TEST OF MULTIPLE FRAME SAMPLING TECHNIQUES FOR AGRICULTURAL SURVEYS: NEW BRUNSWICK, 1978

Barbara Armstrong, Statistics Canada

1. Introduction

The Agriculture Division of Statistics Canada conducts the Agriculture Enumerative Survey (AES) every July. The AES is a multi-purpose survey based on an area sample producing estimates for crops, livestock and expense items for all provinces but the Prairies. A recurring problem with the survey has been that the sample size allocated to smaller provinces has been insufficient to produce good provincial estimates. It was decided, therefore, to test multiple frame sampling in one of these provinces to determine whether this technique could increase the efficiency of estimates and to study the operational problems associated with the technique. The province chosen for testing was New Brunswick.

2. Sample Design for the 1978 New Brunswick Test

A multiple frame survey is one which employs two or more sampling frames to produce estimates for a specific survey population. In the 1978 New Brunswick test an area and a list frame were combined to produce estimates for agricultural items. Use of the area frame was essential to ensure complete coverage of the population. The list frame was introduced to improve the efficiency of the sample design.

The AES, in its present form employs an extreme type of multiple frame sampling. The AES area sample is supplemented by a group of very large farms (large with respect to some key items) taken from the updated 1976 Census of Agriculture list and included in the sample with probability 1. This group of farms, referred to as specified farms, is included in the sample as a separate, complete enumeration stratum for two reasons. Since the AES is a probability sample survey, the values of items for farms in the sample are blown up to represent a larger group of farms. If these specified farms were not identified prior to the survey and were by chance picked up in the area sample the resulting estimates would be blown up out of proportion since they are not typical farms. On the other hand, since these farms contribute a significant amount to the provincial totals, their chance exclusion from the sample would result in lower estimates. In either case--chance inclusion or exclusion--the estimates would vary considerably and so specified farms are included with probability 1. This feature of the AES sample design was retained in the multiple frame test also, as described below.

The list frame for the New Brunswick multiple frame test, was the updated 1976 Census of Agriculture list. Very small farms were excluded from the Census list and the remaining farms formed the list sampling frame. This frame was stratified as shown in Table 1 and a simple random sample selected from each stratum. Starting with stratum 1 and continuing sequentially, a farm was assigned to the first stratum for which it met the stratum criterion. As we shall see later, this method of stratifying the list proved to be inefficient for the estimation of one of the stratifying items.

TABLE 1: Stratification of the List Frame

Stratum	Stratum Definition	Population Size	Sample Size
1	Specified farms as defined in AES	69	69
2	Total sales > $75,000	200	88
3	Total potatoes > 50 acres	261	60
4	Total cattle > 40	448	60
5	Total pigs > 30	83	30
6	Total potatoes > 10 acres	204	15
7	Total cattle > 25	858	30
8	Total pigs > 8	84	15
		2,207	367

Sample allocation to strata was based on a trial and error method. The allocation ultimately chosen was the one which gave the best combination of coefficients of variation for the three key items (i.e. potatoes, cattle, pigs) in New Brunswick.

The design of the area sample was the same as it had been since the last redesign in 1974. Enumeration areas (EA's) as defined in the 1971 Census of Agriculture were the first stage sampling units. EA's were stratified using Census data (summarized at the EA level) as shown in Table 2. A stratified replicated random sample of EA's was then selected. Again, sample allocation was based on trial and error with the chosen allocation giving the best combination of coefficients of variation for key items in the province.

Selected EA's were divided into roughly equi-sized pieces of land or segments. These segments became the second stage sampling units. A sample of one or more segments was selected from each selected EA depending on the size of the EA. All operators with land within selected segments were enumerated.

