United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

NASS Research Report
Number SSB-88-07

December 1988

Modifying the Weighted Estimator To Eliminate Screening Interviews In Residential Areas

Raymond R. Bosecker
Michael S. Clark

MODIFYING THE WEIGHTED ESTIMATOR TO ELIMINATE SCREENING INTERVIEWS IN RESIDENTIAL AREAS. By Raymond R. Bosecker and Michael S. Clark, Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture. Washington, D.C. 20250, December 1988. Staff Report No. SSB-88-07.

ABSTRACT

One estimator available to NASS for the area frame component of the Agricultural Surveys is the "weighted" estimator. A weight is calculated for each farm by dividing the total tract acres by the entire farm acres. In residential areas that fall in a sample segment, it is necessary to go from door to door looking for farm operators. When farm operators are found, the procedure defined above is used to calculate a weight. Since the tract acreage associated with farm operators living in residential segments is usually quite small (less than 1/2 acre), the weight will be quite small. This study simulated a modification of the weight to eliminate these small fractions. Looking for farm operator residents in densely populated areas would then be unnecessary for the weighted estimator. The results showed there were only minor differences between the operational weighted estimate and the modified weighted estimate. In addition, the modification would result in a savings of time and money.

CONTENTS

SUMMARY

This report describes a proposed modification to the weighted estimator used by NASS and simulates the likely impact on survey indications. The data used for this study were collected during the 1987 June Agricultural Survey. No special questions were asked during the survey for use in this study. The idea was to investigate a procedure which would make it unnecessary to screen for farm operators who reside in segments having large numbers of residences. In residential areas, lots are usually less than 1/2 acre. Therefore, the basic modification was to remove from the weight up to 1/2 acre for the residence (house and yard). The use of 1/2 acre would generally ensure that the weight for farm operators living in residential areas would legitimately equal zero. Interviews would therefore become unnecessary for those operators.

The results show that the weighted estimates and coefficients of variation from the use of the modified weight are not significantly different from the current weighted statistics. Thus, the modified weighted estimate could be used in the current estimating program without any loss of quality.

The weighted approach is a form of multiplicity or network sampling especially suited to the measurement of rare populations. The number of farm operators among the general population, particularly in residential areas, certainly qualifies as rare. The proposed modification of the weighted estimator, together with the discontinuance of the estimate based on resident farm operators (RFO), would make it unnecessary to contact anyone in a sample segment having 1/2 acre or less devoted solely to the residence. The RFO estimator has already been demonstrated in other studies to be less precise, inefficient, and biased downward [2, 6].

Further evaluation is recommended during the 1989 June Agricultural Survey. For this study, the estimated fractional acreage in the farm residence should be recorded on the questionnaire by the enumerator. This would enable the use of a more accurate adjustment than using a fixed 1/2 acre. The acreage in the farm residence will then vary from farm to farm.

The elimination of door-to-door canvassing in areas with high density housing will save both time and money. The survey estimate provided by the proposed modification is mathematically unbiased and should be an immediate replacement for the current weighted procedure. If the results from the 1989 study again show only insignificant differences and the RFO estimator can be dropped, then the use of the modified weight should be adopted.

MODIFYING THE WEIGHTED ESTIMATOR TO ELIMINATE SCREENING INTERVIEWS IN RESIDENTIAL AREAS

Raymond R. Bosecker
Michael S. Clark

INTRODUCTION

The weighted estimator currently used in the Agricultural Survey (AS) program uses a weight based on tract acreage divided by entire farm acreage. This makes it necessary to locate any land belonging to a farm operator within the segment. In residential areas, the task of searching for farm operators becomes very difficult and time consuming--and quite often finds no farm operators. If a farm operator is found, the weight is extremely small because the tract size is generally less than 1/2 an acre. Since locating farm operators in residential areas usually involves a considerable amount of time, the task is normally done before the actual collection of data in the June Agricultural Survey (JAS). This task is referred to as prescreening.

The purpose of this study was to investigate a modification to the weighted estimator that would ease prescreening for farm operators in densely populated areas. Since the only land within segments in most residential areas is for houses and yards, the exclusion of land for the residence when calculating the weight eliminates the need for "knocking on doors". In other words, we could look for farm operators through land operated for agricultural purposes rather than where they live. This is a form of multiplicity sampling utilized by most statistical organizations to measure rare occurrences [3]. The number of farmer occupied residences is indeed rare, especially in high density residential areas.

