REDRAWING THE 1993 FARM COST AND RETURNS SURVEY LIST SAMPLE TO REDUCE OVERLAP WITH THREE OTHER 1993 SURVEYS AND THE 1992 FCRS

Charles Perry, Jameson Burt, and William Iwig
Charles Perry, USDA/NASS/RD, 3251 Old Lee Hwy, Fairfax, VA 22030

KEY WORDS: Respondent Burden, Multiple Selection, Cross-classification, Sample Coordination

ABSTRACT

This paper describes the Two Stage Algorithm used by the National Agricultural Statistics Service (NASS) to redraw the 1993 Farm Costs and Returns Survey (FCRS) list frame sample. The algorithm was devised to reduce the number of farm and ranch operators in the FCRS sample that are also in other major surveys' samples. It decreased the number of operators in the 1993 FCRS sample that were also in one or more of the four other major surveys from 4,561 to 1,966 (a decrease of 57 percent). It decreased the number of operators in the 1993 FCRS sample that were also in the previous year's FCRS sample from 469 to 29 (a decrease of 94 percent). It increased the number of operators in the 1993 FCRS sample that were not in any of the four other major surveys from 6,617 to 9,212 (an increase of 39 percent).

INTRODUCTION

Because of the length of the Farm Costs and Returns Survey (FCRS) questionnaire, the detail of the information required to answer many FCRS questions, and the intrusive nature of the information asked for in many FCRS questions, there exists a general agreement among National Agricultural Statistics Service (NASS) statisticians and interviewers that the FCRS causes more burden on a respondent than any other NASS survey.

In the last few years, as the number of public and private surveys has increased, statisticians and other government officials have become increasingly concerned about the burden these surveys place on individual respondents. The widely held belief is that the burden experienced by individual respondents tends to increase nonlinearly as they are asked to participate in multiple surveys over a short time period. This means that the burden placed on the population can be reduced by minimizing the number of times any one individual is contacted. For example, burden is reduced by contacting two individuals for one survey each instead of contacting one individual for two surveys.

In addition to reducing burden, there are two other reasons for spreading the burden over the population to the extent possible. First, responding to surveys is one of the costs of providing agricultural statistics. To the extent that these statistics benefit farmers, it is more equitable that their costs be shared equally among farmers. Second, at some level of burden, farmers may refuse to participate in surveys. By spreading the burden to more farmers, the number of refusals should be reduced.

The points above lead most NASS statisticians to conclude that the FCRS sample should be selected using a method that minimizes the overlap within survey cycles between the FCRS sample and other major survey samples and minimizes the overlap across survey cycles between FCRS samples. The general assumption is that decreasing the number of times NASS contacts individual farm and ranch operators (farmers) should decrease the burden placed on the farm population. This should not only improve the overall response rates and quality of NASS surveys, but it should also improve the cooperation farmers give to NASS and other survey organizations.

In "Methods of Selecting Samples In Multiple Surveys To Reduce Respondent Burden," Perry, Burt, and Iwig (1993) presented two new methods of drawing samples in multiple surveys that minimize the burden on the individual population units sampled. These new methods work by spreading the burden from multiple surveys as uniformly as possible over the population without changing the selection probabilities for any survey.

This paper describes how a special case of the Second-Method presented by Perry et al. (1993) was first used to redraw the 1993 FCRS list frame sample so that its overlap with the 1993 Quarterly Agricultural Survey (QAS), Agricultural Labor Survey (ALS), and Cattle and Sheep Survey (CSS) is minimized and then used again to randomly shift the redrawn sample where necessary so that its overlap with the 1992 FCRS is minimized.

ALGORITHM

There are three steps to the Second-Method described by Perry et al. (1993) for drawing multiple surveys samples that minimize the burden on the individual units sampled:

Step 1. Use an equal probability of selection procedure within each stratum to select independent stratified samples for each survey.

Step 2. Cross-classify the population by the stratifications used in the individual surveys.

Step 3. Within each substratum, randomly reassign the samples associated with sampling units having excess burden to population units having less burden. Repeat the process until the burden on individual units sampled is minimized.

