United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Application
Division

NASS Staff Report
Number SSB-90-05

July 1990

MODIFIED AGRICU
WEIGHTED ESTIMA

Robert G. Pontius, Jr.

MODIFIED AGRICULTURAL WEIGHTED ESTIMATORS. By Robert G. Pontius, Jr., Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture. Washington, D.C. 20250, July 1990. Staff Report No. SSB-90-05.

ABSTRACT

Two alternatives have been proposed for the weighted estimator currently used in the NASS area frame. Both alternatives modify the weight by excluding up to one half acre for the house and yard from the farm's acreage. This would eliminate the need to screen for farm operators in residential areas where the tracts are one half acre or less. One of the modifications requires an additional entry in the survey questionnaire, the other does not. This report uses 1989 June Agricultural Survey data to show that only negligible differences exist between the estimates derived from the current estimator vis-à-vis the modified estimators. To improve operational efficiency, NASS should implement the modified estimator which does not require additional data.

CONTENTS

ACKNOWLEDGEMENTS

The author is indebted to Mike Clark for supplying computer programs and answering questions. The author also thanks Ron Bosecker for his comments and encouragement.

SUMMARY

In an effort to eliminate screening for farm operators in densely populated segments, Bosecker and Clark proposed a modification to the current weighted estimator [1]. The weight currently used to prorate entire farm data to the tract is (tract acres)/(entire farm acres). The modified estimator would exclude up to one half acre for non-agricultural land devoted to residential purposes. For resident agricultural tracts, the residential area would be subtracted from the weight's numerator and denominator; for non-resident agricultural tracts, the residential area would be subtracted just from the weight's denominator. Since the modified weight would be zero for small tracts consisting of only a house and yard, screening for farm operators in residential areas would be unnecessary.

This report investigates two ways to implement the modification. The first is called the "observed" estimator, because it requires the enumerator to approximate the area of the residence to the nearest 1/10th of an acre. The second is called the "modeled" estimator because it models the weight by assigning 1/2 acre for residences.

The purpose of this study is to compare the estimates derived from the current estimator to those from the two modified estimators. This report shows that only negligible differences exist. The analysis uses 1989 June Agricultural Survey data for 16 States. For each State, the items considered are the number of farms estimate and the nonoverlap domain estimates for cattle, hogs, and individual on-farm grain stocks. The number of estimates evaluated across the 16 States is 125. For each estimate, the current weighted expansion is compared to the two corresponding modified weighted expansions. T-tests, which pair segments weighted under the current and observed estimators, show significant differences between the estimates in only 10 of the 125 pairs. In these ten cases, the magnitudes of the differences are of no practical significance. Sign tests for each variable show no tendency across States for the observed estimates to be different from the current estimates. For the modeled estimator, paired t-tests show only 9 of the 125 estimate pairs to be statistically different. Again, none of the significantly different pairs are of practical significance. Sign tests show a tendency for the modeled estimates to be larger than the current estimates only for oats stocks.

A change from the current estimator to a modified estimator would save time and money, with negligible change to the estimates. The modeled estimate would not require additional data collection. Both alternatives would exclude information needed for the resident farm operator (RFO) estimate. It is recommended that the modeled weighted estimator replace the current weighted procedure. The RFO estimator should be phased out as soon as possible to free NASS from intensive screening in residential areas.

MODIFIED AGRICULTURAL WEIGHTED ESTIMATORS

Robert G. Pontius, Jr.

INTRODUCTION

An important estimator used by the National Agricultural Statistics Service (NASS) in multiple frame surveys is the area frame weighted segment estimator. The weight prorates farm data to the sample segment. Each farm with land in a segment is assigned a weight. The current weight is the farm area in the segment (i.e. tract area) divided by the total farm area. Weighted farm data is then summed for all tracts to obtain a segment total. Definitions of terms, including the weighted estimator, are given in Appendix 1.

Before a survey is conducted, segments which have many residences are screened for farm operators. In residential areas, many tracts are less than half an acre and consist of only a house and yard. These small tracts rarely have agriculture associated with them, and even when they do, the tract's weight is usually quite small. Much time and money is spent knocking on doors to locate these small tract residences. There is also substantial downward bias in the number of resident farm operators found in the screening process [5].

