A COMPARISON OF THREE RULES FOR ASSOCIATING FARMS AND FARMLAND WITH SAMPLE AREA SEGMENTS IN AGRICULTURAL SURVEYS.

by

W. A. Hendricks, D. T. Searls and D. G. Horvitz
Research Triangle Institute

The statistical efficiencies of three rules for associating land-use and crop acreages with sample area segments in agricultural surveys are compared. Data from identical sample segments in the 1954 and 1959 Census of Agriculture Evaluation Programs were used to compute relative variances for simple expansion, ratio, regression and difference estimators for three regions of the United States for each of these association rules.

The Closed and Weighted Segment approaches appear to be more efficient than the traditional Headquarters Rule. The Weighted Segment Rule has a slight edge over the Closed Segment Rule both from the standpoint of sampling variability and correlations between 1954 and 1959 data.

1. INTRODUCTION

The application of area sampling techniques to sample farm surveys requires some rule for associating farms and farmland with the selected area segments. The rule that was adopted in the United States about a quarter of a century ago, when area sampling first began to be applied to farm surveys in this country, and which has been used most generally up to the present time, is the so-called "headquarters rule." By this rule a point on every farm which can be defined rigorously, and which can be identified by interviewers in the field, is employed as a reference point. If this point for a particular farm falls within the boundaries of the area segment, the farm is regarded as being "in" the segment. But if the reference point is outside the segment, the farm is considered outside the segment, even though some or most of all land in the farm may fall inside the segment.

The Census Bureau and other agencies involved in farm surveys have developed elaborate definitions of the reference point which is to serve as the "headquarters" of the farm in any particular instance, such as the following:

a. If the operator of the farm lives on the farm, his house is the headquarters.

b. If the operator does not live on the farm but there is one and only one house on the farm, that house is the headquarters.

1962

c. If there is more than one house on the farm and the operator does not live on the farm, the house of greatest value is the headquarters.

d. If there are no houses on the farm but other buildings are present, the building of greatest value is the headquarters.

e. If there are no buildings on the farm, the "main entrance" to the farm is the headquarters.

f. If no point can be identified as the main entrance, the corner of the farm farthest west and farthest north (in that order) is the headquarters.

Such sequences are established to permit the use of reference points that can be identified by interviewers with the least difficulty and possibility of error. However, errors in associating farms with area segments by a "headquarters" rule are usually all too prevalent. The problem is aggravated by difficulties and errors that often arise in deciding which parcels of land constitute the "farm."

To avoid these troublesome problems, the agency of the United States Department of Agriculture now known as the Statistical Reporting Service has in recent years made extensive use of a so-called "Closed Segment" rule. By this rule interviewers must account only for items on land that lies entirely within the segment boundaries. To obtain estimates of numbers of farms, all persons, or a sample of them, living within the segment boundaries must be screened to determine how many are farm operators. This is accomplished by ascertaining the nature and extent of their agricultural operations, if any, regardless of where they are performed. These persons are also interviewed to obtain data that can only be obtained for the farm as a whole, such as sales and purchases of various commodities. Wherever possible, estimates of such items are made to conform to related information obtained from land within the segment boundaries. For example, universe estimates of cattle inventories are obtained from the numbers found within the sample segments on the date of the enumeration. But estimates of sales are obtained by applying the ratio of farm sales to farm inventories, for entire farms as reported by farm operators in the segments, to the estimate of total inventories derived only from numbers within the segment boundaries.

This "closed segment" rule has several advantages over a "headquarters" rule, but it also has some disadvantages. On the plus side we note (1) the rule is less troublesome for interviewers to apply in the field, (2) as the land to be

accounted for in each sample segment is depicted on the interviewers' aerial photos, both interviewers and supervisors can recognize gross errors in reported data that might otherwise pass undetected, (3) reporting errors can be reduced because respondents are asked to report separately for specific tracts of land pointed out to them on the aerial photos, and in many cases where their holdings inside the segment boundaries represent only a portion of all their holdings, they are not required to disclose information about the portions outside the segment, and finally (4) between-segment sampling variation can be reduced because the boundaries of each segment place a limit on the total land to be accounted for in connection with that segment.