A Two Stage Algorithm is given below that first redraws the 1993 FCRS sample to reduce its overlap with the 1993 QAS, ALS, and CSS samples and then redraws the redrawn sample to reduce its overlap with the 1992 FCRS sample. The first stage of the algorithm randomly shifts the 1993 FCRS sample within groups of farmers having the same multivariate stratification for the four 1993 surveys, where possible, to another sample of farmers having lower burden for the four 1993 surveys. The second stage of the algorithm randomly shifts the part of first stage sample that overlaps with the 1992 FCRS sample within groups of farmers having the same multivariate stratification for the 1992 and 1993 FCRS surveys, where possible, to other farmers who were not in the 1992 FCRS sample.

First Stage: Redrawing the 1993 FCRS to Reduce "Overlap" with Three Other Major 1993 Surveys

The sampling procedures used with the 1993 FCRS, QAS, ALS, and CSS satisfy the condition of Step 1 of the Second-Method. The cross-classification of the population required in Step 2 of the Second-Method is produced by sequentially sorting the population records by the individual 1993 FCRS, QAS, ALS, and CSS stratifications.

Throughout the first stage of the algorithm, all of the 1993 samples, except the 1993 FCRS, are assumed to be fixed and hence cannot be reassigned. Thus, the reassignment required in Step 3 of the Second-Method is produced by randomly reassigning the 1993 FCRS sample so that the total burden from all four 1993 surveys on the farmers in the 1993 FCRS sample is minimized.

Since the samples are drawn independently for each individual survey using an equal probability of selection mechanism within each stratum and since within each substratum all farmers have exactly the same stratification on individual surveys, the farmers within each substratum all have the same set of multivariate selection probabilities. Since within each substratum all records have exactly the same probability of experiencing a specified burden level, it follows that they are left in simple random order when they are sorted by ascending (non-decreasing) burden. Since within each substratum the total burden on any individual farmer from the other three 1993 surveys is a non-decreasing function of the final sort order, shifting the 1993 FCRS sample to the first elements of the substratum selects a simple random sample that minimizes the total burden on the individual farmers from all four 1993 surveys. Since the redrawn FCRS sample is selected as a simple random sample within substrata, it has the same expansion factors as the initial FCRS sample. As a result, the first stage of the algorithm can not bias the sample. That is, the FCRS estimates based on the first stage sample have the same expected values as those based on the original FCRS sample.

Table 1 shows by example how the "overlap" of the 1993 FCRS with the other major 1993 surveys is reduced by the first stage of the algorithm. The sample configurations indicate whether the record was sampled (1) or not (0) for the 1993 FCRS, 1992 FCRS, 1993 QAS, 1993 ALS, and 1993 CSS respectively. For example, the sample configurations for the tenth record of the table are respectively 10101 and 00101, which indicate that initially this record was selected for the 1993 FCRS, QAS, and CSS but after the 1993 FCRS was redrawn the tenth record was selected only for the 1993 QAS and CSS. (The 1992 FCRS Survey is not involved in this stage of the redrawing algorithm.)

Since this stage of the redrawing algorithm is independent of the 1992 FCRS sample, some records that are initially in both 1992 and 1993 FCRS samples will end up being only in the 1992 FCRS sample and some records that are initially in only the 1992 FCRS sample will end up being in both the 1992 and 1993 FCRS samples. The first record

Table 1. Example of the Method Used to Redraw the 1993 FCRS in Substrata to Reduce the Burden from the Overlap with Other 1993 Surveys.

