United States
Department of
Agriculture

**Statistical
Reporting
Service**

**Statistical
Research
Division**

**SRS Staff Report
Number AGES 831021**

October 1983

Comparison of the Operational and Cropland Weighted Estimators

Dave Dillard
Jack Nealon

COMPARISON OF THE OPERATIONAL AND CROPLAND WEIGHTED ESTIMATORS. By Dave Dillard and Jack Nealon; Statistical Research Division; Statistical Reporting Service; U.S. Department of Agriculture; Washington, D. C. 20250; October 1983. SRS Staff Report No. AGES831021.

ABSTRACT

A five-state study during the 1982 June Enumerative Survey compared two weighted estimators--the Agency's operational weighted estimator based on a total land weight and a cropland weighted estimator based on a cropland weight. The results of the study support earlier research that indicated that the operational weighted estimator is biased upward in some states. The cropland weighted estimator showed potential as a more reliable estimator.

CONTENTS

SUMMARY

This study compared two weighted estimators in five states using data from the 1982 June Enumerative Survey (JES). The two estimators were the operational weighted estimator and a cropland weighted estimator. The weight for the operational estimator was the ratio of tract acreage to entire farm acreage. The weight for the cropland estimator was the ratio of tract cropland (excluding field waste) to total cropland (excluding field waste). A past study (3) showed that farmers are more likely to underreport their total acreage than to overreport it, giving an upward bias to the operational estimator. The purpose of this study was to compare the expansions and standard errors between the cropland and the operational weighted estimators, to explain the reasons for significant differences between the two estimators, and to determine if the cropland weighted estimator may be a reasonable alternative to the operational weighted estimator.

The operational and cropland weighted estimators provided significantly different expansions at the 5 percent level in several states and in the five states combined for number of farms, planted acreage of corn and soybeans, and certain cattle variables. Differences between the estimators approached significance for certain hog variables. When differences were significant, the operational weighted expansions were always higher than the tract, farm, and cropland weighted expansions. Differences were more highly significant in states with a high proportion of noncropland, such as idleland, woodland, and wasteland.

There were only a few significant differences between the cropland weighted estimator and the tract and farm estimators. Many more significant differences occurred when comparing the operational weighted estimator with the tract, farm, and cropland weighted estimators. The cropland estimator was more precise than the tract and farm estimators, but not as precise as the operational weighted estimator.

One drawback to the cropland weighted estimator is that the weight is undefined whenever a farm does not have any cropland. This occurred on 11 percent of the farms in this study. The cropland weighted estimator also had reporting errors associated with it, but it is unknown what effect these errors had on the expansions.

The results of this study confirmed earlier findings that the operational weight is biased upward in some states, particularly in states with less cropland. However, the operational weighted estimator may not be biased in those states where farms have little noncropland associated with them.

Comparison of the Operational and Cropland Weighted Estimators

Dave Dillard
Jack Nealon

INTRODUCTION

The Statistical Reporting Service (SRS) conducts periodic surveys to estimate number of farms, livestock inventory, and crop acreages. The principal SRS survey is the June Enumerative Survey (JES). This survey uses a nationwide area sample to provide data for agricultural estimates and forecasts.

The JES uses a weighted estimator in ten states to provide statistics for number of farms and livestock inventory. Historically, this weighted estimator has been more precise than both the tract and farm estimators. For each operation, the weight for this estimator is the ratio of tract acreage to entire farm acreage. A past study by SRS (3) showed that farmers are more likely to underreport their entire farm acreage than to overreport it. This underreporting produces an upward bias of unknown magnitude in the weighted estimates for some states. To correct this, SRS has tried to find an alternative weighted estimator with high precision like the operational weighted estimator but without the bias.

This study examined one such alternative, the cropland weighted estimator. The weight for this estimator was based on cropland acreage excluding field waste. For this study, cropland was defined as all land planted or to be planted to crops in 1982 and also idle cropland, summer fallow, and cropland pasture. Appendix B shows the page of the JES questionnaire that asked about entire farm cropland acreage. The purpose of this study was to compare the expansions and standard errors between the cropland weighted estimator and the operational weighted estimator, to explain the reasons for significant differences between the two estimators, and to determine if the cropland weighted estimator may be a reasonable alternative to the operational weighted estimator.

The study used data from the 1982 JES in five states--Georgia, Indiana, Missouri, North Carolina, and Ohio. These states were selected because of their geographic differences and because all use the weighted estimator. The data included the entire area frame sample except for hog and cattle extreme operators.

This report begins by summarizing past SRS research that led to the examination of the cropland weighted estimator. It then describes the estimators used in the analysis. The analysis is divided into three sections. The first section compares the cropland weighted estimator with the operational weighted, tract, and farm estimators for number of farms and certain livestock variables. The second section compares the proportion of the tract acreage used for cropland with the proportion of the entire farm used for cropland. The final section compares the tract estimates of corn and soybeans with the estimates obtained by using the two weighted estimators.

BACKGROUND

A 1974 study in Nebraska (1) involving 193 farm operators showed that area frame respondents made more mistakes reporting entire farm acreage than they did reporting any other land or livestock item. In that study, reinterview responses for total acreage differed from the original responses 32 percent of the time.

A three-state study in 1977 (3) involving 528 operators showed that farmers were more likely to underreport their entire farm acreage than to overreport it. The major reason for underreporting entire farm acreage was farmers' failure to include land not in active use, such as idleland, woodland, and wasteland.

A 1981 study (5) in three states focused on the frequency of reporting errors involving entire farm acreage. That study measured the repeatability of answers for total acres operated. Enumerators reinterviewed 414 respondents to determine how close the reinterview response was to the original response. Only about a third of the responses for entire farm acreage were the same on both interviews; about a third of the reinterview responses differed from the original response by less than 10 percent; and about a third of the responses differed from each other by more than 10 percent.

Because of the reporting errors on entire farm acres, SRS sought alternatives to the operational weighted estimator. In 1981, it examined two alternative weights for a weighted estimator in five states (4). One weight was based on entire farm acreage, excluding continuous parcels of wasteland, woodland, and other nonagricultural land. That weight was subject to many reporting problems and the study recommended not investigating it further. The other weight, based on cropland, was subject to fewer nonsampling errors. The study recommended conducting research on a modification of the cropland weight.