The actual cost of the current prescreening procedure is not available. An estimate of this cost was made, though, using the enumerator payroll summaries available from the Survey Administration Section. The cost was estimated at $100,000 but is probably on the low side since it was based on the assumption that screening costs were incurred only prior to the state training schools. Another estimate of $150,000 was previously made by the Survey Management Branch. The modification to the weighted estimate would require some prescreening costs but they would be much lower. A guess is that it would be no more than 25% of the current cost. The resulting savings could then range from $75,000 to $115,000. Using the proposed alternative weighted estimator in place of special, intensive screening of high density housing areas would save about three times the above estimated amount in those years associated with the Census of Agriculture.

As with any given procedure there are both advantages and disadvantages associated with the proposed change. The advantages of the modification to the weighted estimator include:

1. Eliminates the need to interview residents in residential areas to locate farm operators, i.e., resource savings.

2. Provides indications nearly identical with the current weighted indications so they may be used in the current estimating program.

3. Takes full advantage of multiplicity sampling since the weighted indication is based on land where farming can take place rather than where the operator lives.

4. Minimizes extremely small weights since tracts composed solely of a house and yard are removed from computations.

The disadvantages of the modification include:

1. Enumerator must provide the fractional acreage associated with the residence (unless it can be shown that a simulated adjustment as in this report will give equivalent results).

2. Has a very small reporting bias if the adjustment for acreage in the residence is incorrect.

3. May miss selected specialty agricultural activities where: 1) a backyard operation has no visible signs of agriculture but is grossing in excess of $1,000; worm farms for example, or 2) the operation has all land in PIGA except the residence.

4. Does not require identification of RFO's who, under current procedures, comprise the estimator used for the Farm Costs and Returns Survey (FCRS) and the Farm Labor Surveys.

This study was aimed at eliminating the need to locate farm operators living within high density housing areas where there is no agriculture. Interviewing these operators during the June Agricultural Survey would not be necessary because their weight is zero. The elimination of an interview (zero weight) could actually occur in any area frame stratum when only the farm operator's residence is within a segment. A slight decrease in the number of needed interviews was observed in this study.

Only totals for the 10 states combined are shown in the table below since the results are very similar in the individual states. Some additional savings would accrue from the small decrease in the number of interviews required.

	---NUMBER OF INTERVIEWS---		% OF
STRATUM 1/	WEIGHTED	MODIFIED	WEIGHTED
CULTIVATED	13,339	13,307	99.8
RESIDENTIAL	598	530	88.6
NON-AGRICULTURAL	12	12	100.0

1/ The cultivated stratum includes all of the strata designated with a percent cultivated which includes the range strata. The residential stratum includes all of the strata designated as having more than 20 dwellings per square mile.

DESIGN

Data from the 1987 JAS for the 10 states using the weighted estimator (Georgia, Illinois, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Carolina, and Ohio) were used for this study. Estimates were calculated for hogs, cattle, and number of farms. The estimates were pure area estimates, i.e. data for E.O.'s were not taken out prior to summarization.

Some rules for making adjustments to the weights had to be established for this simulation exercise. The general idea was to exclude the farm operator's residence from the calculation of the weight. It was determined before the analysis began that up to 1/2 acre would be used for the residence. This amount should be enough to create zero weights for farm operators residing in residential segments.

The rules used to simulate adjustments to the weights were as follows:

Rule 1: For RFO operations having tract (and therefore farm) acreages above 1/2 acre, a fixed amount of 1/2 acre was subtracted from both the tract and entire farm components of the weight, i.e., (weight = tract - 1/2 acre)/(farm - 1/2 acre). This accounted for 29.5 percent of the mid-year survey agricultural tracts studied.

Rule 2: For non-RFO reports where subtracting 1/2 acre from farm acreage still equaled or exceeded the recorded tract acreage, the new weight was computed with 1/2 acre for the residence removed from the entire farm acres in the

denominator of the ratio, i.e., (weight = tract acres)/(farm - 1/2 acre). This rule applied to 67.7 percent of the agricultural tracts included in this study.

Rule 3: For non-RFO tracts having entire farm acreage within 1/2 acre of the tract acreage, the weight was set to 1.0. The amount of land ascribed to the residence was therefore set at the difference between the tract and farm acreage up to 1/2 acre. This accounted for 2.1 percent of the survey tracts.