TABLE 2: Stratification of Area Frame

Stratum	Stratum Definition	Population Size	No. of Replicates	No. of EAs per Replicate	Sample Size
	(EA Level)	(# EAs)			(# Farms)
1	Total chickens ≥ 25,000	25	6	2	29
2	Total potatoes ≥ 6,000	30	15	2	212
3	Total pigs ≥ 400	21	7	2	32
4	Total cattle ≥ 500	62	10	2	92
5	X ≥ 47	56	7	2	42
6	X ≥ 14	136	9	2	30
7	Remaining EAs	219	16	1	29
8	OLD nonagricultural EAs	221	6	1	0
9	NEW nonagricultural EAs	57	2	1	0
					466

NOTE: X = [R * Livestock + R * Cropland] * 20,000

$$\text{where } R = \frac{\text{item summed at EA level}}{\text{item summed over all agricultural EAs}}$$

As 1978 was our first test of multiple frame, an adequate sample had to be allocated to the area frame in order to secure the regular AES area sample estimates. Fortunately, due to a F.L.I.P.

grant (Federal Labour Intensive Program) from the federal government, the sample allocation to New Brunswick was increased to such an extent that we were able not only to add a list sample to the existing area sample but to actually increase the area sample above its 1977 level.

3. Estimation Procedures

(a) Area Sample Estimator

As shown in Table 2, selection of first stage sampling units in the AES is replicated. Data from all EA's within a replicate are blown up to the stratum level. The estimator for replicate k of stratum h is:

$$y_{hk} = \frac{M_h}{m_h} \sum_{i=1}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{j=1}^{n_{hi}} y_{hij}$$

where M_h = the number of EA's in stratum h

m_h = the number of EA's selected per replicate for stratum h

N_{hi} = the number of segments in the ith selected EA of stratum h

n_{hi} = the number of selected segments in the ith selected EA of stratum h

y_{hij} = data value for the jth segment of the ith EA in stratum h.

Usually a farm enumerated within a segment has part of its land lying inside the segment and part outside. For such farms, data values are reduced to the segment level by applying a weight equal to the ratio of the farm's land inside the segment (excluding woodland) to total land operated on the farm (excluding woodland). It is this "weighted" data value for each farm within a segment which is summed to give the segment total y_{hij}.

Having obtained estimates for all replicates, the stratum estimate can be calculated and is

$$Y_h = \frac{y_{h1} + y_{h2} \cdots y_{hr_h}}{r_h}$$

where r_h = number of replicates in stratum h.

To obtain provincial estimates, stratum estimates are summed

$$\hat{Y}_A = \sum_{h-1}^{H} Y_h$$

where H = number of strata.

Variances take the form

$$V(\hat{Y}_A) = \sum_{h=1}^{H} \frac{1}{r_h(r_h-1)} \{(Y_{h1} - Y_h)^2 + \ldots + (Y_{hr} - Y_h)^2\}$$

(b) Multiple Frame Estimators

(i) Screening Estimator

In multiple frame estimation, the list sample estimates for that portion of the population covered by the list frame. With the multiple frame screening estimator, the area sample estimates only for that portion of the area frame not covered by the list frame. The screening estimator takes the form:

$$\hat{Y}_S = \hat{Y}_L + \hat{Y}_{NOL}$$

where $\hat{Y}_S$ = screening estimate

$\hat{Y}_L$ = list frame estimate

$\hat{Y}_{NOL}$ = area frame estimate for farms not on the list frame (non-overlap domain).

Two additional terms are defined here as well.

$\hat{Y}_{OL}$ = area frame estimate for segment farms on the list frame (overlap domain)

$\hat{Y}_A$ = area frame estimate (as described in section 3(a)).