On the minus side we observe that farm operators who live in cities and towns can only be counted if some sample segments are allocated to urban areas. The task of identifying farm operators in such areas is often formidable and considerable undercounting may result. Furthermore, there is also the risk that some urban residents classified as farm operators may have tenants on their farms who whould be classified as the operators if they fell into the open-country portion of the sample. This is perhaps the greatest defect in the closed segment procedure. As pointed out previously, the fact that the method is not well adapted to surveys seeking data which can be reported conveniently only for farms as a whole must also be considered on the debit side.

Another approach, which for want of a better name has been called the "weighted segment" procedure, seems to offer a solution that retains many of the advantages of both the "headquarters" and "closed segment" rules and is also free of the most serious objections levelled at those two methods. So far as the authors of this paper are aware, it has not been applied to farm surveys as yet, although an agency of the Department of Agriculture has sponsored a rural land-ownership survey in which this method was employed. The Statistical Reporting Service made use of a similar principle in a survey for estimating the availability of farm grain storage facilities a few years ago.

As applied to farm surveys, the "weighted segment" approach regards every farm with some of its land inside a sample segment as associated with that segment. Data are recorded for every such farm as a whole, but are multiplied by the fraction of the farmland in the farm that falls within the segment before incorporating them into segment totals.

This procedure has a number of advantages over the two previous rules. First, it is a rule that can be applied by interviewers in the field with less

difficulty and possibility of error than a headquarters rule. Also the need for canvassing urban areas is eliminated and the sample can be confined to the open country. All data are recorded for entire farms; hence no special treatment is required for items that can not be reported conveniently for portions of farms as in the closed segment approach. At the same time the weighting of the data for each farm by the fraction of its land falling within the segment can reduce the between-segment variability of segment totals to a level comparable to that obtained with closed segment data. In fact, this variability can be expected to be lower because of the averaging effect achieved by prorating entire-farm data to land within the segment, rather than recording data only for the land within the segment.

Among the disadvantages, as compared with the closed segment procedure, we observe that interviewers are still faced with the problem of deciding which parcels of land must be defined as a "farm" and this is sometimes difficult. However, errors introduced by including too much land, or too little, in defining a farm tend to be partially neutralized by the weighting procedure. Tha fact that all land reported as being in a farm may not be covered by the interviewers' aerial photo eliminates some of the visual verification that can be performed by supervisors on closed segment data. The weighting that must be applied to individual farm data is a computational step that is not required with either the headquarters or closed segment rules, but with automatic data processing procedures that are now in rather general use this is not a serious matter. With sample segments of a given size, data must be reocrded for about twice as many farms by this rule as compared with a headquarters rule. With long questionnaires this can increase the time required to be spent in each sample segment, although it would by no means double the time required with a headquarters rule. The proper application of a headquarters rule requires a complete canvass of each sample segment to ascertain the places eligible for enumeration; considerable time is often consumed in screening out ineligible places.

2. OBJECTIVES OF THE PRESENT STUDY

The purpose of the present study was to investigate the sampling variances encountered with each of the preceding rules for associating farms with sample area segments. Although some fragmentary information on the headquarters and closed segment rules has accumulated over the past few years, no systematic comparison of the two approaches in this respect has every been made. So far as the

weighted segment rule is concerned, objective data on variances are non-existent.

The data employed in this study were obtained in the 1954 and 1959 Census of Agriculture Evaluation Programs. Data from 384 identical segments in 175 primary sampling units (PSU) were available for both of those years in a form that made such an analysis possible. In addition to detailed information about the characteristics of each farm covered in the Evaluation Program, data were recorded separately for the portions of those farms that fell inside the segment boundaries. To investigate the behavior of the headquarters rule, only two alternative reference points on each farm were considered as headquarters: (1) if the operator lived on the farm, his residence was the headquarters, and (2) if the operator did not live on the farm, the point on the farm farthest west and farthest north (in that order) was the headquarters. Sketches of the farm in relation to the segment boundaries were available to make the appropriate determination in each case.