TRACT, FARM, AND WEIGHTED ESTIMATORS

This section describes the estimators used in the data analysis. Appendix C contains the formulas for the estimators and their estimated variances. Each estimator relies on the expansion of a particular value.

Tract Value: For each operation, the tract value is the number of acres or livestock physically located within the tract at the time of the interview. For example, suppose a farmer had 150 hogs, of which 40 were physically located in the tract, and no one else had hogs in the tract. The tract value for number of hogs would be 40.

Entire Farm Value: For each operation , the entire farm value is 0 if the farm operator lives outside the tract and is the number of acres or livestock located on the entire farm if the operator lives inside the tract. In the example above, suppose the farmer had 150 hogs on his entire farm and he was a resident agricultural operator (RAO). His farm value would be 150. If the farmer was not an RAO, his farm value would be 0.

Operational Weighted Value: For each operation, the operational weight is the ratio of tract acreage to entire farm acreage. The operational weighted value is the product of the weight and the number of acres or livestock for the entire farm. Suppose the farmer in our example had 300 acres on his entire farm, 100 of which were inside the tract. His weight would be 100/300, or 1/3. The weighted value would be 1/3x150, or 50.

Cropland Weighted Value: For each operation, the cropland weighted value is the same as the operational weighted value except for the weight. The cropland weight is based on cropland acreage rather than on entire farm acreage. If our farmer had 75 acres of cropland within the tract and 250 acres of cropland on his entire farm, his cropland weight would be 75/250, or 3/10. The cropland value would be 3/10x150, or 45.

A weight is undefined whenever the denominator of the weight is zero or missing. Since entire farm acreage is always positive, the denominator of the operational weight was never zero in this study. When the enumerator could not obtain information on entire farm acreage, the survey statistician subjectively imputed a value, so that entire farm acreage was never missing. Thus, the operational weight was always defined for this study.

Whereas the entire farm acreage was always positive, the total cropland acreage was sometimes zero, making the cropland weight undefined. The cropland weight was also undefined when total cropland was missing because the survey statisticians did not impute values for cropland. It was felt that imputed values would not be as accurate as reported values and should not be used in this analysis. Since data on entire farm cropland acreage was collected for research purposes only and not for any operational estimates, there was no reason to impute values when cropland was missing. When the cropland weight was undefined, the tract or farm value for the variable of interest took the place of the undefined cropland weighted value. The tract value was used for inventory variables, the variables for which enumerators collected tract data.

Otherwise, the farm value was used. The 1981 study (4) used this same approach to handle the problem of undefined weights.

Table 1 shows how often entire farm acreage was imputed by survey statisticians and how often total cropland was missing or zero. Cropland acreage was missing more often than entire farm acreage was imputed, especially in Georgia. Statisticians in the Georgia office deleted entries for total cropland when they appeared incorrect, rather than changing the entries as they did for entire farm acreage. It is possible that statisticians in the other states edited cropland in a similar manner. This type of editing would help explain why cropland was missing so often.

Table 1-- Percent of farm operations with imputed farm acreage and missing or zero cropland data for the five states, using 1982 JES data

State	Entire farm acreage imputed	Total cropland missing	Total cropland zero
		Percent	
Georgia	5.8	16.7	17.2
Indiana	14.8	15.1	8.1
Missouri	8.2	9.6	9.3
North Carolina	7.0	9.6	12.5
Ohio	9.0	11.6	9.3
Five states	9.0	12.2	11.0

Altogether, cropland was missing or zero for 23 percent of the farms in the study. Since the tract or farm values replaced cropland values when the cropland weight was undefined, the cropland expansions would naturally be expected to be closer to the tract and farm expansions than the operational weighted expansions would. This problem of undefined cropland weights would be especially acute in states that have little cropland or much missing data.

Undefined cropland weights were not a problem as long as the variable of interest was zero. For then, all expansions would have the value zero. But when the variable of interest was positive, the various expansions would have a chance to be different. For the five states combined, an undefined cropland weight resulting from zero cropland acres occurred in conjunction with a positive value for the test variable 6.2 percent of the time for number of farms, 1.6 percent for total hogs and pigs, and 7.2 percent for total cattle and calves. Thus, undefined cropland weights were a greater potential problem when estimating number of farms and total cattle and calves than when estimating total hogs and pigs.

ANALYSIS OF ESTIMATES FOR LIVESTOCK VARIABLES AND NUMBER OF FARMS

The first part of the analysis examined the estimates for seven variables: number of farms; total hogs and pigs; expected farrowings in the next 6 months (June 1982-November 1982); previous farrowings in the last 6 months (December 1981-May 1982); total cattle and calves; calves born since January 1, 1982; and cows and heifers expected to calve by December 31, 1982. The purpose of this examination was to compare the operational and cropland weighted estimators and explain the reasons for any differences between them.

Tables 2 and 3 show the estimates and relative errors for the seven test variables at the five-state level. Relative error, also known as the coefficient of variation, is the ratio of the standard error of a survey estimate to the survey estimate. Tables A-1, A-2, and A-3 in Appendix A show the estimates and relative errors for the individual states. For the five states combined, the operational weighted expansion was greater than all the other expansions for all 7 variables. The cropland weighted expansion, on the other hand, was neither consistently above nor below the tract and farm expansions. The cropland weighted expansion was also closer to the tract and farm expansions than the operational weighted expansion was, with the exception of the tract expansion for total cattle and calves. The relative errors were highest for the farm and tract expansions, lower for the cropland weighted expansions, and lowest for the operational weighted expansions. Of course, the relative errors measure only the precision of the estimators and do not account for any biases that may be present.

Table 2--Estimates and relative errors for selected variables for the five states combined, using 1982 JES data

Estimator	Total hogs and pigs		Total cattle and calves	
	Estimate (000)	Relative error (%)	Estimate (000)	Relative error (%)
Tract	7,768	10.3	11,232	4.1
Farm	7,885	10.7	10,745	5.7
Operational weighted	8,679	5.6	11,276	2.8
Cropland weighted	8,211	6.5	10,346	3.5

Table 3--Estimates and relative errors for selected variables for the five states combined, using 1982 JES data1/

Estimator	Number of farms		Expected farrowings		Previous farrowings		Cows remaining to calve in 1982		Calves born since 1/1/82	
	Estimate (000)	Relative error (%)	Estimate (000)	Relative error (%)	Estimate (000)	Relative error (%)	Estimate (000)	Relative error (%)	Estimate (000)	Relative error (%)
Farm	398	2.7	1,149	11.3	978	11.2	1,596	6.4	3,016	6.3
Operational weighted	450	2.1	1,230	6.0	1,093	6.2	1,738	4.1	3,285	3.1
Cropland weighted	414	2.3	1,167	7.4	1,034	7.6	1,560	4.9	2,853	4.0

1/ Tract estimates were not available for these variables.