In 1988, Bosecker and Clark investigated a modified estimator which subtracts up to one half acre for the house and yard from the farm acreage [1]. Under the modified estimator, a tract consisting of one half acre or less of house and yard has a weight of zero, thus there is no need to screen for farmers in segments consisting of such tracts. Bosecker and Clark concluded that the modification to the weight had a negligible effect on estimate levels and precision, but recommended that further research should be done another year. This paper presents that research.

DESIGN

This report analyzes two different ways to modify the weight of the current estimator. The first alternative is called the "observed" estimator because it requires the enumerator to approximate the area of the house and yard to 1/10th of an acre, with a maximum of one half acre. The second is referred to as the "modeled" estimator because it models the weight by assigning 1/2 acre (or less where required) for the house and yard.

For the 1989 June Agricultural Survey, 16 States were selected to compare the current estimator to its alternatives. The States are Alabama, Arizona, California, Colorado, Idaho, Indiana, Iowa, Kentucky, Minnesota, North Carolina, North Dakota, Oklahoma, Oregon, Pennsylvania, Texas, and Wisconsin. Variables considered are grain stocks, number of farms, cattle, and hogs. For grain stocks and livestock, only the nonoverlap (NOL) component of the

estimates was considered. For tracts in which the grain stocks questions were refused, imputed data were used.

No State generated estimates for all of the different kinds of grain stocks. Among the 16 States, the total number of estimates ranged from six to eleven, resulting in 125 estimates to be considered.

Observed Weights

To calculate the observed weights, enumerators were asked to approximate the area of non-agricultural land devoted to residential purposes. A maximum area of one half acre was allowed because the June 1988 Agricultural Survey Interviewer's Manual calls for screening in areas with two or more houses per acre [6]. On the questionnaire, enumerators could enter one of the numbers listed: 0.1, 0.2, 0.3, 0.4, or 0.5 acres.

In some cases enumerators misunderstood the instructions and wrote in areas larger than one half acre. Almost 1% of the 14,416 agricultural tracts had residential area approximations greater than 1/2 acre. All States except three had some approximations between 1/2 and 5 acres. Only one tract had an approximation greater than 5 acres.

Editing of the residence acreage data was required. Residential areas recorded as more than 1/2 acre were set to 1/2 acre. For tracts which had missing residence acreage values, the median value from completed questionnaires of 0.4 acres was imputed. For resident tracts, if the acreage recorded for the farmstead (house, yard, outbuildings, etc.) was less than the acreage for the residence alone, then the farmstead acres was used for the residence. For non-resident tracts, if the total farm land was less than the sum of the tract and residence, then the weight was set to 1 (i.e. farm acres minus residence area set equal to tract acres).

To obtain the observed weights, the following calculations were performed:

$$\text{for residents,} \quad \frac{\text{total tract area} - \text{residence area}}{\text{total farm area} - \text{residence area}}$$