All relevant information was placed on punch cards to facilitate the computations, which were performed on an IBM 1401 computer. The items studied in the analysis were:

1. Numbers of farms (omitted in closed segment analysis)
2. Acres of farmland
3. Acres of cropland harvested
4. Acres of corn harvested
5. Acres of wheat harvested
6. Acres of cotton harvested
7. Acres of soybeans harvested
8. Acres of oats harvested
9. Acres of hay cut
10. Acres of tobacco harvested

Between segment variability was estimated for the above items within primary units, separately for each of three regional strata of the United States and for each of the three rules employed to associate sample farms with sample area segments. Coefficients of correlation between 1954 and 1959 data were also computed for each rule of association. Variances and covariances were computed within primary units under the assumption that a large sample of farms in the United States would be of a one-stage design with sample segments allocated proportionally to all PSU's in the universe rather than a two-stage design of the kind used in the Evaluation Program. To indicate how much improvement could be

attained by basing 1959 estimates of agricultural items upon 1954 data through the use of difference, ratio, and regression estimators, the variability of such estimates was computed on a per-segment basis for comparison with the per-segment variances of the 1959 segment totals. To investigate possible gains in precision by excluding large farms from the area sample, all variances and covariances were computed with data for farms of 2000 acres or more included and excluded.

The three regional strata and the number of sample segments in each are as follows:

Region I-North (154 segments)	Region II-South (116 segments)	Region III-West (114 segments)
Connecticut	Alabama	Arizona
Illinois	Arkansas	California
Indiana	Delaware	Colorado
Iowa	Florida	Idaho
Maine	Georgia	Kansas
Massachusetts	Kentucky	Montana
Michigan	Louisiana	Nebraska
Minnesota	Maryland	Nevada
Missouri	Mississippi	New Mexico
New Hampshire	N. Carolina	N. Dakota
New Jersey	S. Carolina	Oklahoma
New York	Tennessee	Oregon
Ohio	Virginia	S. Dakota
Pennsylvania		Texas
Rhode Island		Utah
Vermont		Washington
West Virginia		Wyoming
Wisconsin		

These regions do not conform exactly to the regions for which official agricultural statistics are customarily summarized separately. The strata employed here were adopted mainly to achieve greater equality in the number of sample segments per stratum.

3. COMPUTATIONAL METHODS

The 175 primary units in the three regions had been selected with unequal probabilities and the sampling rates within those PSU's were adjusted correspondingly to arrive at a self weighted sample. For the present analysis some PSU's containing only one sample segment each needed to be combined with other PSU's to permit the computation of within-PSU variability. After those combinations were made, the 384 segments were contained in 124 new PSU's which were regarded as being selected with probabilities comparable to the original PSU. The within-PSU sampling rates were regarded as proportional to the reciprocals of those probabilities. The average 1959 variance between segments within PSU's for a given region was therefore computed from the formula

$$s_y^2 = \frac{1}{n} \sum_i \frac{n_i}{n_i - 1} \sum_j (y_{ij} - \bar{y}_i)^2 \qquad (1)$$

in which

y_{ij} = a 1959 segment total for the j-th segment in the i-th PSU.

$\bar{y}_i$ = the per-segment average of the y_{ij} for all segments in the i-th PSU.

n_i = the number of sample segments in the i-th PSU.

n = the total number of sample segments in the region.

Average covariances between 1954 and 1959 data were computed in a similar fashion:

$$s_{xy} = \frac{1}{n} \sum_i \frac{n_i}{n_i - 1} \sum_j (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) \qquad (2)$$

in which x_{ij} and y_{ij} are comparable segment totals in 1954 and 1959.

Clearly, if an estimated universe total of an agricultural item is to be made for a current year, such as 1959, without reference to related data for previous years, the relvariance (RV) of that estimate will be equal to the relvariance of the per-segment average for that item in the sample. If such an estimate is represented by Y_1, we have

$$RV(Y_1) = \frac{s_y^2}{n\bar{y}^2} \qquad (3)$$

in which s_y^2 is the between segment variance of the sample segment totals, $\bar{y}$ is the per-segment average of all segment totals, and n is the number of segments in the sample.

If a difference estimate, which makes use of data for the universe and the sample in a previous year such as 1954 is computed, such an estimate takes the form

$$Y_2 = X + (Y_1 - X_1) \tag{4}$$

in which X is the universe total in the base year, Y_1 is the current year estimate obtained by applying the reciprocal of the sampling rate to the current year sample total, and X_1 is the corresponding estimate of X derived from base year data in the sample. The relvariance of Y_2 is given by

$$RV(Y_2) = \frac{s_x^2 + s_y^2 - 2s_{xy}}{n\bar{y}^2} \tag{5}$$

in which the various terms are self-explanatory.