Table 4 shows the variables for which the operational weighted expansion was significantly different from the cropland weighted expansion. In Table 4, paired t-tests provided the significance levels for the univariate tests and Hotelling's t^2 statistic provided the significance levels for the multivariate tests. Appendix E describes the tests used.

Table 4-- Significance levels for the univariate and multivariate test statistics comparing the operational and cropland weighted expansions for the seven test variables for the five states, using 1982 JES data

State	Multivariate test on all variables	Number of farms	Total hogs and pigs	Expected farrowings	Previous farrowings	Total cattle and calves	Cows remaining to calve in 1982	Calves born since 1/1/82
Georgia	<.01*	<.01*	.99	.20	.30	.28	.05*	.01*
Indiana	.53	.58	.22	.27	.45	.71	.31	.60
Missouri	<.01*	<.01*	.12	.67	.62	.01*	<.01*	<.01*
North Carolina	<.01*	<.01*	.62	.30	.45	<.01*	<.01*	<.01*
Ohio	.05*	.02*	.76	.98	.50	.17	.77	.70
Five states	<.01*	<.01*	.08	.15	.13	<.01*	<.01*	<.01*

The symbol * denotes significance levels less than or equal to .05.

At the 5 percent level, the multivariate test on the seven variables showed that the operational and cropland weighted expansions were significantly different for each state but Indiana. Univariate tests showed the following results: (1) Number of farms was significantly different for each state except Indiana; (2) the cattle expansions were significantly different for the five states combined and for the individual states of Georgia, Missouri, and North Carolina; and (3) none of the hog expansions were significantly different at the 5 percent level. However, a few states had low significance levels for hogs, and total hogs was significant at the 10 percent level for the five states combined. When significant differences occurred between the two weighted estimators, the operational weighted estimator was always higher.

The most highly significant results were for number of farms. One explanation for this concerns the fact that the variable can be only zero or one. The farm variable took the value one when gross value of sales for the past year was at least $1,000 and zero otherwise. The livestock variables took on a much wider range of values. As a result, the variance associated with number of farms was smaller than for the other variables, and the tests were therefore able to detect smaller differences.

A positive value for the farm variable occurred about 92 percent of the time for the five states, compared with only 54 percent for total cattle and calves and 23 percent for total hogs and pigs. Thus, the expansions for number of farms had a chance to be different on 92 percent of the tracts, much more often than for the other variables. This higher percent provides another explanation why number of farms was so highly significant. It also helps explain why the cattle expansions were significant but the hog expansions were not.

Indiana showed no significant differences for any of the tests. Indiana also had the highest proportion of cropland to total land, 77 percent. Ohio followed with 69 percent, then Missouri, North Carolina, and Georgia with 49 percent, 43 percent, and 38 percent, respectively. Therefore, the operational weight and cropland weight were more nearly equal for Indiana than for the other states. Conversely, Georgia, Missouri, and North Carolina had the lowest proportion of cropland to total land; and they are the states that showed significant differences for number of farms and the cattle items. These findings are consistent with those of the 1977 study (3), which showed that farmers tend to underreport entire farm acreage when noncropland is present.

The next step in the analysis was to determine whether the significant differences between the two weighted estimators were the result of the contributions from the tract and farm expansions to the undefined domain of the cropland weighted estimator and the office imputations for total land--a component of the operational weighted estimator--or the result of differences in the weighting procedures. The undefined domain of the cropland weighted estimator was all tracts with zero or missing values for entire farm cropland acreage. Although the operational weighted estimator was still defined for tracts with imputed values for entire farm acreage, these tracts were deleted from this part of the analysis because imputed values were considered to be not as accurate as reported values. All tracts with imputed values for entire farm acreage will be referred to as the undefined domain of the operational weighted estimator. The data was reanalyzed after deleting all tracts belonging to the undefined domains of the two weighted estimators to determine if the significant differences between the two weighted estimators still remained. In this way, the undefined domains were eliminated from the comparison of both estimators. Appendices D and E provide the test procedure and the formulas for the estimates and variances used in this part of the analysis.

Table 5 summarizes this analysis for the three primary variables--number of farms, total cattle and calves, and total hogs and pigs. Whenever the weighted expansions were significantly different, the operational weighted expansion was higher than the cropland weighted expansion. Differences were significant in Georgia, Missouri, North Carolina, and the five states combined for number of farms and total cattle and calves and approached significance in Ohio. The results of the statistical tests in Table 5 are very similar to the results in Table 4, which included the undefined domain. Therefore, these results illustrate that the two weighted estimators differ significantly even when the contributions to the cropland weighted estimates from the farm and tract estimates and the imputed operational weights are eliminated.

Table 5-- The relative difference 1/ and significance level for the three variables tested when using tracts with a defined cropland weighted value and no imputed data for entire farm acreage for the five states, using 1982 JES data

State	Number of farms		Total cattle and calves		Total hogs and pigs	
	Relative difference	Significance level	Relative difference	Significance level	Relative difference	Significance level
Georgia	-7.4	.01*	-19.4	<.01*	-2.2	.83
Indiana	0.3	.81	-2.0	.38	0.6	.76
Missouri	-5.3	<.01*	-10.8	<.01*	2.2	.67
North Carolina	-13.0	<.01*	-29.6	<.01*	1.9	.80
Ohio	-2.4	.07	-5.3	.09	0.8	.79
Five states	-5.5	<.01*	-11.3	<.01*	1.1	.63

1/Relative Difference $= 100 * \frac{(\text{Cropland weighted expansion} - \text{Operational weighted expansion})}{\text{Operational weighted expansion}}$

The symbol * denotes significance levels less than or equal to .05.