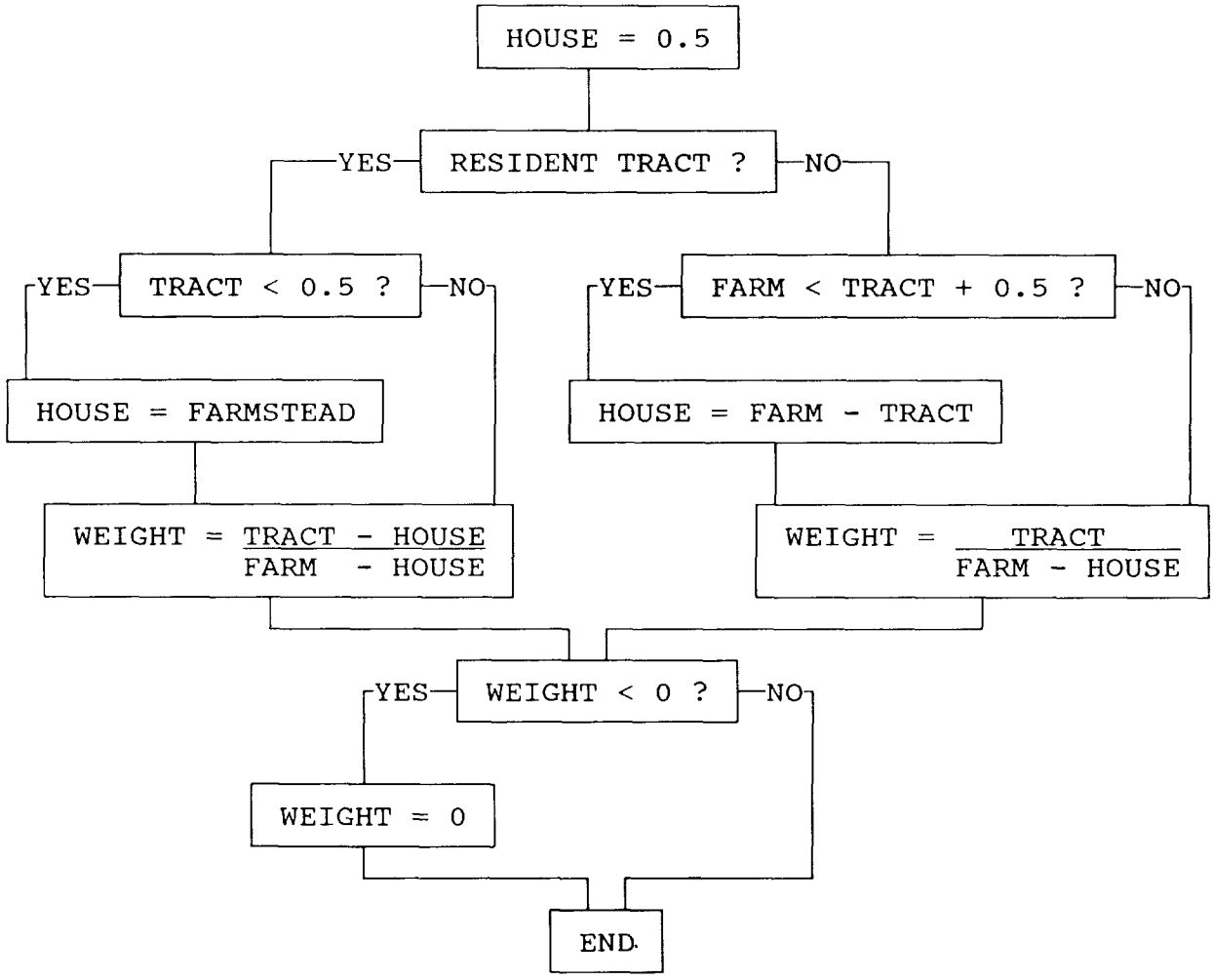
$$\text{for non-residents,} \quad \frac{\text{total tract area}}{\text{total farm area} - \text{residence area}}$$

Modeled Weights

The second alternative, using the modeled weights, does not require the area of the house and yard to be recorded. The modeled estimator assumes 1/2 acre for all residences, except where it is known that the farmstead is less than 1/2 acre. The formulas for the modeled weights are the same as those given above. The exact flow diagram is given in Table 1. In the table, "HOUSE" means the

house and yard acreage, "FARMSTEAD" means farmstead acreage, "TRACT" means tract acreage, and "FARM" means total farm acreage.

TABLE 1 - DECISION DIAGRAM TO CALCULATE THE MODELED WEIGHT



ANALYSIS

Identical analyses are used to compare the current estimates to the observed estimates and the current estimates to the modeled estimates. Two tailed t-tests, which pair corresponding segments, compare each current estimate with its corresponding modified estimate. The formula for the effective number of degrees of freedom for the stratified design was developed by Satterthwaite, obtained via Cochran [2].

Furthermore, nonparametric sign tests were used to detect any difference between the current estimates and their alternatives across the States. For each variable, the number of States for which the current estimate was less than the modified estimate was

compared to the number of States for which the current estimate was greater than the modified estimate. Details of the test are described by Hollander and Wolfe [4].

