United States Department of Agriculture

National Agricultural Statistics Service

Research and Applications Division

SRB Research Report Number SRB-92-07

July 1992

EVALUATION OF AN ALTERNATIVE SAMPLING DESIGN FOR THE FCRS AREA FRAME SAMPLE

Orrin Musser

EVALUATION OF AN ALTERNATIVE SAMPLING DESIGN FOR THE FCRS AREA FRAME SAMPLE, by Orrin R. Musser, Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC 20250, Research Report No. SRB 92-07, June 1992.

ABSTRACT

This research project was conducted, using data from the 1989 and 1990 June Agricultural Surveys(JAS) and the 1990 Farm Costs and Returns Survey(FCRS), to provide information useful in understanding current problems of the high incidence of interviews with nonfarms and possible undercoverage in the FCRS area frame sample. The current FCRS area sample includes all fourth and fifth year JAS resident farm operator(RFO) agricultural tracts, even those which failed to qualify as farms in June. There is no screening for FCRS, and the determination of farm-nonfarm status is not made until the completion of the interview. Thus full interviews are often conducted for nonfarms, and the data are zeroed out in the summary.

The results show that there are a disproportionate number of such wasted interviews with the JAS nonfarm ag tracts, and that a sampling design which subsamples nonfarm tracts at a low rate would reduce costs with little impact on the overall precision of survey indications. Also, currently no JAS non-agricultural tracts are enumerated, even those with potential for agriculture in the FCRS reference year, and there are concerns that those non ag tracts with potential that do become farms are missed. The data from consecutive June surveys suggest that about 4% of the non-ag tracts with potential do qualify as farms the following year, and thus that there may indeed be a coverage problem.

ACKNOWLEDGEMENTS

The author would like to thank Dave Dillard of the Objective Yield Surveys Section for his advice and assistance in obtaining FCRS data and summaries. Thanks to Jack Rutz and Chris Cadwallader of the Survey Administration Section for their help in providing answers and analysis on FCRS. Thanks also to Phil Kott and to Charles Perry for technical advice. And thanks to Dale Atkinson for his assistance on all stages of this project.

INTRODUCTION

This research brings together data from the 1989 and 1990 June Agricultural Surveys(JAS) and the 1990 Farm Costs and Returns Survey(FCRS) to provide information useful in understanding current problems of the high incidence of interviews with nonfarms and possible undercoverage in the FCRS area sample. This information may be useful in evaluating the potential and/or feasibility of an alternate sampling approach for FCRS area frame tracts.

The Farm Costs and Returns Survey is an annual face-to-face survey conducted in February and March, which provides regional and national estimates of income, expenses, assets, and liabilities associated with farm operation during the previous year. The target population is the official United States Department of Agriculture farm population, which is defined as "all establishments that sold or would normally have sold at least $1000 of agricultural products during the year". Farms that report sales of less than $1000 are assigned values(points) for crop acreages and livestock inventories. If these values indicate that the farm would normally have sold more than $1000 of agricultural products during the year, then it qualifies as a farm.

The FCRS is a multiple frame survey using a list frame of medium, large, and specialty farms and a supplemental area frame, to provide complete coverage. The area sample is a subset of the JAS sample segments. The reporting units are all resident farm operators within the segment who are not on the list(referred to as "nonoverlap").

The current FCRS area frame sample consists of all JAS segments in the fourth and fifth years of their five year rotation cycle. Segments with no resident farm operators in June are treated as valid zeroes and are not enumerated. Tracts determined to be non agricultural in June are not enumerated. All agricultural tracts, even those not satisfying the USDA definition of a farm in June, are enumerated.

After the FCRS interview is conducted, tracts which do not qualify as farms have their data set to zero. The farm determination is not made until the end of the interview. While this ordering of the questionnaire is deemed necessary to avoid inaccurately declaring a respondent a nonfarm at the beginning of an interview, it does result in wasted interviews. Alternate sampling techniques could reduce the number of wasted visits and interviews. One such alternative would involve stratifying the JAS tracts based on value of sales and subsampling lower sales strata at lower sampling rates. Non-ag tracts which currently are not enumerated could also be sampled at a low rate, to improve coverage. Currently, non-ag tracts with potential for agriculture(in June AS) which become agricultural tracts for the FCRS reference year, are missed.