It should be noted that

$$\hat{Y}_A = \hat{Y}_{OL} + \hat{Y}_{NOL}.$$

The variance of the screening estimate is simply

$$\mathrm{Var}(\hat{Y}_S) = \mathrm{Var}(\hat{Y}_L) + \mathrm{Var}(\hat{Y}_{NOL}).$$

(ii) The Hartley Estimator

Although the screening estimate will result in smaller sampling errors than the area estimate there is one major drawback. It makes no use of data for area sample farms falling in the overlap domain. A second type of multiple frame estimator--the Hartley estimator--does make use of this data by combining estimates of the overlap domain from the two frames. The Hartley estimator is:

$$\hat{Y}_H = \hat{Y}_{NOL} + q\,\hat{Y}_{OL} + p\,\hat{Y}_L$$

where p = weight given to the list estimate

q = weight given to the area estimate of the overlap domain

$p + q = 1.$

An optimum value of p can be determined (see Appendix 1 for derivation of p_{opt}). Using this value of p, the sampling error for $\hat{Y}_H$ will be less than for our first estimator $\hat{Y}_S$. Previous studies have shown, however, that the reduction is small. The weight attached to the list frame is usually so large relative to the weight given to the area frame (the list estimate being far more efficient than the area) that little efficiency is gained by including the area estimate of the overlap. However, if the area data for overlap is collected regardless (as was done in the New Brunswick test) it makes sense to use the Hartley estimate even if the reduction in sampling error is small. The formula for the variance estimate of the Hartley estimator is given in Appendix 1.

4. 1978 New Brunswick Estimates

The screening estimates for the 1978 New Brunswick test are presented in Table 3. Comparing regular AES weighted estimates with multiple frame estimates, there were two important observations to note. First of all, coefficients of variation for multiple frame estimates were significantly lower than for weighted--often by as much as 50%. The second observation was that the level of multiple frame estimates appeared to be generally higher than for weighted estimates. In only 4 out of the 21 estimates displayed in Table 3 were multiple frame estimates lower than the weighted.

The first observation needs little comment. The list frame is a more efficient sampling frame and we therefore expected the coefficients of variation to decrease sharply with the introduction of a list sample.

It was with the second observation with which there was the most concern. However, although there did appear to be a tendency for multiple frame estimates to be higher than the weighted estimates, it is interesting to note that the same tendency could be seen with the published estimates. Published estimates are compiled by subject matter experts taking into account estimates from all their sources (of which the AES is one). Of the 17 cases where the multiple frame estimate was higher than the weighted, 11 of the published estimates were also higher than the weighted estimate. Of the 4 cases where multiple frame estimated lower than the weighted, the published estimates for all 4 were lower than the weighted as well. Thus the "level" problem of multiple frame estimates does not appear as extreme taking this into account although we shall be investigating it further in Section 10.

TABLE 3: New Brunswick Estimates - 1978

Item	Weighted Estimate	C.V.	Multiple Frame Estimate	C.V.	Published Figure
Total area	1,090,235	8.5	1,271,419	6.0	1,090,200
Potatoes	63,355	15.4	70,318	7.2	58,000
Mixed grains	3,945	46.5	5,077	37.8	6,800
Oats	37,357	17.2	45,771	9.4	42,000
Barley	5,210	26.6	5,946	19.5	7,200
Tame hay	158,628	12.0	197,167	7.5	180,000
Spring wheat + Winter wheat	7,903	-	8,405	-	9,600
Corn for grain	380	31.9	510	33.8	500
Total crops	298,238	9.8	356,746	5.4	298,200
Improved land for pasture	97,987	13.3	119,155	10.0	-
Total cattle	109,350	12.0	118,844	5.4	113,000
Milk cows > 2 yrs	35,277	20.2	34,801	9.6	31,000
Beef cows > 2 yrs	19,300	13.8	23,815	11.8	23,000
Dairy heifers (1-2 yrs)	11,071	23.4	9,804	11.0	7,600
Bulls	2,600	16.6	2,743	16.1	2,700
Steers	7,200	12.1	8,821	11.1	8,200
Calves	25,500	11.1	29,365	11.1	29,700
Total pigs	47,610	18.8	60,925	12.1	43,000
Sows & gilts	10,205	44.1	7,741	14.1	5,000
Boars	700	38.5	587	17.8	400
Market pigs (<3 mon) ..	21,200	16.9	29,148	14.1	18,000

5. Comparison of Screening and Hartley Multiple Frame Estimates

Table 4 shows the weighted, screening and Hartley estimates for the four key items in New Brunswick. As well, the p and q values of the Hartley estimate are given.