RESULTS

A salient feature of the analyses for both alternative estimators is the lack of practical significance in the differences from the current estimates. Appendix 2 gives estimates and 16 State totals for NOL CATTLE, NOL HOGS, NOL CORN STOCKS, and NUMBER OF FARMS. Appendix 3 gives the coefficients of variation. For the 16 State totals, the current totals are within 0.1% of the modified totals with one exception. Deviations of 0.5 and 0.6 percent between the current and the modified totals for hogs is primarily the result of one large NOL hog operation with a small amount of land. The differences in the hog expansions caused by this outlier were not significant for either the individual State or the 16 State total.

The remainder of this section will examine the size and direction of estimate changes in the 16 States studied for each of the alternative weights. Throughout this report, percent change is defined as:

$$\% \text{ change} = \frac{(\text{modified estimate} - \text{current estimate}) * 100}{\text{current estimate}}$$

Observed Weights

The paired t-tests reveal that in only 10 of the 125 estimate pairs are the observed and current estimates statistically different, i.e. p-value < 0.05. This is close to what is expected under the hypothesis of no difference in any estimate pairs. Because there are 125 tests each with alpha = 0.05, approximately 6 estimate pairs are expected to show a difference when there are no differences. Furthermore, of the ten statistically different pairs, none are different by more than 0.3 percent (Table 2), and none have CV differences of more than 0.01 percentage points. This shows that, although the differences in a few estimates are statistically significant, this is to be expected and the differences are of no practical significance.

TABLE 2 - PERCENT CHANGE OF CURRENT ESTIMATES WHICH ARE STATISTICALLY DIFFERENT (P-VALUE < 0.05) FROM THEIR CORRESPONDING OBSERVED ESTIMATES

	State							
Estimate	CO	ID	IN	MN	NC	ND	OK	PA
NUMBER OF FARMS			0.26		0.25			0.12
SORGHUM STOCKS	0.04							
SOYBEAN STOCKS					0.11	0.04		
WHEAT STOCKS	0.05	0.04		0.08			0.04	

Table 3 contains the results of formal nonparametric sign tests on each of the variables. The current estimate is less than the observed estimate in approximately one-half of the 125 estimate pairs. This is evidence that the observed estimator does not systematically produce different results than the current estimator. The category entitled "WHEAT STOCKS" includes combinations of individual wheat types (i.e. durum, spring, or winter) for states with separate expansions by type. Although NUMBER OF FARMS and SOYBEAN STOCKS have p-values near 0.05, none of the tests are significant at the 5% level. This indicates there is no tendency toward higher or lower expansions between the observed and current estimates among the States.

TABLE 3 - NONPARAMETRIC SIGN TEST FOR DIFFERENCE BETWEEN CURRENT AND OBSERVED ESTIMATORS ACROSS STATES BY VARIABLE

Estimate	# of States	# of States for which Current < Observed	Sign Test P-value
NUMBER OF FARMS	16	12	0.08
CATTLE	16	8	1.00
HOGS	16	7	0.80
WHEAT STOCKS	14	10	0.18
BARLEY STOCKS	12	7	0.77
CORN STOCKS	15	9	0.61
OATS STOCKS	15	6	0.61
SOYBEAN STOCKS	11	9	0.07

Practical changes in the estimates are nonexistent. Only 6 of the 125 estimate pairs have more than a 1% change, and none of these changes are statistically significant. The maximum relative change among paired estimates of appreciable size was 2.3 percent. The relative change in the estimates was less than 0.1% in 82 of the 125 cases. Nearly all of the CV changes are less than 0.1 percentage points.

Modeled Weights

Comparing the modeled estimates to the current estimates, paired t-tests reveal that only 9 of the 125 estimate pairs are statistically different. Just as in the observed versus current study, to have a few significantly different pairs is little concern because even if no pairs were truly different, the tests are expected to show that approximately 6 pairs are different. Of the nine statistically different pairs, none are different by more than 0.4 percent (Table 4), and none have CV differences of more than 0.02 percentage points. This again shows that, although the differences in some estimates are statistically significant, these differences are of no practical significance.

