USE OF NETWORK SAMPLING IN AN AGRICULTURAL SURVEY

Michael A. Steiner, U.S. Department of Agriculture

Introduction

The Statistical Reporting Service (SRS), United States Department of Agriculture (USDA) conducts many agricultural surveys in order to provide data on a wide range of items relating to the United States agricultural economy. It is the policy of the USDA to support the small farm as a continuing component of American agriculture. One problem in implementing the USDA's small farm policy has been the acute lack of information on the social and economic characteristics of the population. In order to provide needed information, SRS conducted a survey in 1981 called the Family Farm Survey.

The purpose of this survey was to obtain information about the households of small farm operators. Of particular concern were the characteristics of black farm operators. Since a great number of small farm operations are in the South, a study area consisting of a rural region of Mississippi and Tennessee was selected. Although there are a large number of small farm operators in the study area, black farm operators are a relatively rare population in the study area. According to 1974 Census of Agriculture data, approximately 14 percent of the farms in the study area were operated by blacks. Since 1974 the percentage of black farm operators is believed to have decreased significantly.

An adequate universe list of small farm operators was not available, so an area frame sample was used for the Family Farm Survey. Since black farm operators were a rare population in the study area, network sampling was used in conjunction with the area frame sample, which is not efficient for such populations. Network sampling is a procedure designed for sampling rare populations. Although network sampling has been used for many health surveys [1, 2] this is its first use in agricultural surveys.

This report discusses survey procedures involved with using network sampling in addition to an area frame sample. Four different network counting rules are described. Precision of estimates along with data collection costs are compared for the area frame sample units and the units identified by a network rule based on family relationships.

SAMPLE DESIGNS

Area Frame

An area frame was used so that all small farm operations in the study area would have a chance for selection. Use of a list frame was ruled out because of incompleteness of small farm operations on available list frames. The existing area frame constructed and maintained by SRS for economic surveys was used. In this frame all land was classified into strata consisting of areas of homogeneous land use. Size of sampling unit or segment varied depending on the particular stratum, ranging from one tenth of a square mile in the residential stratum to two square miles in area for the intensive agricultural land use strata. A two-stage stratified cluster sample

was used. The first stage of sampling was the segment and the second stage was the farm household.

Network

The Family Farm Survey is designed to obtain information by race of farm operator. Since black farm operators are a rare population, network sampling was used along with the area frame sample in order to obtain a larger sample of black farmers. In network sampling, elements possessing the rare attribute within a population may be linked to one or more enumeration units. The counting rule defines the network of households to be enumerated for the survey. The counting rule specified that a black farm operator was eligible to be identified for enumeration at his household and at the households of prescribed relatives living within the same county. Relatives included were grandparents, parents, brothers, sisters or children of the head of household or their spouse. All households within the boundaries of an area frame segment were interviewed. In addition, black farm households outside of segments which were identified by the counting rule were also interviewed.

DATA COLLECTION

General

Prior to field work, enumerators were trained on questionnaire content, survey procedures and purpose of the survey. Special emphasis was put on network sampling procedures and the importance of obtaining information from black households. Segments in the area frame sample were identified on U.S. Geological Survey Quadrangle maps and on county highway maps. All households within the sampled segments were screened. An in-depth interview was conducted at households containing a farm operator. Black households, both farm and non-farm, were asked to identify certain types of relatives within the county who were farm operators. Follow-up interviews were conducted with the relatives identified.

Questionnaire Design

Four forms were used to obtain data necessary for the network sampling procedures. A screening form (Appendix 1) was used to identify households within sampled segments and to determine if someone in the household qualified as a farm operator. Black non-farm head of households and their spouses were asked if they had any grandparents, parents, siblings or children who operated a farm within the same county. If the black non-farm head of household or spouse had relatives, their names were recorded on a listing sheet (Appendix 2). If individuals in segments qualified as farm operators, they and their family members were interviewed with a household questionnaire. The household questionnaire (Appendix 3) contained questions about the number of relatives within

496

the county and the names of any relatives who were farm operators. Names from the household questionnaire and the listing sheet were transferred to a record sheet (Appendix 4) which was a composite for the segment of relatives who are black farm operators. The record sheets were turned over to supervisory enumerators who coordinated the assignment of names for follow-up interviews.