The first three rules accounted for more than 99 percent of all survey tracts. Additional rules were established to account for all other possibilities in this simulation study. The remaining situations that could occur involve RFO tracts where tract acreage is 1/2 acre or less. The action taken would depend upon other available information as follows:

Rule 4: If RFO, and 1/2 acre $\geq$ tract > farmstead acres, the acres in the field designated as farmstead were subtracted from both the tract and entire farm. No cases were found for application of this rule.

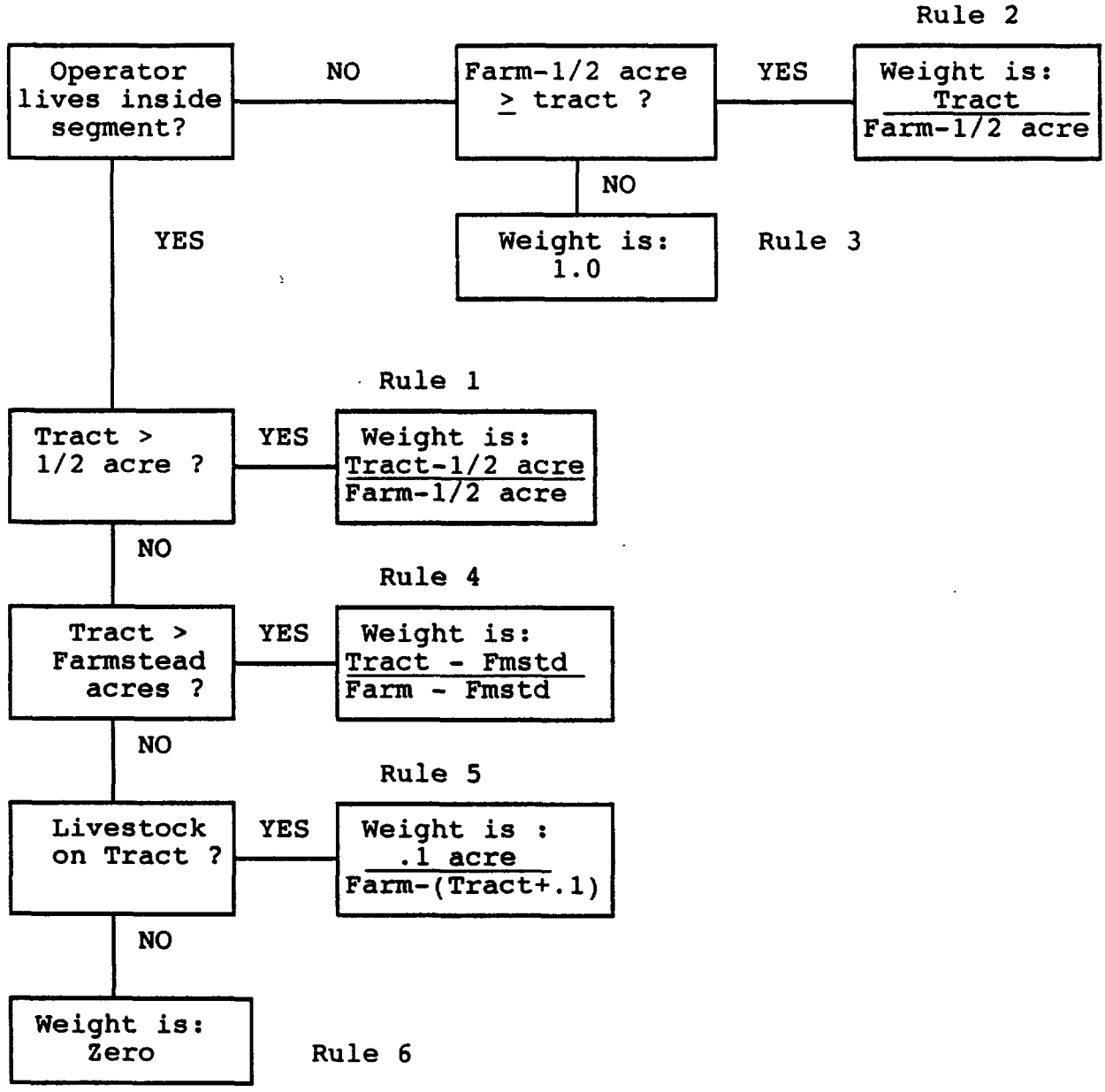
Rule 5: If RFO, and 1/2 acre $\geq$ tract = farmstead acres, then the weight adjustment depended upon the presence of hogs or cattle on the small tract. In order to avoid the loss of any agricultural data under the modified weight, some land (.1 acre) was retained in the tract to preserve a positive weight for any land having an agricultural use. No cases for this rule were found.