TABLE 4 - PERCENT CHANGE OF CURRENT ESTIMATES WHICH ARE STATISTICALLY DIFFERENT (P-VALUE < 0.05) FROM THEIR CORRESPONDING MODELED ESTIMATES

	State						
Estimate	CO	ID	IN	MN	NC	ND	PA
NUMBER OF FARMS			0.33		0.27		0.14
SORGHUM STOCKS	0.05						
SOYBEAN STOCKS					0.12	0.04	
WHEAT STOCKS	0.08	0.04		0.10			

Table 5 contains the results of nonparametric sign tests on each of the variables. Just as with the observed estimator, the current estimate is less than the modeled estimate in about half of the 125 pairs. The difference in the Soybean Stocks variable is significant at the 10% level, but the only difference which is significant at the 5% level is in oats stocks. This indicates a tendency for the modeled estimator to give larger oats stock estimates than those given by the current estimator.

TABLE 5 - NONPARAMETRIC SIGN TEST FOR DIFFERENCE BETWEEN CURRENT AND MODELED ESTIMATORS ACROSS STATES BY VARIABLE

Estimate	# of States	# of States for which Current < Modeled	Sign Test P-value
NUMBER OF FARMS	16	11	0.21
CATTLE	16	9	0.80
HOGS	16	6	0.45
WHEAT STOCKS	14	10	0.18
BARLEY STOCKS	12	7	0.77
CORN STOCKS	15	9	0.61
OATS STOCKS	15	3	0.04
SOYBEAN STOCKS	11	9	0.07

Differences in the estimates give no need for practical concern. Only 10 of the 125 estimate pairs have more than a 1% change, and none of these changes are significant. The largest relative changes are due to the small values for grain stocks in Arizona. Outside of Arizona, all relative changes are below 3%. Relative change in the estimates is less than 0.1% in 77 of the 125 cases. Nearly all of the CV changes are less than 0.1. With the exception of some Arizona grain stocks, no CV change is more than 1 percentage point.

DISCUSSION

Two of the concerns noted by Bosecker and Clark are mentioned here. First, some specialty agricultural activities may be missed if enumerators were to pass over small residences. Backyard operations, such as worm farms, are difficult to detect, but can gross over $1,000. Other operations might have all land in PIGA except for the residence. However, the area frame sample is not

well suited to adequately measure rare operations with any estimator.

Second, both alternative estimators do not require resident farm operator (RFO) identification which is needed for the Farm Costs and Returns Survey (FCRS) and the Agricultural Labor Surveys. One solution is to screen for resident operators in the 40% of the June area segments which are in the FCRS. Another option is to use a different estimator for FCRS. A modified weighted estimator described in this report could be used.

Although the exact numbers are not available, the Survey Management Branch has estimated screening costs at $150,000. Bosecker and Clark used administrative records to estimate screening costs incurred prior to the State training schools to be $100,000. A modified estimator will not completely eliminate screening costs, but will probably reduce the costs by 75%. If this reduction is realized, the resulting savings could be from $75,000 to $112,000.

Since the modified estimators exclude collection of information needed to calculate RFO estimates, NASS should seriously consider abandoning RFO estimates. RFO estimates are now used as one indication for number of farms from the June Agricultural Survey and for the NOL domain estimates for Agricultural Labor Surveys and FCRS. Reports from the mid 1980's have shown that the RFO estimates are biased downwards [3,5,8]. The Crop Reporting Board Policy and Procedures Working Group and the Livestock, Dairy and Poultry Branch have already suggested that the RFO estimate be dropped.

CONCLUSIONS AND RECOMMENDATIONS

The above analyses indicate that both the observed estimator and the modeled estimator are practically indistinguishable from the current estimator. It seems the observed estimator is only slightly better than the modeled estimator at simulating the current estimator. This is probably due to the fact that the observed estimator requires more information than does the modeled estimator. Considering the effort required to accommodate the observed estimator, the modeled estimator is the better choice.