Instead of employing a difference estimator, one might invoke a ratio estimate of the form

$$Y_3 = \frac{Y_1}{X_1} X \tag{6}$$

The relvariance of Y_3, written in terms of the relvariances of X_1 and Y_1, together with the relative covariance (RCV) of X_1 and Y_1 is approximately:

$$RV(Y_3) = RV(X_1) + RV(Y_1) - 2RCV(X_1Y_1) \tag{7}$$

A third alternative would be to employ a regression estimator of the form

$$Y_4 = Y_1 + b(X - X_1) \tag{8}$$

in which b is the regression coefficient of y on x and the other symbols have the same meaning as before. The relvariance of Y_4 is approximately

$$RV(Y_4) = (1-r_{xy}^2)\, RV(Y_1) \tag{9}$$

in which r_{xy} is the coefficient of correlation between 1954 and 1959 segment totals within PSU's.

4. NUMERICAL RESULTS

Simple Expansion Estimates

The relvariance of a simple expansion estimate, as indicated previously, is identical with the relvariance of the per-segment average of that item. The estimated relvariances for all farms for each of the selected items under consideration are shown in Table 1 on a per-segment basis (n=1) by region for each of the three association rules, and also with farms of 2000 acres or more excluded. No such large farms were in the Region I sample.

The table indicates that the exclusion of large farms from the sample (Regions II and III) did not reduce the relative variances to any great degree. Aside from farmland itself, some reduction was effected in the relative variances of a few items such as corn, oats, and hay acreages in Region II.

The Closed Segment and Weighted Segment Rules tend to yield the lowest variances, with the Weighted Segment showing a slight edge over the Closed Segment. Table 2 shows that the use of the Weighted Segment Rule about doubles the number of farms from which information is obtained as compared with the Headquarters approach. This does not double the interview cost because interviewers must account for all land in a segment to identify farms with headquarters in the segment.

1954-59 Correlation Coefficients

Average correlation coefficients between segments within PSU are shown in Table 3 for the items covered in Table 1. Correlations are highest in Region I and lowest in Region III. In Region II they are a bit larger when large farms are retained in the sample, but in Region III the reverse seems to be true in several instances. One extremely large place of about 128,000 acres in Region III was omitted from the computations because it tended to dominate the results unduly. Generally speaking, correlations are highest with the Closed and Weighted Segment Rules of association. Of these two, the difference is again in favor of the Weighted Segment Rule.

The correlations are of sufficient magnitude, almost in general, to yield a considerable increase in statistical efficiency with estimation procedures which include prior Census or base year information as compared to simple

Table 1. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Simple Expansion Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.388	*	0.381	0.388	*	0.381
Farmland	.544	.420	.420	.544	.420	.420
Cropland Harvested	.720	.561	.567	.720	.561	.567
Corn Acreage	.913	.858	.746	.913	.858	.746
Wheat Acreage	2.150	2.625	1.998	2.150	2.625	1.998
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	2.418	1.630	1.343	2.418	1.630	1.343
Oats Acreage	.974	1.038	1.123	.974	1.038	1.123
Hay cut Acreage	.870	.844	.781	.870	.844	.781
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.598	*	.713	.596	*	.710
Farmland	1.775	1.194	1.194	1.335	.992	.992
Cropland Harvested	3.914	3.174	2.585	4.019	3.185	2.654
Corn Acreage	3.241	2.776	1.950	2.092	1.749	1.418
Wheat Acreage	11.142	9.581	8.238	11.046	9.495	8.148
Cotton Acreage	6.188	2.228	2.017	6.135	2.218	2.003
Soybean Acreage	20.219	18.021	17.975	20.043	17.863	17.818
Oats Acreage	12.685	13.320	9.105	8.194	5.794	5.323
Hay cut Acreage	7.296	6.585	5.405	6.567	5.620	5.304
Tobacco Acreage	7.837	5.902	6.804	7.769	5.850	6.745
			Region III			
Farms	.553	*	.440	.546	*	.436
Farmland	5.612	.878	.878	.721	.688	.688
Cropland Harvested	1.086	.604	.577	1.098	.555	.563
Corn Acreage	1.142	.750	.643	1.092	.740	.641
Wheat Acreage	1.326	1.186	1.208	1.347	1.134	1.129
Cotton Acreage	12.289	15.264	11.742	12.182	14.455	12.039
Soybean Acreage	29.449	31.518	26.694	29.196	31.246	26.450
Oats Acreage	2.536	2.026	2.130	2.473	2.018	2.136
Hay cut Acreage	1.133	.961	.619	1.042	.944	.639
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