BACKGROUND

A serious concern for the FCRS is the large number of interviews with operations too small to qualify as farms. Of the 2550 completed area frame interviews in the 1990 FCRS, 460 reported sales and estimated potential sales which failed to qualify the tract as a farm. The data for these tracts are set to zero in the summary. There are also a large number of tracts with a completion code of 4, "nonfarm screenouts", which should be tracts determined to be nonfarm prior to interview, but may include some nonfarm interviews.

Until recently, FCRS used a separate economic area frame and all selected segments were enumerated without tract information from a previous survey. There was a screening section on the questionnaire and tracts which were non-agricultural or did not meet the farm definition were "screened out" and not interviewed. These were referred to as "non-farm screenouts". The current FCRS area frame sample is a subset of the JAS sample containing all fourth and fifth year agricultural tracts. All of these tracts had reported crop acreages, livestock inventories, crop storage, agricultural sales, or government agricultural payments on the screening form in June and completed the questionnaire. For this reason, screening is not considered necessary for the FCRS and an interview is conducted for each of these tracts. However, there is still a completion code 4 for non-farm screenouts. If a prior determination is made that a tract is non-agricultural or that an error was made in June and that the land operated by a resident operator is definitely not a farm, then the questionnaire is given a completion code of 4 and no interview is conducted. Valid use of this code should be rather unusual, and in questionable cases, the interview should be conducted. However, there is a large number of completion code 4 records in the '89 and '90 data. Many of these records may represent completed interviews with nonfarms, which should have been given a completion code of 1, but were coded 4 to avoid having to key enter data that would be zeroed out in the summary. In 1990 there were 624 nonfarm screenouts(completion code 4), resulting in a total of 1084 tracts from the June agricultural tracts which were determined to be nonfarm for the FCRS. If all of these tracts had been interviewed, they would have accounted for nearly one third of the area frame interviews.

Another concern for the FCRS area frame sample is complete coverage. All tracts determined to be non-agricultural in June are excluded from the FCRS sample, even those determined to have agricultural potential for the FCRS reference year. While most of these tracts would not qualify as farms if sampled, a few would, and their exclusion from sampling eligibility raises coverage concerns.

METHODS AND RESULTS

Since the 1990 FCRS segments and tracts are a subset of those in the previous JAS(1990), we can look at data from both surveys for each tract in the FCRS. It is interesting to examine the FCRS data for certain subpopulations, or domains, based on the June "value of sales"(either reported sales from the previous year, or if reported sales < $1000, estimated potential sales value for the current year from crop acreages and livestock inventory). Sales are grouped into 12 categories on the June questionnaire, and these will be our domains of interest:

DOMAIN	JUNE(1989) VALUE OF SALES
1	< $1000
2	$1000-2499
3	$2500-4999
4	$5000-9999
5	$10K-19999
6	$20K-24999
7	$25K-35999
8	$40K-49999
9	$50K-99999
10	$100K-250K
11	$250K-500K
12	>= 500,000

We could think of these domains as strata for tracts, from which we could have subsampled, in the 1990 FCRS. This research focuses interest on the domain 1 tracts, which were the agricultural tracts which did not qualify as farms in the JAS.

The table below displays the numbers and percentages of nonfarm interviews and screenouts by value of sales domain(collapsed into groups of roughly equal numbers of total interviews). As we might expect, a disproportionate number of the nonfarm interviews and nonfarm screenouts were domain 1 tracts. While less than one fifth of all interviews were domain 1 tracts, more than two thirds of the nonfarm interviews and nearly one half of the nonfarm screenouts are from domain 1.