In order to eliminate the tedious and unproductive task of screening in densely populated segments, the modeled weighted estimator should be implemented as soon as possible. Conversion to the modified estimator further supports phasing out the RFO estimator in favor of alternative weighted estimators for the FCRS and Agricultural Labor Survey. In addition, highly specialized operations of interest with very little land and ranchers operating solely on PIGA land must be covered via the list frame.

APPENDIX 1: Glossary [7]

Area Sampling Frame - All land area in the State divided into sampling units called segments.

Farm - Land under one operating arrangement on which there were or are expected sales of at least $1,000 worth of crops, livestock, poultry, or other agricultural products during the year.

Farmstead - That part of the farm consisting of the main set of buildings (including the house and yard), adjacent pens or corrals, family garden and orchard.

Grain Stocks on Farm - All whole grains and oilseeds on hand or stored on the total acres operated, regardless of ownership or intended use. This includes grain and oilseeds in temporary or permanent storage facilities or on the ground.

List Sampling Frame (LSF) - A list of agricultural operators in a given State. Each classified operation name becomes a sampling unit. The name may be an individual, manager, farm or ranch, corporation, institution, etc.

Nonoverlap (NOL) - A tract operator in an area frame sample whose name is either not on the list sampling frame or is not eligible to be selected (not classified), from the list sampling frame.

PIGA - Public and Industrial Grazing Land Association.

Resident Farm Operator (RFO) - A farm operator whose residence is located within the boundaries of a segment. In the case of a partnership, there is only one operator per farm. The partner making most of the day-to-day decisions is considered to be the operator. When partners share equally in decision making, the oldest partner is considered to be the operator.

Tract - An area of land inside a segment under one type of land operating arrangement. It may have one or more parcels of land. A tract may consist of agricultural land (ag tract), nonagricultural land, residential areas, and or some other land use. A tract consists of land the operator owns and operates as well as land rented form others.

Weighted Estimator - The estimator used to prorate entire farm data to the tract [7]. For each tract, it requires the tract acreage, farm acreage, and data for each survey item. For example, if an operator has 80 hogs on his entire operation of 40 acres, and operates 10 acres in the tract, then the weighted value for the operation is: 80*10/40 = 20 hogs. The weighted estimate is computed by: (1) adding weighted values from the tracts to the segment level, (2) multiplying segment level data by the expansion factor, and (3) adding expanded segment data to the State level.

APPENDIX 2: Selected Commodity Estimates

	NOL CATTLE			NOL HOGS		
STATE	Current	Observed	Modeled	Current	Observed	Modeled
AL	677001	677167	677180	80163	79799	79427
AZ	54403	54361	54290	4129	4129	4116
CA	1040793	1040705	1040356	29778	29781	29752
CO	1039937	1039869	1039858	46520	46488	46492
ID	304262	304207	304182	33711	33703	33169
IN	267759	268034	268049	1015469	1038947	1045835
IA	1016101	1016143	1016194	2586142	2583019	2583535
KY	833277	833257	833354	284797	283822	283863
MN	317065	317105	317091	753694	762629	762826
NC	448039	448894	449031	196063	196940	196924
ND	357505	357431	357455	23288	23271	23271
OK	1401538	1401810	1401883	49746	49084	49007
OR	401387	401316	401184	16262	16268	16270
PA	338092	338203	338247	250514	250475	250356
TX	3725775	3726481	3726791	89598	89647	89681
WI	797190	797135	797166	269498	269647	269624
TOTAL	13020124	13022118	13022311	5729372	5757649	5764148
CHANGE		0.015%	0.017%		0.494%	0.607%