TABLE 4: Comparison of Different Estimates for N.B., 1978
(C.V.'s in brackets)

	Weighted Estimate	Standard Error	Screening Estimate	Standard Error	Hartley Estimate	Standard Error	p	q
Total Area ...	1,090,238 (8.5%)	92,730	1,271,419 (6.0%)	75,726	1,217,572 (5.9%)	71,487	0.70	0.30
Potatoes	63,355 (15.4%)	9,788	70,318 (7.2%)	5,050	69,900 (7.2%)	5,023	0.94	0.06
Total Cattle .	109,351 (12.0%)	13,124	118,847 (5.4%)	6,410	117,577 (5.2%)	6,159	0.75	0.25
Total Pigs ...	47,610 (18.8%)	8,960	60,925 (12.1%)	7,365	55,581 (10.4%)	5,751	0.59	0.41

Results for total area, potatoes, and total cattle are as expected. Standard errors have been reduced slightly by using the Hartley estimate and levels of the Hartley and screening estimates are comparable. It is the total pig estimates which are interesting. The Hartley estimate is lower than the screening by 5,000. The standard error for the Hartley estimate is also substantially lower,thereby producing a coefficient of variation of 10.5% (as opposed to 12.1% for the screening estimate). The reason for this lies with the list estimate for total pigs. Since the variance of the list estimate was relatively high (although still about half that of the area estimate for the overlap domain), the result was that a lower weight of p was assigned to the list estimate of the overlap domain and consequently gains in efficiency were realized using the combination of estimates. The value of p for total pigs was 0.59 and the Hartley estimate for pigs was therefore:

$$\hat{Y}_H = \hat{Y}_{NOL} + 0.41\ \hat{Y}_{OL} + 0.59\ \hat{Y}_L.$$

6. List Estimates--Comparison with 1976 Census

For multiple frame sampling to be effective in producing good estimates, it is essential that the list sample provides a good estimate for that portion of the population covered by the list frame. As a first step in evaluating list estimates, 1978 list estimates were compared by stratum with the corresponding totals from the updated 1976 Census.

TABLE 5: Comparison of 1976 Updated Census Totals
and 1978 List Estimates, N.B.

Stratum	Total Area		Potatoes		Total Cattle		Total Pigs	
	1976	1978	1976	1978	1976	1978	1976	1978
1	50,954	68,888	2,123	3,976	13,416	13,359	18,443	19,427
2	99,644	93,498	17,721	19,148	7,145	5,082	4,191	1,809
3	87,284	102,155	26,243	25,461	4,005	2,462	561	3,306
4	203,822	245,601	678	866	39,186	37,438	3,293	12,484
5	19,518	19,973	168	357	1,895	1,494	9,278	9,907
6	43,453	58,303	6,193	8,418	2,790	3,006	268	1,210
7	242,836	229,229	237	86	34,303	26,255	1,745	4,976
8	13,964	12,634	39	34	912	1,058	1,228	885
Total	761,475	830,281	53,402	58,336	103,652	90,154	39,007	54,005
C.V.	-	5.5%	-	6.3%	-	4.6%	-	12.9%

Table 5 shows corresponding figures from 1976 and 1978 for four key items in New Brunswick. As was mentioned in the last section, the list estimate for total pigs had a high coefficient of variation. This is not surprising since at the design stage, pigs were given the lowest priority of all stratification variables. This low priority may have been responsible for the large difference between the census total and the list estimate. Notice that the cattle strata (strata 4 and 7) contributed 17,460 to the total pig estimate while the census total for these two strata was only 5,038. Since cattle had a higher priority than pigs at stratification, if a farm met the criteria for both the large cattle and the large pig strata, it was put in the large cattle stratum. Thus, the resulting stratum was homogeneous for total cattle but not for total pigs. The result was that, because some large pig farms were picked up in the sample, the stratum produced an inflated estimate for total pigs. For future surveys it would be wise to consider giving pigs an equal priority with cattle.