Initial Sample Configuration (1)	Redrawn Sample Configuration (2)	Burden Other 1993 Surveys (3)	Initial 1993 FCRS Sample (4)	Uniform Random Number Attached (5)	Redrawn 1993 FCRS Sample (6)	Burden		Rep Code		Burden	
						Initial 1993 Samples (7)	Redrawn 1993 Samples (8)	Initial FCRS Sample (9)	Redrawn FCRS Sample (10)	Initial All Samples (11)	Redrawn All Samples (12)
01000	11000	0	0	0.46313	1	0	60	0	11	60	120
10001	10001	25	1	0.13284	1	85	85	11	25	85	85
00001	10001	25	0	0.05881	1	25	85	0	72	25	85
00001	00001	25	0	0.28983	0	25	25	0	0	25	25
01001	01001	25	0	0.36130	0	25	25	0	0	85	85
00001	00001	25	0	0.47441	0	25	25	0	0	25	25
00001	00001	25	0	0.93427	0	25	25	0	0	25	25
00100	00100	45	0	0.22914	0	45	45	0	0	45	45
00100	00100	45	0	0.86787	0	45	45	0	0	45	45
10101	00101	70	1	0.07412	0	130	70	25	0	130	70
10101	00101	70	1	0.32697	0	130	70	72	0	130	70
00101	00101	70	0	0.61579	0	70	70	0	0	70	70
00101	00101	70	0	0.69437	0	70	70	0	0	70	70
00101	00101	70	0	0.77151	0	70	70	0	0	70	70
00101	00101	70	0	0.93982	0	70	70	0	0	70	70

of Table 1 demonstrates the later case. The second stage of the algorithm reduces the overlap of the 1993 FCRS sample with the 1992 FCRS sample.

Second Stage: Redrawing the 1993 FCRS to Reduce "Overlap" with the 1992 FCRS

The sampling procedures used with the 1992 FCRS and 1993 FCRS satisfy the conditions of Step 1 of the Second-Method, provided no information from the 1992 FCRS sample is used to update the 1993 FCRS frame. The cross-classification of the population and sample required in Step 2 of the Second-Method is produced by sequentially sorting the population records by the 1993 FCRS and 1992 FCRS stratifications.

Since the 1992 FCRS sample is predetermined, the reassignment required in Step 3 of the Second-Method is produced by randomly reassigning the 1993 FCRS sample redrawn at the end of the first stage of the algorithm so that the total burden from the 1992 and 1993 FCRS on the farmers in the 1993 FCRS sample is minimized.

Table 2 shows how the "overlap" of the 1993 FCRS with the 1992 FCRS is reduced by the second stage of the algorithm.

Since the second stage of the algorithm is independent of the other three 1993 survey samples, some of the reduction of overlap between 1993 FCRS sample and the other three 1993 survey samples that was achieved in the first stage of the algorithm will be lost. However, since only about four percent of the 1993 FCRS sample that was selected at the end of the first stage of the algorithm overlaps with the 1992 FCRS, which are the records being redrawn in this stage of the algorithm, the loss of overlap reduction with respect to the other 1993 survey samples will be small.

Reasons for Using a Two Stage Algorithm Instead of a Single Stage Algorithm to Redraw the 1993 FCRS

The first stage of the Two Stage Algorithm can be converted to a one stage algorithm that minimizes the "overlap" of the 1993 FCRS sample with the 1993 QAS, ALS, and CSS samples and the 1992 FCRS sample. So, why choose a two stage algorithm?

The primary reason the Two Stage Algorithm was used instead of a single stage algorithm to redraw the 1993 FCRS was to minimize the potential for introducing bias into the FCRS estimates. Since the first stage of the Two Stage Algorithm is unconditionally unbiased, bias can only be introduced in the second stage of the Two Stage Algorithm. Potential for bias will exist if 1992 FCRS data have been used to update the 1993 FCRS frame.

In the second stage of the Two Stage Algorithm, the only part of the sample that is redrawn is the part that overlaps with the 1992 FCRS sample. Since

Table 2. Example of the Method Used to Redraw the 1993 FCRS in Substrata to Reduce the Burden from Overlap with the 1992 FCRS Sample.