	NOL CORN STOCKS			NUMBER OF FARMS		
STATE	Current	Observed	Modeled	Current	Observed	Modeled
AL	297075	296657	296584	54130	54160	54155
AZ	27	27	27	5154	5131	5126
CA	2	2	2	83871	83991	83960
CO	2490636	2494066	2495071	28326	28358	28355
ID	56137	56024	56039	21063	21056	21048
IN	8208575	8222280	8225145	69342	69523	69572
IA	80074866	80056989	80067998	102648	102672	102688
KY	996406	995867	996036	90633	90758	90813
MN	21726352	21728284	21727744	94559	94610	94629
NC	1595999	1596178	1596257	65540	65700	65719
ND	2976745	2977481	2977527	34557	34550	34552
OK	30304	30204	30175	68102	68107	68109
OR	0	0	0	37844	37827	37812
PA	3383671	3383726	3383895	56180	56245	56261
TX	1361704	1363125	1363084	204370	204453	204440
WI	16813248	16813705	16815929	82278	82282	82275
TOTAL	140011747	140014615	140031513	1098597	1099423	1099514
CHANGE		0.002%	0.014%		0.075%	0.083%

APPENDIX 3: Selected Commodity Coefficients of Variation

	NOL CATTLE			NOL HOGS		
STATE	Current	Observed	Modeled	Current	Observed	Modeled
AL	11.11	11.11	11.11	26.25	26.43	26.61
AZ	44.58	44.61	44.67	53.92	53.92	54.07
CA	14.15	14.15	14.16	37.66	37.66	37.70
CO	23.32	23.33	23.33	54.32	54.31	54.30
ID	12.93	12.93	12.93	59.78	59.79	60.72
IN	15.85	15.84	15.84	69.92	69.53	69.38
IA	10.39	10.39	10.40	16.54	16.54	16.54
KY	9.82	9.83	9.83	26.19	26.27	26.27
MN	14.98	14.99	14.99	32.51	32.55	32.55
NC	12.78	12.83	12.84	22.21	22.19	22.21
ND	16.81	16.81	16.81	71.58	71.63	71.63
OK	11.78	11.78	11.78	38.37	38.86	38.93
OR	15.79	15.79	15.79	29.61	29.59	29.59
PA	12.40	12.40	12.40	45.10	45.09	45.08
TX	5.81	5.81	5.81	21.83	21.82	21.82
WI	10.74	10.74	10.74	33.71	33.68	33.68

	NOL CORN STOCKS			NUMBER OF FARMS		
STATE	Current	Observed	Modeled	Current	Observed	Modeled
AL	34.07	34.06	34.06	5.52	5.52	5.52
AZ	83.32	83.49	84.18	11.76	11.80	11.81
CA	98.84	98.84	98.84	7.68	7.68	7.68
CO	29.00	28.99	29.00	7.16	7.16	7.16
ID	48.50	48.51	48.50	6.81	6.81	6.81
IN	25.47	25.47	25.46	4.15	4.16	4.17
IA	17.07	17.08	17.08	2.91	2.91	2.91
KY	23.11	23.09	23.09	3.97	3.98	3.98
MN	24.64	24.64	24.64	3.96	3.96	3.96
NC	42.54	42.59	42.60	5.93	5.93	5.94
ND	50.28	50.28	50.29	3.29	3.30	3.30
OK	39.47	39.61	39.65	3.89	3.89	3.89
OR	---	---	---	8.91	8.91	8.91
PA	18.62	18.63	18.62	4.51	4.51	4.51
TX	41.78	41.77	41.77	3.81	3.81	3.82
WI	21.43	21.45	21.45	3.54	3.54	3.54

REFERENCES

[1] Bosecker, Raymond R. and Michael S. Clark. Modifying the Weighted Estimator To Eliminate Screening Interviews In Residential Areas. U.S. Department of Agriculture, National Agricultural Statistics Service, Staff Report No. SSB-88-07, December 1988.

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

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

[4] Hollander, Myles and Douglas A. Wolfe. Nonparametric Statistical Methods, New York: John Wiley and Sons, Inc., 1973.

[5] Matthews, Ralph V. Screening Residential Tracts for Agricultural Activity. U.S. Department of Agriculture, National Agricultural Statistics Service, Staff Report No. SSB-88-05, April 1988.

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

[7] National Agricultural Statistics Service. NASDA ENUMERATOR - Survey Interviewer's Manual. U.S. Department of Agriculture, NASS, 1989.

[8] 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.

* U.S. Government Printing Office 1990 - 261-161.20082