Table 2. Farms with Land in Segment by Location of Headquarters, 1959 EPA Rural Area Sample

Region[1/]	Number of farms in sample		
	With Hq. in Segment	With Hq. Not in Segment	Total
North	1192	1251	2443
South	1072	1033	2105
West	291	240	531
U. S.	2555	2524	5079

[1/] The regions are as defined for the 1959 Census, not as for this study.

Table 3. Estimated Average Within Primary Sampling Unit Correlations between 1954 and 1959 Area Segment Totals obtained with Three Rules of Associating Farmland with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.716	*	0.798	0.716	*	0.798
Farmland	.744	.942	.942	.744	.942	.942
Cropland Harvested	.828	.960	.966	.828	.960	.966
Corn Acreage	.810	.839	.922	.810	.839	.922
Wheat Acreage	.623	.857	.905	.623	.857	.905
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	.686	.705	.793	.686	.705	.793
Oats Acreage	.750	.722	.895	.750	.722	.895
Hay cut Acreage	.732	.751	.752	.732	.751	.752
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.799	*	.686	.799	*	.686
Farmland	.887	.932	.932	.832	.918	.918
Cropland Harvested	.851	.947	.936	.828	.930	.928
Corn Acreage	.828	.851	.834	.657	.704	.758
Wheat Acreage	.608	.731	.656	.610	.730	.657
Cotton Acreage	-.070	.685	.734	-.070	.405	.594
Soybean Acreage	.574	.820	.702	.574	.824	.699
Oats Acreage	.697	.769	.728	.182	.374	.360
Hay cut Acreage	.660	.671	.690	.360	.379	.538
Tobacco Acreage	.962	.957	.960	.962	.957	.960
			Region III			
Farms	.691	*	.714	.704	*	.718
Farmland	.855	.319	.319	.429	.378	.378
Cropland Harvested	.597	.817	.730	.684	.772	.740
Corn Acreage	.430	.559	.661	.491	.582	.663
Wheat Acreage	.446	.892	.682	.537	.827	.665
Cotton Acreage	.663	.743	.741	.663	.652	.712
Soybean Acreage	.250	.578	.344	.250	.578	.344
Oats Acreage	-.013	.572	.576	.571	.654	.738
Hay cut Acreage	.388	.388	.479	.500	.367	.453
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

expansion estimation. This is verified in Tables 4, 5 and 6. The average correlations between segments within PSU's for the years immediately following a Census year should be even larger than those obtained for the 5-year interval in this study. It should be pointed out, however, that base year data assembled for the sample segments during a Census should be in a form comparable to the data that will be collected according to the specified association rule in intercensal sample surveys.

The negative correlations for cotton acreage in Region II with the Headquarters Rule were due to two segments with fairly large acreages in 1954, but small cotton acreages, according to this rule in 1959. This was due to a shift of the headquarters of one or two farms out of these segments in 1959, rather than a larger reduction in cotton acreage between 1954 and 1959. The correlations for cotton acreage in this region for the Closed and Weighted Segment Rules remained substantially positive.

Difference Estimates

The relative variances of 1959 difference estimates, based on 1954 data from the same segments with the same rules of association, are shown in Table 4. As compared with the variances of the simple expansions shown in Table 1, the results are as would be expected from the magnitudes of the correlation coefficients involved. Most items show some improvement in Regions I and II. Improvement was least noticeable in Region III.

The superiority of the Closed and Weighted Segment approaches is quite noticeable and as expected from the higher correlations. In a number of instances, the indicated gain in statistical efficiency would more than offset the increased cost, if any, of the Weighted Segment approach.

Ratio Estimates

The relative variances of 1959 ratio estimates, comparable to the difference estimates of the preceding section, are given in Table 5. As anticipated there are no striking differences between the results in Tables 4 and 5. In some cases the difference estimates are better but in others the ratio estimates have less variability. There seems to be a small edge in favor of ratio estimates.