The list estimate for potato acreage was 58,336. The portion of the census list used as a list frame in New Brunswick accounted for 96.18% of all potatoes in the province in 1976. Thus, if we had had only the list sample from which to estimate, results would have been similar to the 1978 New Brunswick potato survey (a survey based on a list sample). The fact is, however, that we had an area sample as well which showed that area sample farms not found on the list frame contributed 11,982 acres of potatoes to the provincial estimate. Farms which were on the census list but not on the reduced list frame used in New Brunswick accounted for 4,680 acres. This means that the estimate coming from farms which were not found on the census list was 7,302 acres. It is not surprising, therefore, that multiple frame estimates were higher than the New Brunswick potato survey estimates, the latter being based solely on a list frame.

7. Condition of List Frame

The main problem in using any list frame to select a sample is that such frames become out of date very quickly. The census data upon which stratification in New Brunswick was based was two years old when the survey was run. Table 6 shows for list sample farms to what stratum each farm was assigned before the survey and to what stratum it should have been assigned based on survey data.

TABLE 6: List Sample Stratum Changes; 1976-1978

1976 Stratum	1978 Stratum										Total
	1	2	3	4	5	6	7	8	9*	Out of Business	
1	51	8	1	9	-	-	-	-	-	-	69
2	2	66	9	2	1	-	-	-	6	2	88
3	1	11	35	-	-	4	-	-	-	9	60
4	2	3	1	52	-	-	-	-	-	2	60
5	1	-	1	4	14	-	5	1	2	2	30
6	-	-	4	-	-	6	-	1	2	2	15
7	-	-	-	8	1	-	11	-	5	5	30
8	-	-	-	-	2	-	-	4	8	1	15
Total .	57	88	51	75	18	10	16	6	23	23	367

*Stratum 9 refers to farms which no longer meet the criteria of any list stratum.

As shown in Table 6, the strata for large farms (strata 1 - 5) tended to be more stable than the lower strata thereby reducing or eliminating the benefits of stratification for these lower strata. Thus, if may be advisable to put stricter limits on strata to be included in the list frame. The smaller sized farms are the ones which can more readily "jump" strata so that the area frame would estimate almost as well for these farms. This would also leave the entire list sample to estimate more efficiently for the larger sized farms.

8. Removing Medium Strata from the List

Eliminating a stratum from list frame coverage changes the multiple frame sample in two ways. First, list sample farms in that stratum do not contribute to the list estimate. Secondly, area sample farms which overlap with that stratum in the population frame (and were, therefore, part of the area overlap domain) are now part of the non-overlap domain. For the screening estimate described in Section 3, then, the number of farms contributing to the list portion of the estimate decreases while the number of farms contributing to the area portion of the estimate increases.

Table 7 shows the effect on multiple frame (screening) estimates of removing medium-sized strata (strata 6, 7, 8) from list frame coverage. Coefficients of variation for all items increased only slightly (over the full-multiple frame estimates) but this increase was significant considering that the sample size increased as well. Coefficients were still lower than for weighted estimates. Also worth noting is the fact that allowing the area frame to estimate for these strata reduced the level of resulting estimates for all items. Therefore the area frame estimates were lower than the list frame for these list strata. This may have been due to the problem of out-of-date stratification in the lower list strata as mentioned in the previous section. Since the area frame appeared to estimate better than the list for lower strata the conclusion reached in Section 7 is reinforced here, i.e. the medium-sized strata should be removed from the list frame coverage for the 1979 survey.