Pretest

The survey questionnaires and sample design were pretested in thirty segments in the study area. One of the purposes of the pretest was to study alternative network counting rules. Four different counting rules were used in each of the thirty segments in the pretest. The counting rules are explained in the following section of this report. A primary concern was respondents willingness and ability to identify black farm operations. Pretest results indicated that respondents were willing to identify names of black farm operators. Segments used in the pretest were not part of the sample used for the actual survey.

COUNTING RULE

Four different counting rules which identified black farm households in addition to those in the sampled area segments were considered for the Family Farm Survey. All four were used during the survey pretest. One rule specified relatives within the same area as the sample unit was located. The other three rules were area rules not restricting names to relatives. The area rules would be appropriate if blacks tended to know each other and if black farmers tended to live in clusters. We hoped to find a workable rule so that when a segment within the area frame sample happened to be located in an area containing a cluster of black farm operators, the one or more black farm operators living within the segment could identify other black farm operators in the immediate area.

The four counting rules tested are as follows:
1) Residences of black farm operators located adjacent to a sampled segment and on the opposite side of the segment boundary as a black household inside the segment.
2) Residences of black farm operators adjacent to any of the sampled segment boundaries.
3) Residences of black farm operators located in any segment adjacent to a segment selected in the area frame sample.
4) Residences of black farm operators where either the farm operator or spouse is a grandparent, parent, sibling or child of a head of household or their spouse living within a segment in the sample.

Diagram 1 illustrates the counting rules. Households Y1 and Y4 would qualify under counting rule 1 since they are adjacent to a segment and are on the opposite side of the segment boundary as a black household inside of the segment. Households Y2 and Y5 along with households Y1 and Y4 would qualify under counting rule 2 since they are all located adjacent to the segment. Households Y2 and Y5 do not qualify for counting rule 1 since they are not opposite the segment boundary to X1 and X2, respectively. Households Y1, Y2, Y3, Y4, Y5 and Y6 would qualify under counting rule 3 since they are all within segments adjacent

to segments selected in the area frame sample. Rule 3 required that boundaries of segments adjacent to sampled segments be drawn prior to enumeration. If Y2 was the son of the head of household in X1, and Y7 and Y8 were brothers to the head of household in X2, then Y2, Y7 and Y8 would qualify under counting rule 4.

The criteria for selecting a counting rule was that it must generate additional names of black farm operators without causing a significant bias on the estimates. The pretest indicated that rule 1 would not generate a sufficient number of black farm operators. Since rule 1 did not satisfy the selection criteria it was not used for the actual survey. As expected rule 2 generated more names than did rule 1. However, rule 2 required that respondents understand where segment boundaries were and also be able to identify households adjacent to those segment boundaries. This proved to be difficult for some respondents, which resulted in eligible names not being reported. Respondents also had difficulty in reporting all eligible names using rule 3, resulting in the same type of response bias as rule 2, but to a greater extent. For both rules 2 and 3, respondents tended to report names of successful farm operators who were prominent in the area, and those who were active in local agricultural programs. Poor, small-farm operators who often did not participate in agricultural programs were frequently not reported. Since rules 2 and 3 caused considerable bias, they were not used for the actual survey. Rule 4 generated additional names of black farm operators without producing the response bias that rules 2 and 3 did. With rule 4 respondents were not required to ascertain the location of segment boundaries and to determine the proximity of farm households to those boundaries. Rather, a respondent could identify certain types of relatives living anywhere in the county. Having been explained the importance of the study for assisting black farm operators, respondents seemed willing and able to identify names of relatives. Rule 4 was the counting rule selected for the actual survey.