First Stage Sample Configuration (1)	Final Redrawn Sample Configuration (2)	Burden 1992 FCRS Survey (3)	Sampled Stage 1 1993 FCRS (4)	Uniform Random Number Attached (5)	Sampled Redrawn 1993 FCRS (6)	Burden		Rep Code		Burden	
						Stage 1 92 & 93 FCRS (7)	Redrawn 92 & 93 FCRS (8)	Stage 1 FCRS Sample (9)	Redrawn FCRS Sample (10)	Stage 1 All Samples (11)	Redrawn All Samples (12)
10001	10001	0	1	0.12824	1	60	60	5	5	90	90
10001	10001	0	1	0.61343	1	60	60	7	7	90	90
10101	10101	0	1	0.89720	1	60	60	13	13	135	135
00001	10001	0	0	0.37579	1	0	60	0	2	30	90
00001	00001	0	0	0.59949	0	0	0	0	0	30	30
11001	01001	60	1	0.06257	0	120	60	2	0	145	85

this part of the sample amounts to only about four percent of the sample, it follows that when the Two Stage Algorithm is used only about four percent of the redrawn sample has any potential for biasing the 1993 FCRS estimates.

In contrast, when the one stage algorithm is used every element in the redrawn sample is drawn conditional on the 1992 FCRS sample. Thus, with the one stage algorithm, every element of the redrawn sample has a potential for biasing the 1993 FCRS estimates.

In summary, unless it can be shown that no information from the 1992 FCRS sample is used to update the 1993 FCRS frame, both the single and two stage algorithms have potential for biasing the 1993 FCRS estimates. With the single stage algorithm, every element is redrawn conditional on the 1992 FCRS sample; but, with the two stage algorithm only about four percent of the sample is redrawn conditional on the 1992 FCRS sample. Thus it follows that there is much less potential for biasing the 1993 FCRS estimates with the Two Stage Algorithm than there is with the single stage algorithm.

In addition to the greatly reduced potential for biasing the 1993 FCRS estimates, the Two Stage Algorithm provides considerably more burden reduction than the single stage algorithm.

RESULTS

The burden reduction achieved by redrawing the 1993 FCRS can be summarized by the changes (or percentage changes) that occurred from the initial to the redrawn sample in the number of farmers selected for none, one, two, three, and four of the other surveys (1992 FCRS, 1993 QAS, ALS, and CSS). The next paragraph summarizes the burden reduction achieved by the redrawing algorithm for the complete 1993 FCRS sample.

Redrawing the 1993 FCRS sample of 11,178 farmers:

1. Increased the number of 1993 FCRS samples that were not selected for any of the other four surveys from 6,617 to 9,212 (an increase of 39 percent),
2. Decreased the number of 1993 FCRS samples that were selected for only one of the other four surveys from 3,603 to 1,598 (a decrease of 56 percent),
3. Decreased the number of 1993 FCRS samples that were selected for exactly two of the other four surveys from 843 to 325 (a decrease of 61 percent),
4. Decreased the number of 1993 FCRS samples that were selected for exactly three of the other four surveys from 107 to 41 (a decrease of 62 percent),
5. Decreased the number of 1993 FCRS samples that were selected for all of the other four surveys from 8 to 2 (a decrease of 75 percent), and
6. Decreased the number of 1993 FCRS samples that were also in the previous year's FCRS sample from 469 to 29 (a decrease of 94 percent).

These statistics show the effectiveness of the two stage redrawing process.

Table 3 provides detailed summary statistics by sampling stage for the initial, first stage, and redrawn sample. In Table 3, the 11,178 farmers in each sample are first broken down according to the number of times they were selected for other surveys. Then the farmers in each sample are further broken down according to the specific combination of surveys for which they were selected.

Table 3. For All Strata, U.S.: The Number of FCRS Samples and Percent of Total FCRS Sample by Sampling Configuration for the Initial, First Stage Redrawn, and Final Redrawn 1993 FCRS Samples.