Rule 6: If RFO, and 1/2 acre $\geq$ tract = farmstead acres (as in b), but no livestock were present, the weight was assumed to be a legitimate zero. Any crop acres would have been recorded in a separate field other than farmstead. The remaining tracts not included in the first three rules (less than 1 percent) fell under this rule.

These rules were for simulation purposes only in order to provide a reasonable approximation for the fractional acreage decreases that might be found when enumerators supply the necessary data to modify the weights.

A schematic representation of the simulation rules is given in Figure 1 below.

Figure 1: Decision diagram for simulating modified weighted estimator.



This, or a similar automated adjustment, might serve as a way to model the modified weight so enumerator input would not be required. However, data are needed on the appropriate size of the adjustment (land used for residential purposes) before an appropriate model can be applied.

ANALYSIS

Estimates and coefficients of variation (C.V.'s) were calculated using the modified weight and the current weight. The resulting estimates and C.V.'s were compared. In addition, a paired t-test was calculated. The paired t-test was based on 95% confidence limits for the difference of the estimates. A two-tailed hypothesis was tested since the modified estimate could be either higher or lower than the current estimate.

RESULTS

The estimates and C.V.'s for the current and modified approaches are shown in Tables 1, 2, and 3 for hogs, cattle, and number of farms respectively. Paired t-test results with significant differences at the 5% level are indicated by a "*" in the tables. The C.V.'s were virtually unchanged.

The biggest difference in the estimates was just over 1/2% for cattle in Indiana. This amounted to only 7,385 which is negligible when compared to the current estimate of 1,411,427. Only four other estimates were different by more than 1/10%.

TABLE 1 - ESTIMATES AND C.V.'S FOR HOGS

	------ESTIMATE------		% OF	--------CV--------	
STATE	WEIGHTED	MODIFIED	WEIGHTED	WEIGHTED	MODIFIED
GA-13	1,045,666	1,045,138	99.95	31.9	31.9
IL-17	5,661,666	5,661,940	100.00	13.3	13.3
IN-18	3,988,517	3,998,071	100.24	15.0	15.0
IA-19	14,559,474	14,573,834	100.10	8.2	8.2
KS-20	1,229,246	1,229,372	100.01	16.5	16.5
MN-27	4,156,392	4,156,593	100.00	10.6	10.6
MO-29	2,469,887	2,471,457	100.06	11.9	11.9
NE-31	3,746,493	3,754,157	100.20	17.1	17.0
NC-37	2,719,645	2,721,878	100.08	37.1	37.1
OH-39	1,738,302	1,743,070	100.27	13.3	13.3
10-ST	41,315,288	41,355,510	100.10 *	5.0	5.0

TABLE 2 - ESTIMATES AND C.V.'S FOR CATTLE

STATE	ESTIMATE WEIGHTED	ESTIMATE MODIFIED	% OF WEIGHTED	CV WEIGHTED	CV MODIFIED
GA-13	1,583,305	1,583,860	100.04	7.5	7.5
IL-17	2,220,213	2,220,390	100.01	12.4	12.4
IN-18	1,411,427	1,418,812	100.52	7.8	7.9
IA-19	4,397,474	4,397,774	100.01	5.4	5.4
KS-20	7,429,513	7,430,570	100.01 *	25.5	25.5
MN-27	2,851,239	2,853,498	100.08	6.4	6.4
MO-29	5,300,568	5,300,253	99.99	4.7	4.7
NE-31	6,544,256	6,542,661	99.98	6.5	6.5
NC-37	1,080,111	1,080,852	100.07 *	9.3	9.3
OH-39	1,915,663	1,916,008	100.02	10.1	10.1
10-ST	34,733,769	34,744,678	100.03	5.8	5.8