NETWORK COUNTING RULE WEIGHT

Every farm household enumerated in the survey was assigned a network counting rule weight which was used to adjust the probability of selecting the area frame segment. The counting rule weights are determined on the basis of information collected from the household during the enumeration. The counting rule weight assigned to each black farm household is a ratio of the number of times the household was identified for enumeration divided by the number of households eligible to report the farm household. Consider the following two examples.

Example 1

A black farm household within an area segment reported on the household questionnaire that the head of household had parents living within the county and also a brother in a separate house within the county, and that the spouse of head of household also had parents living within the county. Since there were three households containing specified relatives in the county in

addition to the household itself, there was a total of four households which could have identified the black farm households had all four been selected in the area frame sample. The brother was located in another sampled area segment in the county and identified his brother as a farm operator, while neither set of parents happened to be located within an area segment selected for the sample. The black farm household in this example was identified for enumeration twice. because there was one household containing a relative in the area sample which identified the household, and since the household was within a sampled segment, the household identified itself for enumeration. This black farm household would be assigned a network counting rule weight of one-half because it was identified twice for enumeration while having four ways in which it could have been identified.

Example 2

A black non-farm household within an area segment reported on the screening form that the head of household had a son who operated a farm within the county. The name and address of the son was recorded on a listing sheet. A follow-up interview using the household questionnaire was conducted at the son's household. The son reported on the household questionnaire that his parents were his only relatives within the county. His wife also had parents living within the county. Since there were two households containing specified relatives in the county in addition to the household itself, there was a total of three households which could have identified the son's household had all three been selected in the area frame sample. Neither the son's household nor his wife's parents happened to be located within an area segment selected for the sample, leaving the son's parents as the only household which actually identified the farm household for enumeration. The son's farm household would be assigned a network counting rule weight of one-third because it was identified once for enumeration while having three ways in which it could have been identified.

DIAGRAM 1

County Boundary
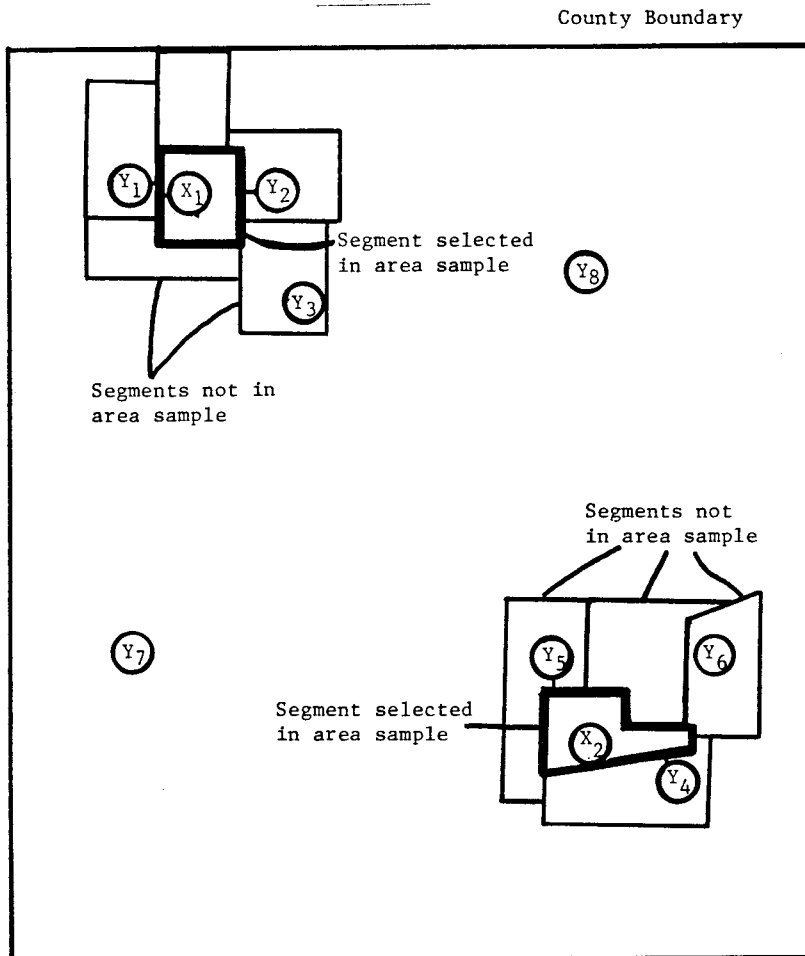
$\textstyle\bigcirc\kern-0.6em X_i$ = Black household within a segment in area sample