Table 6 compares the cropland weighted expansion with the tract expansion for total hogs and pigs and for total cattle and calves, and with the farm expansion for the other five test variables. At the 5 percent level, 2 of the 6 multivariate tests and 3 of the 42 univariate tests showed significant differences. With so many tests performed, one would expect a few significant differences to arise purely by chance. Multivariate tests comparing the operational weighted expansions with the tract and farm expansions found significant differences in Georgia, Missouri, North Carolina,

Table 6--Signficance levels for the univariate and multivariate test statistics comparing the tract or farm expansions with the cropland weighted expansions for the seven test variables for the five states, using 1982 JES data

State	Multivariate test on all variables	Number of farms	Total hogs and pigs	Expected farrowings	Previous farrowings	Total Cattle and calves	Cows remaining to calve in 1982	Calves born since 1/1/82
Georgia	.87	.74	.60	.72	.64	.33	.49	.91
Indiana	.67	.21	.23	.27	.14	.52	.90	.60
Missouri	.04*	.10	.71	.21	.31	.23	.90	.47
North Carolina	.18	.85	.96	.13	.12	.01*	.56	.16
Ohio	.48	.29	.16	.23	.14	.18	.99	.90
Five states	<.01*	.03*	.53	.87	.55	.02*	.67	.33

The symbol * denotes significance levels less than or equal to .05.

and the five states combined. Univariate tests found 7 significant differences out of 42, more than twice as many as in Table 6. In addition, when differences between the operational weighted expansions and the tract and farm expansions were significant, the operational weighted expansion was always greater than the tract or farm expansion. For the 3 significant differences in Table 6 involving univariate tests, the cropland weighted expansion was greater than the farm expansion for number of farms for the five states combined and less than the tract expansion for total cattle and calves in North Carolina and the five states combined. So, differences between the cropland weighted expansions and the tract and farm expansions were not unidirectional as the differences between the operational weighted expansions and the tract and farm expansions were.

For the five states combined, the cropland weighted expansions for the cattle variables were below the tract and farm expansions. One possible explanation for this is that farmers incorrectly included permanent pasture as cropland acreage when reporting for the entire farm. Including permanent pasture as cropland acreage would inflate the value of entire farm cropland acreage, giving a downward bias to the cropland weight. However, the study found no objective evidence that farmers reported permanent pasture incorrectly. The only significant differences between the cropland weighted expansions and the tract and farm expansions for the cattle variables were for total cattle and calves in North Carolina ($\hat{\alpha}$=.01) and for the five states combined ($\hat{\alpha}$=. 02). Also, the cropland weighted expansions for cows remaining to calve in Indiana and Ohio and for calves born in Georgia were virtually the same as the farm expansions. Thus, the cattle expansions provided no evidence of bias associated with reporting cropland acreage as they did for reporting entire farm acreage.

ANALYSIS OF TRACT AND FARM RATIOS

The second part of the analysis compared ratios based on acreage within the tract with the corresponding ratios based on entire farm acreage. When collecting data on tract acreage, enumerators first partitioned the tract into fields, each of which was devoted to a single crop or land use. They then questioned the respondent about the acreage in each field to arrive at a total for the tract. This field-by-field partitioning was not done for the entire farm acreage. Instead, the enumerators simply asked the respondent to report the entire farm acreage. Since the enumerators were more meticulous about collecting tract data, the assumption was made that the tract data was more accurate than the entire farm data. The purpose of this part of the analysis was to identify problems with reporting cropland and total farmland.

Table 7 compares two estimates of the ratio of cropland to total farmland. The first estimate is the ratio of tract cropland acreage to tract total acreage. The second estimate is the ratio on the entire farm of cropland acreage to total acreage. Both ratios estimate the same proportion, so they should be approximately equal. For all states, however, the farm ratio exceeded the tract ratio. Differences between the two ratios were significant in Missouri and the five states combined and approached significance in Ohio. Appendix F presents the significance tests used in Table 7.

Table 7-- Ratios of cropland to total farmland when using tracts with a defined cropland weight and no imputed data for entire farm acreage, for the five states, and significance levels for testing whether the ratios are the same, using 1982 JES data

State	Ratio of cropland to total farmland		Significance level
	Tract	Farm	
	Percent		
Georgia	41.8	45.1	.32
Indiana	79.3	81.5	.20
Missouri	52.0	58.4	.03*
North Carolina	46.7 1/	51.3 1/	.21
Ohio	71.3	74.2	.11
Five states	58.2	63.7	<.01*

1/ The North Carolina area sample included an outlier, a large farm that appeared in one tract. Deleting the farm from the analysis changed the tract ratio for North Carolina from 46.1 percent to 46.7 percent and the farm ratio from 40.7 percent to 51.3 percent. The tract ratio for the five states changed from 58.1 percent to 58.2 percent and the farm ratio changed from 57.6 percent to 63.7 percent.

The symbol * denotes significance levels less than or equal to .05.

One explanation for the farm ratio being consistently higher than the tract ratio is that farmers underreported their total acreage. This explanation is consistent with there being a bias in the operational weight. Another explanation is that farmers overreported cropland--possibly by including waste or permanent pasture--although the study found no evidence to support this. In fact, the JES provided data on cropland waste in Missouri to determine whether incorrectly including waste as part of cropland could seriously bias the cropland weight. The ratio of cropland waste to cropland was only 2.0 percent, probably not enough to cause significant differences, and certainly not enough to explain the differences in Table 7.

ANALYSIS OF CROP ACREAGE ESTIMATES

At present, SRS uses tract expansions from the JES to estimate crop acreages. It could, however, estimate crop acreages by using weighted expansions similar to the ones discussed earlier. For research purposes only, the five states in this study collected entire farm data for corn and soybeans to make weighted estimates for those crops (see Appendix B). The purpose was to compare the weighted expansions with the tract expansions and thereby identify any further problems with the weighting procedures. Once again, the tract data was assumed to be more accurate than the entire farm data. Therefore, differences between the tract expansions and weighted expansions were considered more indicative of problems with the weights.

When comparing the operational weighted and cropland weighted expansions, all tracts with an undefined cropland weight or imputed data for entire farm acreage were deleted. In addition, tracts with missing data for entire farm corn acreage were deleted when computing the corn expansions, and tracts with missing entire farm soybean acreage were deleted for the soybean expansions. This procedure is similar to the one used for comparing the weighted estimators in Table 5 for the 7 test variables discussed earlier. Appendix D explains the procedure for computing these corn and soybean expansions.