TABLE 7: Estimates for Reduced List Frame Coverage (C.V.'s in brackets)

Item	Weighted Estimate	Full Multiple Frame Estimate	Removing Strata 6, 7, 8 from List Frame
Total area .	1,090,235 (8.5%)	1,271,419 (6.0%)	1,220,732 (6.4%)
Potatoes ...	63,355 (15.4%)	70,318 (7.2%)	66,600 (7.2%)
Total Cattle	109,350 (12.0%)	118,844 (5.4%)	114,483 (6.5%)
Total Pigs .	47,610 (18.8%)	60,925 (12.1%)	57,082 (12.1%)
Sample Size.	535	569	591

9. Eliminating Strata by Commodities

Table 8 shows what happened as potato strata (3 and 6), cattle strata (4 and 7), and pig strata (5 and 8) were, in turn, dropped from list frame coverage. As strata based on a certain item are dropped from the list frame one would expect the efficiency of the estimate for that item to drop, since the area sample is now estimating almost entirely for it. As shown in Table 8 this was, in fact, what happened. As potato strata were removed the coefficient of variation for potatoes increased from 7.2 to 11.1; as cattle strata were removed the coefficient for cattle increased from 5.4 to 11.7; and as pig strata were removed the coefficient for total pigs increased from 12.1 to 16.6.

What is interesting to note in Table 8 is what happened to the estimate for total pigs as the cattle strata (4 and 7) were removed from list coverage. By allowing the area sample to estimate for the cattle strata, the estimate for total pigs was reduced to the level of the 1978 weighted estimate. As well, the coefficient of variation for this estimate was less than for both the weighted and the full multiple frame estimate. This tended to support the suspicion raised in Section 6 that the list sample for the cattle strata over-estimated for total pigs. It also reinforced the recommendations that, for stratification purposes, total pigs should be given equal or higher priority with total cattle for future design of list frames.

TABLE 8: Eliminating List Strata for Certain Commodities (C.V.'s in brackets)

Item	Weighted Estimate	Full Multiple Frame Estimate	Eliminate Strata 3, 6 (Potatoes)	Eliminate Strata 4, 7 (Cattle)	Eliminate Strata 5, 8 (Pigs)
Total Area ...	1,090,235 (8.5%)	1,271,419 (6.0%)	1,216,582 (6.0%)	1,145,174 (8.1%)	1,263,446 (6.0%)
Potatoes	63,355 (15.4%)	70,318 (7.2%)	62,191 (11.1%)	70,151 (7.1%)	69,944 (7.2%)
Total Cattle .	109,350 (12.0%)	118;844 (5.4%)	118,585 (5.2%)	106,582 (11.7%)	119,271 (5.6%)
Total Pigs ...	47.610 (18.8%)	60,925 (12.1%)	56,868 (12.0%)	48,100 (9.7%)	60,945 (16.6%)
	535	569	591	585	537

10. Overlap Determination

The high level of multiple frame estimates indicated that there could be a problem with the determination of overlap between the area sample and list frame. Recall that only area sample farms which are not found on the list frame contribute to the area portion of the multiple frame estimate. If area sample farms which appear on the list frame are not identified as such, then resulting multiple frame estimates will be inflated.

To check on the overlap determination, a list of area sample farms not matched to the list frame at head office was sent to the regional office to verify that they were true "non-matches" to the list frame. The list was returned with comments indicating that several farms were on the list frame but were now operating under different names. While the regional office assumed that we were matching farms (i.e. pieces of land) we were in fact matching farm operators. If an area sample farm appeared on the list frame under a different operator's name, it was next to impossible for head office to identify these as being the same farm. Therefore we had to be content with matching farm operators rather than pieces of land.