$\textstyle\bigcirc\kern-0.6em Y_i$ = Black household outside of segments in area sample

498

RESULTS

In order to compare the results of conventional area frame sampling to network sampling used in conjunction with an area frame sample, two sets of estimates were computed. One set of estimates used strictly the area frame data as if network sampling was not used. The second set of estimates included data from the area sample along with data obtained using network sampling. Table 1 shows the coefficients of variation for selected variables of black farm operators for both sets of data. The addition of data obtained by use of network sampling procedures reduced CV's for all variables in Table 1, ranging from a 4.2 percent reduction for number of beef cattle to a 25.4 percent reduction for number of family members.

TABLE 1

Coefficients of Variation for Black Farm Operators
In Study Area

Variable	Coefficient of Variation		Reduction In CV Due to Network Sampling (%)
	Area Sample	Area and Network Sample	
Number of Farms	12.2	9.2	24.6
Number of Family Members	13.0	9.7	25.4
Total Acres In Farms	14.6	13.2	9.6
Number of Beef Cattle on Farms	18.9	18.1	4.2
Total Production Expenses in 1980 On Farms	23.0	21.2	7.8

DATA COLLECTION COSTS

The decreases in sampling variances using network counting rules were obtained at a much lower cost than using conventional area frame sampling. Use of network counting rules allows for enumeration of particular black farm households whose name and address were obtained from black households living in sampled area segments. This pinpointing of particular black farm households for enumeration costs much less than canvassing several sampled areas in order to locate black farm households for enumeration. Table 2 shows the savings in data collection costs due to network sampling. Completed interviews were obtained from seventy-five additional black farm households by using network sampling procedures. In order to have attained interviews with seventy-five black farm households by area sampling it would have been necessary to select an additional 456 area segments. These additional segments would have been screened for black households and

interviews would have been conducted at households containing a farm operator. The extra cost of sampling additional segments to get the same number of interviews obtained by network procedures would have added thirty-four percent to the data collection costs for the survey.

TABLE 2

Savings In Data Collection Costs Due to Network Sampling

Item	Number
Number of Additional Completed Interviews with Black Farm Households Obtained Using Network Procedures	75
Additional Segments Required to Achieve Number of Interviews Obtained by Network Procedures	456
Increase in Data Collection Cost of Using Area Sampling to Reach Number of Interviews Obtained by Network Procedures	+ 34 %

CONCLUSION

The results showed that use of network sampling in conjunction with an area frame sample improved the precision of estimates for a rare population. The technique of using network counting rules in addition to an area sample allows for use of an existing area frame, and does not require the expensive and time-consuming task of building a specialized list which would have limited usefulness. Although use of network counting rules does require collection of additional information, enumerators were able to collect the names, addresses and counts of relatives needed for this survey with a minimum of problems. The improved precision of estimates with network sampling was obtained at a cost of tens of thousands of dollars less than if conventional sampling techniques had been used.

REFERENCES

[1]   Sirken, Monroe G., "Household Surveys With Multiplicity," Journal of the American Statistical Association, Vol. 65, No. 329, March 1970, pp. 257-266.

[2]   _____, "Survey Strategies for Estimating Rare Health Attributes," Proceedings of the Sixth Berkeley Symposium on Statistics and Probability, 1970, pp. 135-144.