TABLE 3 - ESTIMATES AND C.V.'S FOR NUMBER OF FARMS

STATE	ESTIMATE WEIGHTED	ESTIMATE MODIFIED	% OF WEIGHTED	CV WEIGHTED	CV MODIFIED
GA-13	50,797	50,782	99.97	5.5	5.6
IL-17	86,453	86,529	100.09 *	3.8	3.8
IN-18	70,604	70,701	100.14 *	4.6	4.6
IA-19	111,693	111,776	100.07	2.9	2.9
KS-20	75,345	75,366	100.03	4.2	4.2
MN-27	92,119	92,190	100.08	3.5	3.5
MO-29	118,986	118,877	99.91	4.5	4.5
NE-31	62,069	62,111	100.07	3.5	3.5
NC-37	75,328	75,376	100.06	5.5	5.6
OH-39	87,359	87,376	100.02	4.5	4.5
10-ST	830,753	831,084	100.04 *	1.4	1.4

None of the tests for state hog estimates were significant at the 5% level. However, the combined ten state estimate was significant. It should be pointed out that the confidence limits were very small due to the large sample sizes and the precision of a paired t-test. Therefore, the tests could detect extremely small differences. The interval for the combined ten state test for hogs was less than 2/10 of one percentage point.

The estimates for cattle showed significant differences for Kansas and North Carolina at the 5% level. The intervals here were even smaller than that previously mentioned for hogs. The combined ten state estimate was not significant for cattle.

The estimates for number of farms were the most interesting. The tests for Illinois, Indiana, and the ten states combined were all significant at the 5% level. In addition, Illinois and Indiana also tested significant at the 1% level. But again, the intervals were very small: only 1/10% for Illinois and less than 2/10% for Indiana. Thus, although the differences were statistically significant, the differences were too small to be of practical significance.

A review of the strata level estimates showed little change from the operational weighted estimates. This is as expected since there are farming operations located within residential strata other than the operator's residence. In addition, the residential strata have very little impact on the overall estimates. Thus, even major changes in the residential strata would not significantly affect the state level estimates.

The estimates shown in Tables 1, 2, and 3 are almost always higher for the modified than for the current weight. This indicates the use of a fixed 1/2 acre adjustment produced a small upward bias evidenced by the consistency of the direction of change. For nonresident farm operators, the adjustment to the weight makes it slightly larger resulting in higher estimates. For resident farm operators, the weight becomes smaller and may go to zero. Around 70 percent of farm operators were nonresidents. The bias from using a constant 1/2 acre would have to be extremely small since the bias contribution was only a part of the small difference in the indications. Using the actual size of the residence to adjust the weight, probably less than .5 on average, should show even smaller differences between the weighted estimates, both positive and negative.

DISCUSSION

The Agricultural Survey Manual for June 1988 says that "Segments...with "built up" areas having a density of 2 or more houses per acre should be thoroughly screened for farm operators...". This statement recognizes that in residential segments the residence is generally less than or equal to 1/2 acre. The justification for using 1/2 acre in this study for making adjustments to the weights comes from the above statement.

The purpose of this study was to find some way to avoid costly prescreening in residential segments by making the weight zero for farm operators residing in those segments. The procedure must apply to all operations in all segments uniformly in order to be statistically sound. Using <u>up to</u> 1/2 acre for the residence seemed a promising way to proceed.

The problem with a fixed adjustment is that when the residence is actually less than the adjustment, the weights were slightly biased because they no longer added to 1. However, this bias proved extremely small in this study since the estimates using the modified weight did not differ much from the estimates using the current weight. The bias was a component of the small total difference between the estimates.

The alternative to a fixed adjustment is a variable adjustment for the residence. This could be calculated in two ways:

(a) The farm operator could be asked the acreage in his residence

or

(b) An estimate of the acreage in the residence could be made by the enumerator.

We believe the first option would not make sense to the farmer and should be avoided. Making an estimate for the residence would be more reasonable since the estimates would be in the range of 1/10 to 5/10 of an acre. The enumerator should be able to provide a reasonable approximation. There is no guarantee that the enumerator will always see the farm operator's residence since some interviews are conducted in the field where the farm operator is working. However, his best estimate based on experience or an imputed value based on other reports could not be off by very much. The enumerator would have the potential to differentiate between lots up to 5/10 acre.