Table 4. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Difference Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.286	*	0.178	0.286	*	0.178
Farmland	.350	.049	.049	.350	.049	.049
Cropland Harvested	.274	.045	.039	.274	.045	.039
Corn Acreage	.335	.257	.112	.335	.257	.112
Wheat Acreage	2.564	1.214	.584	2.564	1.214	.584
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	1.775	.909	.640	1.775	.909	.640
Oats Acreage	.663	.713	.244	.663	.713	.244
Hay cut Acreage	.445	.446	.367	.445	.446	.367
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.589	*	.481	.584	*	.476
Farmland	.453	.222	.222	.540	.272	.272
Cropland Harvested	1.226	.328	.320	1.420	.437	.374
Corn Acreage	1.213	.857	.734	1.614	1.088	.848
Wheat Acreage	7.030	4.473	4.854	6.945	4.446	4.765
Cotton Acreage	14.314	4.494	2.195	12.353	5.975	2.703
Soybean Acreage	16.882	11.100	12.383	16.736	11.094	12.393
Oats Acreage	15.347	13.032	12.215	29.185	30.813	21.817
Hay cut Acreage	4.197	3.620	2.957	6.690	5.454	4.280
Tobacco Acreage	1.742	1.816	1.322	1.727	1.800	1.310
			Region III			
Farms	.391	*	.309	.374	*	.296
Farmland	1.514	2.551	2.551	.974	.716	.716
Cropland Harvested	1.309	.262	.367	1.051	.305	.322
Corn Acreage	1.120	.574	.419	1.010	.548	.423
Wheat Acreage	4.029	.486	.780	2.546	.591	.754
Cotton Acreage	48.395	13.032	14.932	47.972	21.355	17.584
Soybean Acreage	29.054	21.442	24.622	28.804	21.258	24.397
Oats Acreage	8.296	2.174	1.799	3.093	2.045	1.189
Hay cut Acreage	2.114	1.544	.924	1.587	1.573	.974
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

Table 5. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Ratio Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.209	*	0.140	0.209	*	0.140
Farmland	.309	.048	.048	.309	.048	.048
Cropland Harvested	.278	.045	.038	.278	.045	.038
Corn Acreage	.374	.282	.123	.374	.282	.123
Wheat Acreage	2.198	1.320	.540	2.198	1.320	.540
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	2.035	1.206	.825	2.035	1.206	.825
Oats Acreage	.507	.533	.236	.507	.533	.236
Hay cut Acreage	.454	.393	.348	.454	.393	.348
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.257	*	.378	.255	*	.376
Farmland	.394	.162	.162	.458	.176	.176
Cropland Harvested	1.118	.400	.381	1.301	.524	.446
Corn Acreage	1.022	.767	.600	1.230	.922	.639
Wheat Acreage	7.573	4.953	5.463	7.442	4.899	5.326
Cotton Acreage	9.817	1.514	1.028	9.732	2.828	1.538
Soybean Acreage	13.564	6.915	9.618	13.447	6.603	9.486
Oats Acreage	8.388	6.379	4.987	16.951	11.466	8.512
Hay cut Acreage	4.168	3.647	2.859	6.295	5.263	3.945
Tobacco Acreage	.602	.517	.540	.596	.512	.535
			Region III			
Farms	.302	*	.234	.287	*	.227
Farmland	1.742	2.052	2.052	.787	.705	.705
Cropland Harvested	.924	.226	.320	.756	.259	.290
Corn Acreage	1.041	.556	.392	.931	.530	.397
Wheat Acreage	2.448	.314	.708	1.722	.421	.684
Cotton Acreage	24.565	7.005	6.383	8.830	9.022	6.531
Soybean Acreage	58.914	42.573.	48.035	58.388	42.174	47.613
Oats Acreage	4.615	1.523	1.470	2.017	1.361	.979
Hay cut Acreage	1.324	1.112	.742	1.014	1.120	.786
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

Regression Estimates

Regression estimates should have less variability than difference or ratio estimates because sampling fluctuations in base data have only a negligible effect, whereas in difference and ratio estimates such fluctuations exert considerable effect. For the present computations the effects of sampling fluctuations in 1954 base data on the regression estimates were ignored completely. The results are shown in Table 6.