Table 8 compares each of the weighted expansions with the tract expansions for corn and soybeans. Appendix E explains the derivation of the statistics used to obtain the significance levels in Table 8.

Corn planting was nearly complete by the time of the survey. Soybean planting was only about 75 percent complete in Indiana and Ohio and less than 50 percent complete in the other three states. Thus, farmers were required to report their planting intentions much more often for soybeans than for corn. This may have led to more reporting errors for soybeans. Therefore, less importance will be placed on the soybean results.

Table 8-- For corn and soybeans, relative differences 1/ and significance levels between the weighted and tract expansions when using tracts with a defined cropland weighted value and no imputed data for entire farm acreage, for the five states, using 1982 JES data

For Corn	Operational weighted vs. tract		Cropland weighted vs. tract	
	Relative difference (%)	Significance level	Relative difference (%)	Significance level
Georgia	25.7	<.01*	12.1	.15
Indiana	-2.7	.28	-4.2	.07
Missouri	11.8	.10	1.9	.75
North Carolina	11.8	.04*	-5.6	.16
Ohio	3.0	.24	0.7	.77
Five states	3.9	.03*	-1.4	.38

For Soybeans	Operational weighted vs. tract		Cropland weighted vs. tract	
	Relative difference (%)	Significance level	Relative difference (%)	Significance level
Georgia	-0.9	.80	-8.6	<.01*
Indiana	4.0	.24	3.0	.37
Missouri	-1.4	.67	-6.2	.03*
North Carolina	12.9	<.01*	-4.5	.18
Ohio	-1.2	.71	-2.2	.48
Five states	1.5	.38	-3.5	.02*

1/ Relative difference = $100 * \frac{(\text{Weighted expansion} - \text{Tract expansion})}{\text{Tract expansion}}$

The symbol * denotes significance levels less than or equal to .05.

The operational weighted expansion for corn was significantly greater than the tract expansion for the five states combined and individually for Georgia and North Carolina. This reinforces earlier findings that the operational weight is biased upward in states with a great deal of noncropland. The cropland expansion was not significantly different from the tract expansion in any state. In fact, the tract expansion for corn was closer to the cropland weighted expansion than to the operational weighted expansion for all states but Indiana.

Only in North Carolina were there significant differences between the operational weighted and tract expansions for soybeans. There were significant differences between the cropland weighted and tract expansions for the five states combined and individually for Georgia and Missouri. Thus, the results for soybeans contradict the results for corn and other variables in the study, but because soybeans were only partially planted, less importance will be placed on this contradiction.

Table 9 compares the operational weighted and cropland weighted expansions for corn and soybeans. The differences in Table 9 are more striking than those in Table 8. There were highly significant differences between the two weighted expansions for both corn and soybeans in Georgia, Missouri, North Carolina, and the five states combined. In addition, corn differences were significant in Ohio and approached significance in Indiana.

Table 9-- For corn and soybeans, relative differences 1/ and significance levels between the operational weighted and cropland weighted expansions when using tracts with a defined cropland weighted value and no imputed data for entire farm acreage, for the five states, using 1982 JES data

State	Corn		Soybeans	
	Relative difference (%)	Significance level	Relative difference (%)	Significance level
Georgia	-10.8	< .01*	-7.8	< .01*
Indiana	-1.5	.09	-1.0	.19
Missouri	-8.8	.01*	-4.8	.01*
North Carolina	-15.5	< .01*	-15.4	< .01*
Ohio	-2.3	.03*	-1.0	.22
Five states	-5.1	< .01*	-4.9	< .01*

1/ Relative difference $= 100 * \frac{(\text{Cropland weighted expansion} - \text{Operational weighted expansion})}{\text{Operational weighted expansion}}$

The symbol * denotes significance levels less than or equal to .05.

CONCLUSIONS AND RECOMMENDATIONS

Multivariate tests comparing the two weighted estimators showed highly significant differences for all states except Indiana. The most highly significant differences for the univariate tests occurred for number of farms, which was significantly different in all states except Indiana. There were significant differences for the cattle variables in Georgia, Missouri, North Carolina, and the five states combined. None of the hog differences were significant, but they approached significance at the five state level. When differences were significant, the operational weighted expansion was always higher than the cropland weighted expansion. This result is consistent with the findings of earlier research showing that the operational weight is biased upward.

Multivariate tests comparing the cropland weighted estimator with the tract and farm estimators showed significant differences only for Missouri and the five states combined. Univariate tests showed very few significant differences for the seven test variables. Number of farms was significant for the five states combined, and total cattle and calves was significant for North Carolina and the five states combined. None of the hog differences were significant. Many more significant differences occurred in comparisions involving the operational weighted estimator.

The study identified possible reporting errors associated with cropland during the office editing process. Most of these errors centered around farmers' failure to report all the cropland located outside the tract. It is unknown what effect these errors had on the expansions.

One drawback to the use of the cropland estimator is that the weight was undefined whenever there was no cropland anywhere on the entire farm. This occurred on 11 percent of the farms in this study. The relative errors for the cropland weighted expansions were also slightly higher than those for the operational weighted expansions, but were not as high as the relative errors for the tract and farm expansions.

The study showed no evidence of any bias associated with the cropland weight. When there were significant differences between the cropland expansion and the tract and farm expansions, the cropland expansion was neither consistently above nor below the other expansions. The cropland weighted expansion was also closer to the tract and farm expansions than the operational weighted expansion was.

SRS has tried to improve the reporting of entire farm acreage on the JES by reminding respondents to include woodland, pastureland, and wasteland and by emphasizing to enumerators the importance of obtaining accurate data on acreage. In spite of these efforts, many reporting errors remain. Although states should continue to minimize the number of such errors, it is unlikely that this will eliminate the bias in the estimator. Although the current weighted estimator is biased in states with much of their acreage in noncropland, it may not be biased in all states. Even in states where the estimator is biased, it may still provide an accurate indication of change from year to year. However, if the magnitude of the bias is some states varies from year to year, thereby making time trends unrecognizable, SRS should conduct additional research with the cropland weighted estimator for possible use in these states.