However, after this confusion with regional office over "farms" versus "farm operators", it occurred to us that the same confusion had probably arisen with enumerators of list sample farms. Again, since it is a farm operator that is selected from the list frame, the list sample farm should be classified as being out-of-business if the operator is no longer the same. As this had never been explicitly told to enumerators they tended not to make this distinction. Farms were enumerated as usual even if the farm operator had changed. Thus, it was necessary to go through the list questionnaires to pick out all such farms (of which there were 12). Removing the contribution of these farms from multiple frame estimates, Table 9 giving revised estimates was produced.

TABLE 9: Revised Multiple Frame Estimates

Item	Original Multi-frame Estimate	Revised Multi-frame Estimate	Weighted Estimate	Published Estimate
Total area .	1,271,419 (6.0)	1,241,888 (6.2)	1,090,235 (8.5)	1,090,200
Potatoes ...	70,318 (7.2)	66,479 (7.5)	63,335 (15.4)	58,000
Total Cattle	118,844 (5.4)	115,334 (5.6)	109,350 (12.0)	113,000
Total Pigs .	60,925 (12.1)	58,388 (12.3)	47,610 (18.8)	43,000

The revised multiple frame estimates do show improvement over the original estimates in terms of the level of estimates. Thus much of our "overlap" problem was largely definitional. More explicit instructions and clearer definitions of out-of-business list farms will be given to enumerators next year. Further work, however, will have to be done to verify the quality of overlap determination for the 1979 survey.

11. Conclusions and Recommendations

A great deal of information was gained from the 1978 New Brunswick test. Generally speaking, the test ran smoothly and results are encouraging. It appears that multiple frame sampling is a viable technique for collecting data and producing estimates in a province such as New Brunswick. It has already been decided that use of multiple frame will expand to all three Maritime provinces for the 1979 survey. It will be a year for further testing of multiple frame since once again weighted estimates will be produced in addition to multiple frame estimates.

In light of the analysis outlined in this paper, the following recommendations were made and adopted for the 1979 survey.

1. Since data is to be collected for overlap area farms in 1979, Hartley estimates as well as screening estimates will be produced. Although for most items the gain in efficiency when using the Hartley estimates will be slight, if the data is available the additional computations necessary should be made. This will prove especially worthwhile for items for which the list estimate is poor for it is with these items that the Hartley estimate shows the greatest improvement.

2. The medium sized strata will be dropped from the list frame. List estimates for these strata are not as good as for the larger strata. It has been shown that a large number of sample farms selected in these strata have changed strata by survey time, thus reducing the effectiveness of stratification. As well, small to medium sized

list frame farms with limited data available may be harder to match with area sample farms. Thus, the crucial step of overlap determination could be endangered by including these farms in the list frame.

3. At the stratification stage, pigs are to be given equal priority with cattle when defining list strata. There is evidence from 1978 data that giving cattle priority has caused problems with list estimates for pigs. Equal priority will be given to cattle and pigs through use of multiple (or deep) stratification techniques.

4. Interviewers will attempt to determine overlap between the area sample and the list frame in the field in 1979. They will carry with them a copy of the list frame and, after every interview with a segment farm, will record whether or not this farm is on the list frame. It is hoped that interviewers will be able to make this determination more accurately than head office personnel were able to do in 1978. For 1979, however, the entire area sample/list frame match will be checked at head office to evaluate just how well interviewers were able to perform this step.

   In addition, a question has been added to the 1979 questionnaire which will help the interviewer determine whether the area sample farm is likely to be found on the list frame. The question is *"Was this farm operating under the present name at June 1, 1976?"* If the answer to this is no, the farm is automatically a non-overlap farm; if yes, the farm should have been included on the census list (but not necessarily the list frame) and the list frame would therefore be carefully checked for a match.

In conclusion, it is felt that multiple frame testing has certainly been worthwhile. We are continuing the testing in 1979 and if the changes to the 1979 survey (as listed above) are successful, we will be able to significantly improve our estimates in future surveys.

APPENDIX 1: Multiple Frame Formulae

a) Notation

$\hat{Y}_L$ = estimate for list frame population from list sample.

$\hat{Y}_A$ = area sample estimate for entire population.