Combination of Surveys (1)	Sampling Configuration† 93 FCRS 92 FCRS 93 QASS 93 ALS 93 CSS (2)	Count Initial (#) (3)	Count Stage 1 (#) (4)	Count Redrawn (#) (5)	Percent Initial (%) (6)	Percent Stage 1 (%) (7)	Percent Redrawn (%) (8)	Change Initial to Stage 1 Count (#) (9)	Change Initial to Stage 1 Percent (%) (10)	Change Stage 1 to Redrawn Count (#) (11)	Change Stage 1 to Redrawn Percent (%) (12)	Change Initial to Redrawn Count (#) (13)	Change Initial to Redrawn Percent (%) (14)
93 FCRS alone	1 0 0 0 0	6617	9095	9212	59.2	81.4	82.4	+2478	+37%	+117	+1%	+2595	+39%
	Total	6617	9095	9212	59.2	81.4	82.4	+2478	+37%	+117	+1%	+2595	+39%
93 FCRS plus one other survey	1 0 0 0 1	1110	597	662	9.9	5.3	5.9	−513	−46%	+65	+11%	−448	−40%
	1 0 0 1 0	336	74	93	3.0	0.7	0.8	−262	−78%	+19	+26%	−243	−72%
	1 0 1 0 0	1966	668	836	17.6	6.0	7.5	−1298	−66%	+168	+25%	−1130	−57%
	1 1 0 0 0	191	305	7	1.7	2.7	0.1	+114	+60%	−298	−98%	−184	−96%
	Total	3603	1644	1598	32.2	14.7	14.3	−1959	−54%	−46	−3%	−2005	−56%
93 FCRS plus two other surveys	1 0 0 1 1	74	24	30	0.7	0.2	0.3	−50	−68%	+6	+25%	−44	−59%
	1 0 1 0 1	414	189	225	3.7	1.7	2.0	−225	−54%	+36	+19%	−189	−46%
	1 0 1 1 0	147	39	60	1.3	0.3	0.5	−108	−73%	+21	+54%	−87	−59%
	1 1 0 0 1	59	44	4	0.5	0.4	0.0	−15	−25%	−40	−91%	−55	−93%
	1 1 0 1 0	12	6	1	0.1	0.1	0.0	−6	−50%	−5	−83%	−11	−92%
	1 1 1 0 0	137	73	5	1.2	0.7	0.0	−64	−47%	−68	−93%	−132	−96%
	Total	843	375	325	7.5	3.4	2.9	−468	−56%	−50	−13%	−518	−61%
93 FCRS plus three other surveys	1 0 1 1 1	45	26	31	0.4	0.2	0.3	−19	−42%	+5	+19%	−14	−31%
	1 1 0 1 1	6	2	0	0.1	0.0	0	−4	−67%	−2	−100%	−6	−100%
	1 1 1 0 1	33	21	6	0.3	0.2	0.1	−12	−36%	−15	−71%	−27	−82%
	1 1 1 1 0	23	9	4	0.2	0.1	0.0	−14	−61%	−5	−56%	−19	−83%
	Total	107	58	41	1.0	0.5	0.4	−49	−46%	−17	−29%	−66	−62%
93 FCRS plus four other surveys	1 1 1 1 1	8	6	2	0.1	0.1	0.0	−2	−25%	−4	−67%	−6	−75%
	Total	8	6	2	0.1	0.1	0.0	−2	−25%	−4	−67%	−6	−75%
Total		11,178	11,178	11,178	100.0	100.0	100.0	0	0%	0	0%	0	0%

† Note: The **Sampling Configuration** indicates the sampling pattern for the 93FCRS, 92FCRS, 93QAS, 93ALS and 93CSS (Cattle and Sheep). For example, a 10001 indicates sample units selected for only the 93FCRS and 93CSS.

Table 4. For All Strata, U.S.: The Estimated 1993 Population Control Variable Totals for the Initial, First Stage Redrawn, and Final Redrawn 1993 FCRS Sample.

Control Variable† (1)	Population Total (#) (2)	90% Lower Bound (#) (3)	90% Upper Bound (#) (4)	Estimated Total Initial Sample (#) (5)	Estimated Total Stage 1 Sample (#) (6)	Estimated Total Redrawn Sample (#) (7)	Exact CV of Estimated Total (%) (8)	Estimated CV Initial Sample (%) (9)	Estimated CV Stage 1 Sample (%) (10)	Estimated CV Redrawn Sample (%) (11)
Total Land in Farm	862.52	47.04	1678.01	791.28	798.17	790.29	57.30	2.31	2.46	2.43
Farm Value of Sales	150.53	148.44	152.62	151.96	151.93	151.76	0.84	0.86	0.85	0.84
Total Cropland	363.62	327.21	400.03	363.54	365.83	364.74	6.07	1.11	1.07	1.06
On-Farm Grain Storage	947.58	902.34	992.81	992.67	990.91	997.68	2.89	2.98	2.88	2.99
All Cattle & Calves	850.13	811.07	889.19	821.80	829.07	840.22	2.78	3.05	3.21	3.61
Total Hogs & Pigs	547.15	457.17	637.12	588.76	576.22	585.09	9.97	5.33	4.08	4.29
All Sheep	114.96	88.60	141.32	106.15	103.90	103.23	13.90	13.02	13.13	13.20
Farm Workers Hired	127.21	58.80	195.61	102.40	102.54	102.05	32.59	5.22	5.31	5.33