One final comment concerns the Farm Costs and Returns Survey (FCRS). The only estimate currently produced from the FCRS is the RFO estimate. The modified weighted estimate studied here would eliminate the need to identify the RFO's except as currently needed for the FCRS. There are two possible alternatives. The first would be to screen residential segments that are in the sample for the FCRS. This would be less work than is currently done for the JAS since only 40% of the segments are in the FCRS sample. The second alternative would be to use a weighted estimate for the FCRS. Research has been conducted on a commodity weighted estimate [8], or the weighted estimator proposed in this study might be used. Although a weighted approach has not yet been adopted, it should be given serious consideration as a viable alternative to the less precise RFO estimator.

RECOMMENDATIONS

The modified weighted estimate has been shown to be practically indistinguishable in level and precision from the current weighted estimate. It would be beneficial to adopt such an

approach since it would save time and money in screening built-up residential areas.

The recommendation to adopt the modified weighted estimator is dependent upon two conditions:

1) a willingness to move away from the RFO estimator, and
2) a demonstration of the feasibility to modify the current weighted estimator.

The current use of the RFO estimator for the Agricultural Surveys (AS) is limited to indications for hogs and number of farms in the June quarter only. Proposals to drop the hog RFO estimate have already come from the Crop Reporting Board Policy and Procedures Working Group and the Livestock, Dairy and Poultry Branch.

The RFO indication for number of farms has been used since the beginning of area frame sampling. These survey estimates have long been recognized to understate the actual level of farm numbers (downward bias). This was demonstrated in 1987 when intensive screening of high density housing areas revealed additional farm operators who had been missed under traditional screening procedures. The intensive procedure was a very expensive and arduous task. Even periodic measurements of the number of missed farms due to faulty screening would add substantially to the long term costs associated with the resident farm operator concept.

Until the 1988 June survey there was no nationwide alternative to the RFO estimate. That was the first survey in which the necessary data were collected in all states to generate the weighted estimate. Continuation of a weighted estimator, having a lower coefficient of variation, should provide a more stable indication of year to year change in number of farms than the RFO procedure has been able to do in the past.

Area frame stratification and sampling have not been designed for the RFO estimator since NASS began moving away from the original Master Sample of Agriculture in the early 1960's. It is recommended that NASS finally abandon the RFO approach.

The 1989 June area questionnaire should contain a question for the enumerator to estimate the acreage devoted to the house and yard. This would provide a feasibility evaluation and another comparison of the modified weighted estimate with the current weighted estimate. Enumerator estimates for the acreage in the residence may reduce the slight bias noted in this study. In addition, average sizes for residences can be obtained by state and stratum for possible future modeling of the weight adjustments.

The land to be excluded for the residence should be defined as "nonagricultural land devoted to residential purposes". This would include the house and yard, detached garages, gardens, and sheds for lawn and garden equipment. Any land, including the yard, with livestock or grain storage would then fail the test of being nonagricultural and would therefore not be excluded from the weight. The minimum acreage for the residence would be .1 acre.

If the estimates again show very little difference and the RFO concept is dropped, then the June area survey should be conducted using fewer resources to screen high density housing areas by adopting the proposed modification to the weighted estimator.

REFERENCES

[1] Cochran, William G. Sampling Techniques, 3rd Edition, New York: John Wiley and Sons, Inc., 1977.

[2] Crop Reporting Board Policy and Procedures Working Group. Crop Reporting Board Standards. U.S. Department of Agriculture, National Agricultural Statistics Service, July 1985.

[3] Czaja, Ronald F., Cecelia B. Snowden, and Robert J. Casady. Reporting Bias and Sampling Errors in a Survey of a Rare Population Using Multiplicity Counting Rules. Journal of the American Statistical Association, Vol. 81, June 1986, pp. 411-419.

[4] National Agricultural Statistics Service. Agricultural Surveys Interviewer's Manual. U.S. Department of Agriculture, NASS, 1988.

[5] National Agricultural Statistics Service. Enumerative and Agricultural Surveys Supervising and Editing Manual. U.S. Department of Agriculture, NASS, 1988.

[6] Nealon, John Patrick. Review of the Multiple and Area Frame Estimators. U.S. Department of Agriculture, National Agricultural Statistics Service, Staff Report No. 80, March 1984.

[7] Snedecor, George W. and William G. Cochran. Statistical Methods, 7th Edition, Ames, Iowa: The Iowa State University Press, 1980.

[8] McClung, Gretchen. A Commodity Weighted Estimator. U.S. Department of Agriculture, National Agricultural Statistics Service, Staff Report No. SRB-88-02, January 1988.