REFERENCES

(1)Bosecker, Raymond R., and William F. Kelly. Summary of Results from Nebraska Survey Concept Study. U.S. Department of Agriculture, Statistical Reporting Service, November 1975.

(2)Hansen, Morris H., William N. Hurwitz, and William G. Madow. Sample Survey Methods and Theory, Vol. 1. New York: John Wiley & Sons, Inc., 1953.

(3)Hill, George W., and Martha S. Farrar. Impact of Nonsampling Errors on Weighted Tract Survey Indications. U.S. Department of Agriculture, Statistical Reporting Service, December 1977.

(4)Nealon, Jack. An Evaluation of Alternative Weights for a Weighted Estimator. U.S. Department of Agriculture, Statistical Reporting Service, October 1981.

(5)____________. Response Errors in the Weighted Estimator. U.S. Department of Agriculture, Statistical Reporting Service, July 1982.

(6)Tatsuoka, Maurice M., Multivariate Analysis, New York: John Wiley & Sons, Inc., 1971.

APPENDIX A

Table A-1-- Estimates and relative errors for selected variables and estimators for the five states, using 1982 JES data

Variable	State	Estimator					
		Tract		Operational Weighted		Cropland Weighted	
		Est.	R.E.	Est.	R.E.	Est.	R.E.
		(000)	(%)	(000)	(%)	(000)	(%)
Total hogs and pigs	Georgia	906	18.7	850	14.5	851	16.8
	Indiana	2,277	19.5	2,873	9.7	2,713	10.5
	Missouri	2,763	19.7	2,820	10.1	2,576	13.0
	North Carolina	739	33.8	782	19.6	753	23.6
	Ohio	1,083	22.9	1,354	14.2	1,319	15.5
	Five states	7,768	10.3	8,679	5.6	8,211	6.5
Total cattle and calves	Georgia	1,683	9.8	1,679	6.6	1,581	9.0
	Indiana	1,449	10.6	1,391	6.8	1,374	7.2
	Missouri	5,290	6.6	5,455	4.2	4,925	5.5
	North Carolina	860	10.1	895	8.1	701	10.2
	Ohio	1,951	8.9	1,855	7.5	1,766	8.0
	Five states	11,232	4.1	11,276	2.8	10,346	3.5

Table A-2-- Estimates and relative errors for number of farms for selected estimators for the five states, using 1982 JES data

State	Estimator					
	Farm		Operational weighted		Cropland weighted	
	Est.	R.E.	Est.	R.E.	Est.	R.E.
	(000)	(%)	(000)	(%)	(000)	(%)
Georgia	52	6.5	60	4.9	52	5.9
Indiana	76	5.9	81	5.0	80	5.5
Missouri	107	5.2	123	4.2	114	4.7
North Carolina	79	6.6	94	4.9	80	5.6
Ohio	85	5.5	92	4.0	88	4.4
Five states	398	2.7	450	2.1	414	2.3

Table A-3-- Estimates and relative errors for selected variables and estimators for the five states, using 1982 JES data

Variable	State	Estimator					
		Farm		Operational Weighted		Cropland Weighted	
		Est.	R.E.	Est.	R.E.	Est.	R.E.
		(000)	(%)	(000)	(%)	(000)	(%)
Expected farrowings	Georgia	96	24.2	122	15.1	103	15.2
	Indiana	289	18.7	358	10.2	337	11.3
	Missouri	540	19.1	437	10.2	424	13.0
	North Carolina	65	22.3	116	19.7	106	25.7
	Ohio	160	31.8	197	18.4	198	22.2
	Five states	1,149	11.3	1,230	6.0	1,167	7.4
Previous farrowings	Georgia	98	22.7	122	15.6	106	16.8
	Indiana	267	18.3	336	10.7	322	11.9
	Missouri	445	19.7	378	10.7	365	14.1
	North Carolina	48	23.1	95	22.3	88	29.8
	Ohio	120	31.3	161	18.0	152	20.7
	Five states	978	11.2	1,093	6.2	1,034	7.6
Cows remaining to calve in 1982	Georgia	260	14.4	277	8.6	239	11.0
	Indiana	221	15.1	213	11.4	224	12.7
	Missouri	582	10.7	697	6.4	575	7.1
	North Carolina	135	17.8	161	10.9	125	13.7
	Ohio	397	14.5	390	10.7	397	12.2
	Five states	1,596	6.4	1,738	4.1	1,560	4.9
Calves born since 1/1/82	Georgia	462	12.9	546	7.3	467	10.0
	Indiana	348	14.1	336	8.5	328	9.0
	Missouri	1,552	10.4	1,709	4.6	1,442	6.1
	North Carolina	217	16.2	253	9.8	184	12.9
	Ohio	437	11.6	442	8.3	432	10.2
	Five states	3,016	6.3	3,285	3.1	2,853	4.0

SECTION D — ACRES OPERATED

Refer to Face Page for Type of Operation

Individually .... ☐ } *Go to item 1.*
Partnership or Joint .. ☐ }

Managed Land ...... ☐ *Go to item* ⑤

1. Now I would like to ask you about the **total acres you operate** under this land arrangement. Include all cropland, woodland, pastureland and wasteland.

How many acres do you:

a. **Own?** .... 901 [.]

b. **Rent from others?** *(Exclude all land used on an AUM basis)* .... 902 [.]

c. **Rent to others?** .... 905 [.]

Then the **total land you operate is** *(items a + b — c)*................. 900 [.]

Office Use 908 [ ]

Does this include woodland, pastureland and wasteland?

☐ **YES** · *Continue.* ☐ **NO** — *Make corrections and then continue.*

3. Next I would like to know how many **cropland acres** you operate. Cropland includes land planted and to be planted to crops during 1982, cropland to be used only for pasture, hay acreage, idle cropland and land in summer fallow.

Of the ________.______ total acres you operate, how many are cropland
(item 1 or 5)
acres when you exclude waste, woods, roads and ditches in the cropland fields?.. 903 [.]

4. My next question is about the **corn** and **soybean** acres you operate, excluding waste, woods, roads and ditches in the fields.

Of the ________.______ total acres you operate, how many acres,
(item 1 or 5)
excluding field waste, are planted or will be planted this year to:

a. **corn?** *(for all purposes)* .... 906 [.]

b. **soybeans?** *(for all purposes)* .... 907 [.]

