}{n_{ij}})} \sum_{k=1}^{n_{ij}} \left(y'_{ijk} - y'_{ij\cdot} \right)^2 \quad (8)$$

where:

$$y'_{ijk} = e_{ij} y_{ijk} \quad (9)$$

$$y'_{ij\cdot} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y'_{ijk} \quad (10)$$

The standard error is then:

$$SE(Y^*) = \{\text{Var}(Y^*)\}^{1/2} \quad (11)$$

In Earlier Issues

This fact that a sample can be more accurate than a census, under certain conditions, is becoming widely accepted. The explanation is simple. With the exception of rather unusual cases, surveys and censuses are subject to many errors which have little, if anything, to do with the way a sample is selected. The challenging problem is often how to get accurate and useful information from respondents, or how to keep errors due to causes other than sampling at a minimum—not how to design an efficient and adequate sample.

Earl E. Houseman
Vol. 2, No. 3, July 1950