DOMAIN	TOTAL NONFARM INTERVIEWS		NONFARM SCREENOUTS	
1 (< 1000)	319	(69.3%)	302	(48.6%)
2 (1000 - 2500)	111	(24.1%)	173	(27.8%)
3+4 (2500 - 10,000)	25	(5.4%)	88	(14.1%)
5-12 (>10,000)	5	(0.1%)	59	(9.5%)
TOTAL:	460	(100.0%)	622	(100.0%)

It is clear from the data that sampling the domain 1 tracts at a low rate would decrease the proportion of nonfarm interviews and screenouts, but we must consider the resulting increase in the variance of the estimates. Like the other follow-on surveys, the FCRS area frame sample design would be two-phase, with the June selection of segments being the first phase, and the sampling of domain 1 tracts the second phase(for other domains all tracts would be sampled). The variance of an estimator $\hat{\theta}$, in this two phase design, would have two components(see Cochran, p.276):

$$V(\hat{\theta}) = V_1(E_2(\hat{\theta})) + E_1(V_2(\hat{\theta}))$$

The first term in the equation represents the portion of the variance of the estimator due to first phase sampling of segments, and thus would be the variance that would be obtained if all second phase sampling units(tracts) were sampled. This corresponds to the current FCRS design. The second term is the portion of the variance of the estimator due to second phase sampling, and thus would correspond to the increase in variance if FCRS were to adopt a two phase design. Suppose y is a variable whose population total we wish to estimate. We currently estimate this total with the sum of the expanded y values for all tracts in sampled segments. If we think of the second phase sample as a simple random sample of m of these n expanded values, then the variance due to this sampling, and thus the increase with the two phase design, is just the formula for the variance of the Horvitz-Thompson estimator of a total under simple random sampling:

$$V_2(\hat{\theta}) = n^2\left(1-\frac{m}{n}\right)\frac{S^2{}_{yexp}}{m}$$

We can thus estimate the increase in variance for two-phase sampling with the '90 FCRS data. Since the '90 data includes all tracts in sampled segments, we can actually compute the exact second phase variance of the estimator of the total for this particular sample and use this as an estimate for the increase in variance for any first phase sample. This increase would be the expected increase in variance of the area frame nonoverlap estimator. Since we would only sample from the domain 1 tracts, we just compute the formula above over these tracts at the state level and sum to the national level to obtain estimated increases in variances and cvs, expressed in percentages, for various sampling rates:

SAMPLING RATE	% **INCREASE IN VARIANCE** OF NOL ESTIMATES FOR: Number of Farms		Total Expenses		Total Income	
	VAR	CV	VAR	CV	VAR	CV
1/2	3.69%	(1.8%)	2.26%	(1.1%)	.29%	(.15%)
1/3	7.30%	(3.5%)	4.47%	(2.2%)	.59%	(.29%)
1/4	11.07%	(5.4%)	6.77%	(3.3%)	.89%	(.45%)
1/5	14.76%	(7.1%)	9.03%	(4.4%)	1.19%	(.60%)
1/10	33.22%	(15.4%)	20.30%	(9.7%)	2.69%	(1.34%)

Thus, if we were to sample the domain 1 tracts at a rate of one in four, we might expect about a 5% increase in the cv for the number of farms, 3% for total expenses, and less than 1% for total income. These expected increases to the area nonoverlap cv's would have a very small effect on the overall multiple frame cv's, as shown in the following table which gives the actual 1990 cv's for these variables and the "expected" cv's(the cv's which would result from the expected increase in variance due to sampling domain 1 tracts):

1990 CV'S AND EXPECTED CV'S FOR SAMPLING RATE OF .25

Variable	1990 Area Nonoverlap CV		Expected	1990 Multiple Frame CV		Expected
Number of Farms	3.29	-->	3.45	1.77	-->	1.80
Total Expenses	5.48	-->	5.64	2.17	-->	2.19
Total Income	6.39	-->	6.44	2.37	-->	2.37