(Go to Section E)

⑤ Now I would like to ask you about the **total acres you operate** as a hired manager.

How many acres do you operate as a hired manager? .... 904 [.]

Does this include woodland, pastureland and wasteland?

☐ **YES** · *Complete items 3 and 4 above.* ☐ **NO** · *Make corrections and then complete items 3 and 4 above.*

APPENDIX C

This appendix presents the formulas for the estimated totals for the four estimators discussed earlier. For each estimated total, it also gives the formula for the estimated variance. $\hat{Y}$ represents the estimated total and $\hat{var}(\hat{Y})$ is the estimated variance. These are the same formulas used by Nealon (4).

(1) Tract estimator:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk},$$

where

S = number of land use strata in the state,

P_i = number of paper strata within land use stratum i,

r_{ij} = number of segments within paper stratum j within land use stratum i,

e_{ijk} = expansion factor for segment k in paper stratum j within land use stratum i,

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{f_{ijk}} t_{ijk\ell} & \text{if } f_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

f_{ijk} = number of agricultural tracts in segment k within paper stratum j within land use stratum i,

$t_{ijk\ell}$ = tract value of the variable of interest for tract ℓ within segment k within paper stratum j within land use stratum i, and

$Y'_{ijk} = e_{ijk} Y_{ijk}$.

$$\hat{\text{var}}(\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij.}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \left\{ Y'_{ijk} - \bar{Y}'_{ij.} \right\}^2$$

where

$$\bar{Y}'_{ij.} = \sum_{k=1}^{r_{ij}} \frac{Y'_{ijk}}{r_{ij}} \quad \text{and} \quad \bar{e}_{ij.} = \sum_{k=1}^{r_{ij}} \frac{e_{ijk}}{r_{ij}} .$$

(2) Farm estimator:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk},$$

where

S, P_i, r_{ij}, and e_{ijk} are defined as before, and

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{g_{ijk}} Y_{ijk\ell} & \text{if } g_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

g_{ijk} = number of resident agricultural operators(RAO's) within segment k within paper stratum j within land use stratum i,

$$Y_{ijk\ell} = \begin{cases} \text{entire farm value of the variable of interest for tract } \ell \text{ within segment } k \text{ within paper stratum } j \text{ within land use stratum } i\text{, if the operator of tract } \ell \text{ is an RAO,} \\ 0 \quad \text{otherwise.} \end{cases}$$

$$Y'_{ijk} = e_{ijk} Y_{ijk} \; .$$

$$\hat{var}(\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij.}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \left\{ Y'_{ijk} - \bar{Y}'_{ij.} \right\}^2 ,$$

the same as for the tract estimator.

(3) Operational weighted estimator:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk} ,$$

where S, P_i, r_{ij}, and e_{ijk} are defined as before, and

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{f_{ijk}} a_{ijk\ell} \; Z_{ijk\ell} & \text{if } f_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where f_{ijk} is defined as before, and

$Z_{ijk\ell}$ = entire farm value of the variable of interest for tract ℓ within segment k within paper stratum j within land use stratum i, and

$a_{ijk\ell}$ = the weight for tract ℓ within segment k within paper stratum j within land use stratum i. The weight for each tract is always defined and is equal to the ratio of tract acreage to entire farm acreage.

$$Y'_{ijk} = e_{ijk} Y_{ijk} .$$

$$\hat{\text{var}} (\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{(1- \frac{1}{e_{ij\cdot}})}{(1- \frac{1}{r_{ij}})} \left\{ Y'_{ijk} - \bar{Y}'_{ij\cdot} \right\}^2 ,$$

the same as for the other estimators.

(4) Cropland weighted estimator:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk},$$

where

S, P_i, r_{ij}, and e_{ijk} are defined as before, and

$$Y_{ijk} = \begin{cases} \sum_{\ell}^{f_{ijk}} W_{ijk\ell} & \text{if } f_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$W_{ijk\ell} = \begin{cases} c_{ijk\ell} Z_{ijk\ell} & \text{if } c_{ijk\ell} \text{ is defined,} \\ t_{ijk\ell} & \text{if } c_{ijk\ell} \text{ is undefined and the tract approach is substituted,} \\ Y_{ijk\ell} & \text{if } c_{ijk\ell} \text{ is undefined and the farm approach is substituted.} \end{cases}$$

where $Y_{ijk\ell}$, $Z_{ijk\ell}$ and $t_{ijk\ell}$ are defined as before, and $c_{ijk\ell}$ = the weight for tract ℓ within segment k within paper stratum j within land use stratum i. The weight is equal to the ratio of tract cropland to entire farm cropland, provided there is cropland reported for the entire farm. In case the entire farm cropland value is missing or zero, the weight is undefined.

$$Y'_{ijk} = e_{ijk} Y_{ijk} \quad .$$

$$\hat{var}(\hat{Y}) \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{(1 - \frac{1}{\bar{e}_{ij\cdot}})}{(1 - \frac{1}{r_{ij}})} \left\{ Y'_{ijk} - \bar{Y}'_{ij\cdot} \right\}^2 ,$$

the same as for the other estimators.

APPENDIX D

This report calculated expansions for the defined domains of the cropland and operational weighted estimators for the following variables: number of farms, total hogs and pigs, total cattle and calves, total corn and soybean acres, entire farm acreage, and entire farm cropland acreage. The defined domain of the cropland weighted estimator was all 1982 JES tracts excluding those with zero or missing values for entire farm cropland acreage. The defined domain of the operational weighted estimator was defined as all tracts excluding those with imputed values for entire farm acreage. For the variable total corn (soybean) acres, the defined domain of the cropland weighted estimator also excluded all tracts with missing values for entire farm corn (soybean) acres.

This appendix presents the formulas for the estimated totals for the tract, farm, and weighted expansions for the defined domains of the weighted estimators. It also presents the formula for the estimated variance for each expansion. These formulas are quite similar to the ones shown in Appendix C.