† All missing control values were set to zero before any computations were performed. The units of measure for the population totals and their estimates are:

1,000,000 acres for Total Land;
1,000,000 acres for Total Cropland;
100,000 head for Cattle & Calves, Hogs & Pigs, and Sheep;
$1,000,000,000 for Farm Value Of Sales;
10,000,000 bushels for On-Farm Grain Storage;
10,000 workers Farm Workers Hired.

Control Variable Estimates

The similarity of the redrawn sample to the initial sample can be assessed by examining associated population control values (auxiliary covariates used for stratification during sample selection). Table 4 shows for each of eight population control values the estimated totals and coefficients of variation that are associated with the initial, first stage, and redrawn samples.

Table 4 shows that the three samples are very similar with respect to the estimates derived from them. The three estimates tend to fall at almost the same place in the respective confidence intervals. The same is true for the estimated coefficients of variation. If the estimated coefficient of variation for an item is small (large) relative to the true coefficient of variation for one sample then it is also small (large) for the other two samples.

CONCLUSIONS

Redrawing the 1993 FCRS list frame sample of 11,178 farmers reduced the burden on the individual farmer sampled by:

1. Decreasing the number of 1993 FCRS samples that were in one or more of the other samples (1993 QAS, ALS, CSS or 1992 FCRS) from 4,561 to 1,966 (a decrease of 57 percent),
2. Decreasing the number of 1993 FCRS samples that were in the 1992 FCRS sample from 469 to 29 (a decrease of 94 percent), and
3. Increasing the number of 1993 FCRS samples that are not in any of the other four surveys from 6,617 to 9,212 (an increase of 39 percent).

The Two Stage Algorithm used to redraw the 1993 FCRS has very little potential to introduce bias into the 1993 FCRS estimates. In fact, only at the second stage, which involves less than four percent of the sample, does it have any potential of introducing bias. At the second stage, the potential to introduce bias is limited to those samples which have had their 1992 FCRS sample information used to update the 1993 FCRS frame.

The primary reason that a Two Stage Algorithm was chosen over a single stage algorithm for redrawing the 1993 FCRS was to minimize the potential to introduce bias into the FCRS estimates. A secondary reason for the choice was that the Two Stage Algorithm also provides considerably more burden reduction than the single stage algorithm.

Note

A five stage algorithm, which is an extension of the Two Stage Algorithm used in 1993, will be introduced in 1994. The five stage algorithm will help control the potential for bias in the FCRS estimates that results from application of the Two Stage Algorithm two years in a row. With the five stage algorithm, the potential for bias in the FCRS estimates in 1994 and subsequent years will be only slightly greater than in 1993, while the reduction in burden will remain essentially the same as with the Two Stage Algorithm. For further detail, see Perry, Burt, and Iwig (1994).

REFERENCES

Perry, Charles R., Jameson C. Burt, and William C. Iwig; "Methods Of Selecting Samples In Multiple Surveys To Reduce Respondent Burden"; *Proceedings of the International Conference on Establishment Surveys*; ASA; Alexandria, Virginia; June 1993.

Perry, Charles R., Jameson C. Burt, and William C. Iwig; *Redrawing the 1993 Farm Costs and Returns Survey List Frame Sample to Reduce its Overlap with the 1992 FCRS and Three Other Major 1993 Surveys*; USDA/NASS; Washington D.C.; October 1994.