It is also interesting to look at the contribution of the domain 1 tracts to the area nonoverlap and multiple frame indications for these variables. The domain 1 estimate for number of farms is fairly large, 66,265(7.2%) of the area nonoverlap estimate of 926,086, and 3.8% of the multiple frame estimate of 1,752,125. The relative contributions of domain 1 tracts to the economic variables are much smaller. The domain 1 estimate for total expenses was about 3.2% of the nonoverlap total, but only .7% of the multiple frame total, and for total income 1.1% of the nonoverlap and only .2% of the multiple frame. Clearly the domain 1 tracts have little impact on important economic indications. It should also be pointed out that the alternate sampling plan would be using a design unbiased estimator for the domain 1 strata total and thus would have the same expected value as the current estimator. Thus we would not expect the sampling of domain 1 tracts to introduce any systematic bias into the estimates.

At the .25 sampling rate, there would be substantial cost savings. Our findings suggest that about one fourth of all area nonoverlap tracts are domain 1. About one fifth of the total FCRS "contacts" are area nonoverlap, and thus we might expect about 5% of the total contacts to be domain 1 area tracts. By sampling these at a rate of one in four we should eliminate about 3.75% of the total contacts. The figure usually cited for the cost of FCRS is the total NASDA enumeration cost, which was $2,234,147 for 1990. A conservative reduction of only 2% in NASDA costs would have resulted in savings of about $45,000.

In 1990, there were over 900 contacts with domain 1 tracts, including refusals and inaccessibles. With a .25 sampling rate, and a cost per contact of $75(NASDA average cost per contact=$93) the savings from sampling domain 1 farms would be over $50,000.

Since our data includes reported (or estimated potential) tract farm sales for both June(AS) and FCRS, it is interesting to look at how many tracts show increases or decreases in reported value of sales from June to February. The table contains data for all completed interviews. Of all domain 1 interviews, about 68% resulted in nonfarm determination, while for domain 2 only 15% were nonfarm. Of particular interest is the fact that 32% of the domain 1 tracts report higher sales on the FCRS. Ten domain 1 tracts reported FCRS reported value of sales of over 10,000 and 3 reported sales of over 50,000. This clearly supports the current practice of including JAS ag tracts which do not qualify as farms in June. Domain 2 tracts also show considerable change and about 15% moved down into the nonfarm interview category while about 35% showed higher reported sales. The other domains show low rates of nonfarm interviews.

FCRS REPORTED VALUE OF SALES CATEGORY
(in thousands of dollars)

Counts, (Percent of Row Total)

JUNE REPORTED VALUE OF SALES in thousands (domain)	< 1	1-2.5	2.5-5	5-10	10-20	20-50	> 50	Total
< 1 (1)	319 (67.58)	108 (22.88)	31 (6.57)	4 (0.85)	2 (0.42)	5 (1.06)	3 (0.64)	472
1-2.5 (2)	111 (15.23)	359 (49.25)	160 (21.95)	73 (10.01)	22 (3.02)	2 (0.27)	2 (0.27)	729
2.5-5 (3)	16 (4.18)	111 (28.98)	139 (36.29)	78 (20.37)	22 (5.74)	13 (3.39)	4 (1.04)	383
5-10 (4)	9 (3.23)	51 (18.28)	66 (23.66)	85 (30.47)	51 (18.28)	11 (3.94)	6 (2.15)	279
10-20 (5)	5 (2.44)	15 (7.32)	19 (9.27)	43 (20.98)	72 (35.12)	39 (19.02)	12 (5.85)	205
20-50 (6-8)	0 (0.00)	5 (1.99)	6 (2.39)	22 (8.76)	36 (14.34)	124 (49.40)	58 (23.11)	251
> 50 (8-12)	0 (0.00)	1 (0.43)	6 (2.60)	3 (1.30)	2 (0.87)	16 (6.93)	203 (87.88)	231
Total	460	650	427	308	207	210	288	2550