$\hat{Y}_{OL}$ = area sample estimate of list frame population (overlap domain)

$\hat{Y}_{NOL}$ = area sample estimate of population not covered by list frame (non-overlap domain).

p = weight given to list frame estimate (for Hartley estimate).

q = weight given to area frame estimate of list frame population.

$p+q=1$

b) Weighted Estimate

$$\hat{Y}_A = \hat{Y}_{OL} + \hat{Y}_{NOL}$$

$$\mathrm{Var}(\hat{Y}_A) = \mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Var}(\hat{Y}_{NOL}) + 2\,\mathrm{Cov}(\hat{Y}_{OL}, \hat{Y}_{NOL})$$

c) Multiple Frame Screening Estimate

$$\hat{Y}_S = \hat{Y}_L + \hat{Y}_{NOL}$$

$$\mathrm{Var}(\hat{Y}_S) = \mathrm{Var}(\hat{Y}_L) + \mathrm{Var}(\hat{Y}_{NOL})$$

d) Hartley Multiple Frame Estimate

$$\hat{Y}_H = \hat{Y}_{NOL} + q\,\hat{Y}_{OL} + p\,\hat{Y}_L \qquad p+q=1$$

The area (weighted) estimate is a special case of the Hartley estimate when $p=0$ and $q=1$.

The screening estimate is a special case of the Hartley estimate when $p=1$ and $q=0$.

The Hartley estimate may be rewritten

$$\hat{Y}_H = \hat{Y}_{NOL} + (1-p)\,\hat{Y}_{OL} + p\,\hat{Y}_L$$
$$= \hat{Y}_A + p(\hat{Y}_L - \hat{Y}_{OL}).$$

The variance of this estimate is

$$\mathrm{Var}(\hat{Y}_H) = \mathrm{Var}(\hat{Y}_{NOL}) + q^2\mathrm{Var}(\hat{Y}_{OL}) + p^2\mathrm{Var}(\hat{Y}_L) + 2\,q\,\mathrm{Cov}(\hat{Y}_{NOL}, \hat{Y}_{OL})$$
$$= [\mathrm{Var}(\hat{Y}_{NOL}) + \mathrm{Var}(\hat{Y}_{OL}) + 2\,\mathrm{Cov}(\hat{Y}_{NOL}, \hat{Y}_{OL})]$$
$$- 2\,p[\mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Cov}(\hat{Y}_{OL}, \hat{Y}_{NOL})] + p^2[\mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Var}(\hat{Y}_L)]$$
$$= \mathrm{Var}(\hat{Y}_A) - 2\,p\,[\mathrm{Cov}(\hat{Y}_A, \hat{Y}_{OL})] + p^2[\mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Var}(\hat{Y}_L)]$$

The optimum value of p is now determined.

$$\frac{\delta\ \mathrm{Var}(\hat{Y}_H)}{\delta\ p} = -\ 2\ \mathrm{Cov}(\hat{Y}_A, \hat{Y}_{OL}) + 2\ p[\mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Var}(\hat{Y}_L)] = 0$$

$$\therefore\ p_{opt} = \frac{\mathrm{Cov}(\hat{Y}_A, \hat{Y}_{OL})}{[\mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Var}(\hat{Y}_L)]}$$

Using this value of p the variance of the Hartley estimate becomes:

$$\mathrm{Var}(\hat{Y}_H) = \mathrm{Var}(\hat{Y}_A) - 2\,p_{opt}\,[\mathrm{Cov}(\hat{Y}_A, \hat{Y}_{OL})] + p^2_{opt}[\mathrm{Var}(\hat{Y}_L) + \mathrm{Var}(\hat{Y}_{OL})]$$
$$= \mathrm{Var}(\hat{Y}_A) - p^2_{opt}[\mathrm{Var}(\hat{Y}_{OL}) + \mathrm{Var}(\hat{Y}_L)].$$