As expected, the relative variances are generally smaller than for the other kinds of estimates. The Closed and Weighted Segment Rules of association again show considerable superiority, with the latter being a bit better. Eliminating the large farms from the sample did not change the relative variances appreciably except in a few instances. This was also true for the difference and ratio estimates. The items affected the most by removing the large farms were oats acreage in Regions II and III and farmland in Region III.

5. SUMMARY AND CONCLUSIONS

On the basis of the variances observed in this study, the Closed and Weighted Segment Rules of association are decidely preferable to the Headquarters Rule. This holds true for each of the four methods of estimation considered. Relative variances generally are a bit lower for the Weighted Segment Rule than for the Closed Segment Rule.

The elimination of farms of 2000 acres or more from the sample reduced the relative variances somewhat, but the reduction was not particularly striking. However, the number of large farms in the sample was small. When large farms are eliminated from an area sample and treated separately, the relative variance of the estimate for the two strata combined may be reduced appreciably.

When current estimates are computed by difference, ratio or regression procedures with matching data from a previous Census year, an appreciable reduction can be achieved in the relative variances of the agricultural items studied. The reductions in Region III (where the correlations between the 1954 and 1959 data were lower) would probably be less than in Regions I and II. Ratio and difference estimates would tend to have similar precision, with perhaps a slight edge for the ratio estimates. As expected, regression estimates would be the most precise.

Table 6. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Regression Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.189	*	0.138	0.189	*	0.138
Farmland	.243	.048	.048	.243	.048	.048
Cropland Harvested	.226	.044	.038	.226	.044	.038
Corn Acreage	.314	.254	.112	.314	.254	.112
Wheat Acreage	1.315	.690	.361	1.315	.690	.361
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	1.281	.819	.499	1.281	.819	.499
Oats Acreage	.426	.497	.224	.426	.497	.224
Hay cut Acreage	.403	.368	.339	.403	.368	.339
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.216	**	.378	.215	*	.376
Farmland	.378	.157	.157	.410	.157	.157
Cropland Harvested	1.081	.326	.318	1.264	.431	.367
Corn Acreage	1.020	.483	.594	1.189	.882	.603
Wheat Acreage	7.025	4.465	4.689	6.936	4.440	4.626
Cotton Acreage	6.158	1.181	.932	6.105	1.854	1.296
Soybean Acreage	13.563	5.918	9.129	13.446	5.743	9.104
Oats Acreage	6.515	5.441	4.286	7.922	4.986	4.632
Hay cut Acreage	4.121	3.619	2.831	5.718	4.814	3.772
Tobacco Acreage	.592	.500	.532	.586	.495	.528
			Region III			
Farms	.289	*	.215	.275	*	.212
Farmland	1.511	.789	.789	.589	.590	.590
Cropland Harvested	.699	.201	.270	.585	.224	.255
Corn Acreage	.931	.516	.362	.829	.490	.360
Wheat Acreage	1.062	.242	.646	.959	.358	.630
Cotton Acreage	6.891	6.832	5.300	6.831	8.303	5.933
Soybean Acreage	27.614	20.973	23.534	27.377	20.791	23.319
Oats Acreage	2.535	1.363	1.424	1.666	1.155	.971
Hay cut Acreage	.963	.817	.477	.781	.817	.508
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

The gain in precision to be achieved with the Closed and Weighted Segment Rules as compared to the Headquarters approach is even greater with estimation procedures which make use of base year data, since the correlations are generally higher for these approaches.

Obviously, other considerations in addition to the size of the sampling error must be taken into account when recommending one procedure over another. However, when the farm as a whole is regarded as the unit of observation, the Weighted Segment Rule appears to be preferable to other rules that might be considered for associating farms with sample segments. Not only are sampling errors smaller with this rule; the rule is less likely to be misinterpreted or misapplied by interviewers in the field.

For items where the farm as a whole does not necessarily have to serve as the unit of observation, the Closed Segment Rule has much to commend it. Perhaps the most serious objection to this rule is that some supplemental procedure must be used in conjunction with it to arrive at a count of number of farms and to obtain data on items that apply to the farm as a whole. The rule does have the advantage that interviewers have less difficulty determining the land to be covered in the enumeration. If all of the tracts enumerated are delineated on aerial photos, interviewers and supervisors can perform approximate visual verifications on much of the data reported by respondents. Both of these properties should have the effect of reducing nonsampling errors.