The pattern of considerable changes in the lower domains suggests that we should consider the possibility that non-agricultural tracts in June become farms for the FCRS reference period. We cannot check this directly since these tracts are not sampled for FCRS, but we can look at changes for tracts which were in both the 1989 and 1990 June Ag Surveys.[1] This may indicate how likely it is for a non-agricultural tract to change to agricultural or farm status. The table shows that of the 52,339 '89 non-ag tracts which were also in the '90 sample, 601(1.1%) qualified as farms in 90. Of the 4745 '89 non-ag with potential tracts, 200(4.2%) qualified as farms in 1990, about six times the percentage for non-ag tracts with no potential of .7% and about twice the percentage for non-ag tracts with unknown potential of 2.4%. The fact that 49 non-ag no potential tracts from '89 reported sales over $50,000 for 1989 suggests that there may be some tracts that were incorrectly classified, or that changes occurred during the year, and could support the argument for sampling <u>all</u> non-ag tracts to improve coverage. The current definition of a non-ag tract with agricultural potential as described in the AS interviewer's manual is somewhat vague, leaving a subjective decision for enumerators. If a stronger definition of no potential were employed, it might reduce the non-coverage problem and in this case sampling from the with potential and unknown potential tracts might be adequate.

June Ag Survey Area Frame Tracts Surveyed in Both 1989 and 1990
Cross Classified by 1989 Ag Status and 1990 Reported Value of Sales

1989 Ag Status	1990 Reported Value of Sales: (in thousands of dollars)						
	non-ag	<1	1-5	5-50	>50	Total	Total Farms
non-ag with potential	4426	119	134	49	17	4745	200
non-ag unknown potential	2590	32	34	27	4	2687	65
non-ag no potential	44372	199	175	112	49	4490	336
Total non-ag	51388	350	343	188	70	52339	601

[1] Under the JAS area frame rotation about 80% of the segments overlap from one year to the next. There is a possibility that in some states a few segments were redrawn and that tract numbers were reused. Insofar as this is the case, the numbers in the table above would not be completely accurate.

CONCLUSIONS AND RECOMMENDATIONS

The data show clearly that the domain 1 tracts are the main source of nonfarm interviews and nonfarm screenouts. It is also clear that these tracts do not have a very significant impact on overall multiple frame indications for the variables that we examined and should not have much impact on other economic indications, which are dominated by the larger operations. It was suggested that sampling these tracts would only cause minor increases in the cv's for the indications, and would result in substantial savings. The current practice of sampling all of these tracts appears to be unnecessarily inefficient due to the high percentage of nonfarms. These resources could, most likely, be better utilized on other survey improvement activities.

There is also some evidence that sampling non agricultural tracts would improve coverage. There is a general concern that we are not adequately covering new operations which begin operating after June 1, and sampling these non-ag tracts might pick up some of these.

Based on the findings in this research the Survey Research Branch recommends that NASS give serious consideration to adopting a follow-on area frame sampling approach for the FCRS. There appears to be a significant potential for efficiency improvement which could result in either cost reduction or increased precision by more efficient sampling.

The agency should also consider subsampling non-agricultural tracts with potential. This would increase the coverage of the FCRS, and might reduce any negative undercoverage bias,in particular, that for number of farms. This design would also make the FCRS sampling procedures consistent with those in the Agricultural Survey Program.

REFERENCES

1. Cochran, William G. (1977). Sampling Techniques, Third Edition, Wiley & Sons, New York.

2. Dillard, Dave. (1992). "Long Range Plan for the Farm Costs and Returns Survey", NASS Staff Report SMB-91-02.

3. Kott, Phil. (1990). "Variance Estimation when a First Phase Area Sample is Restratified", Survey Methodology, June, 1990, Vol. 16, No. 1, pp. 99-103.

4. Rutz, Jack L., and Cadwallader, Chris. (1991). "1990 Farm Costs and Returns Survey, Survey Administration Analysis", NASS Staff Report SMB-91-04.

5. Statistical Methods Branch. (1990). "1991 FCRS Specifications".

☆U.S. Government Printing Office : 1992 - 311-355/60523