(1) Tract Estimator:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y_{ijk}' = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk}$$

where

S=number of land use strata in the state,

P_i=number of paper strata within land use stratum i,

r_{ij}=number of segments within paper stratum j within land use stratum i,

e_{ijk}=expansion factor for segment k in paper stratum j within land use stratum i,

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{f_{ijk}} t_{ijk\ell} & \text{if } f_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

f_{ijk}=number of agricultural tracts in segment k within paper stratum j within land use stratum i,

$$t_{ijk\ell} = \begin{cases} \text{tract value of the variable of interest for tract } \ell \text{ within segment k within paper stratum j within land use stratum i, provided tract } \ell \text{ belongs to the defined domain of each weighted estimator,} \\ 0 \text{ otherwise,} \end{cases}$$

$$Y_{ijk}' = e_{ijk} Y_{ijk} \ .$$

$$\hat{var}(\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{(1-\frac{1}{\bar{e}_{ij.}})}{(1-\frac{1}{r_{ij}})} \left\{ Y_{ijk}' - \bar{Y}_{ij.}' \right\}^2 ,$$

where

$$\bar{Y}_{ij.}' = \sum_{k=1}^{r_{ij}} \frac{Y_{ijk}'}{r_{ij}} \text{ and } \bar{e}_{ij.} = \sum_{k=1}^{r_{ij}} \frac{e_{ijk}}{r_{ij}} \ .$$

(2) <u>Farm estimator</u>:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y_{ijk}' = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk},$$

where

S, P_i, r_{ij}, and e_{ijk} are defined as before, and

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{g_{ijk}} Y_{ijk\ell} & \text{if } g_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

g_{ijk} = number of resident agricultural operators (RAO's) within segment k within paper stratum j within land use stratum i,

$$Y_{ijk\ell} = \begin{cases} \text{entire farm value of the variable of interest for tract } \ell \text{ within segment k within paper stratum j within land use stratum i, provided tract } \ell \text{ belongs to the defined domain of each weighted estimator and the operator of tract } \ell \text{ is an RAO,} \\ 0 \text{ otherwise.} \end{cases}$$

$$Y'_{ijk} = e_{ijk} Y_{ijk} .$$

$$\hat{\text{var}}(\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij.}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \left\{ Y'_{ijk} - \bar{Y}'_{ij.} \right\}^2 ,$$

the same as for the other estimators.

(3) <u>Operational weighted estimator:</u>

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk},$$

where S, P_i, r_{ij}, and e_{ij} are defined as above, and

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{f_{ijk}} a_{ijk\ell} Z_{ijk\ell} & \text{if } f_{ijk} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where f_{ijk} is defined as before, and

$Z_{ijk\ell} =$ entire farm value of the variable of interest for tract ℓ within segment k within paper stratum j within land use stratum i, provided tract ℓ belongs to the defined domain of each weighted estimator,

0 otherwise,

and

$a_{ijk\ell}$ = the weight for tract ℓ within segment k within paper stratum j within land use stratum i.

$$Y'_{ijk} = e_{ijk} Y_{ijk} \quad .$$

$$\hat{\text{var}}\,(\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij.}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \left\{ Y'_{ijk} - \bar{Y}'_{ij.} \right\}^2 \quad ,$$

the same as for the other estimators.

(4) Cropland weighted estimator:

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk},$$

where S, P_i, r_{ij}, and e_{ijk} are defined as before, and

$$Y_{ijk} = \begin{cases} \sum_{\ell=1}^{f_{ijk}} C_{ijk\ell} \, Z_{ijk\ell} & \text{if } f_{ijk} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where

$Z_{ijk\ell}$ is defined as on the previous page, and

$c_{ijk\ell}$ = the weight for tract ℓ within segment k within paper stratum j within land use stratum i. The weight is equal to the ratio of tract cropland to entire farm cropland, provided there is cropland reported for the entire farm.

$$Y'_{ijk} = e_{ijk} Y_{ijk} \quad .$$

$$\hat{\text{var}}\,(\hat{Y}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{(1 - \frac{1}{\bar{e}_{ij}})}{(1 - \frac{1}{r_{ij}})} \left\{ Y'_{ijk} - \bar{Y}'_{ij.} \right\}^2 ,$$

the same as for the other estimators.

APPENDIX E

This appendix explains how the univariate and multivariate test statistics were calculated.

The analysis used paired t-tests to calculate the univariate test statistics.

Suppose $\hat{Y}$ and $\hat{Z}$ are estimated totals for a particular item of interest, using two different estimators. Suppose

$$\hat{Y} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Y'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk} \quad \text{and}$$

$$\hat{Z} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} Z'_{ijk} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Z_{ijk} ,$$

where

S = number of land use strata in the state,

P_i = number of paper strata within land use stratum i,

r_{ij} = number of segments within paper stratum j within land use stratum i,

e_{ijk} = expansion factor for segment k in paper stratum j within land use stratum i,

Y_{ijk} = value of the item of interest for segment k within paper stratum j within land use stratum i using one estimator,

Z_{ijk} = value of the item of interest for segment k within paper stratum j within land use stratum i using a different estimator,

$Y'_{ijk} = e_{ijk} Y_{ijk}$ and

$Z'_{ijk} = e_{ijk} Z_{ijk}$.

Let $\hat{D} = \hat{Y} - \hat{Z}$ be the difference between the estimated totals.

Then,

$$\hat{D} = \hat{Y} - \hat{Z} = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Y_{ijk} - \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} Z_{ijk}$$

$$= \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} (Y_{ijk} - Z_{ijk})$$

$$= \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} e_{ijk} d_{ijk}, \text{ where}$$

$d_{ijk} = Y_{ijk} - Z_{ijk}$.

$$\hat{var}(\hat{D}) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij.}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \left\{ d'_{ijk} - \bar{d}'_{ij.} \right\}^2 ,$$

where

$d'_{ijk} = e_{ijk} d_{ijk}$ and

$$\bar{d}'_{ij.} = \sum_{k=1}^{r_{ij}} \frac{d'_{ijk}}{r_{ij}} .$$

If $D = Y - Z$ is the population difference between the totals using estimators Y and Z, then to test H_o: $D=0$ vs. H_A: $D \neq 0$, compute

$$t = \frac{\hat{D}}{\hat{var}(\hat{D})}$$ and reject H_o if t is too large in absolute value.

Tabular values of t exist in most statistical references. The multivariate tests are generalizations of the univariate tests. This analysis used Hotelling's multivariate test (6).

Suppose one calculates $\hat{Y}$ and $\hat{Z}$ as above for q items of interest, using the same two estimators each time. Let $\hat{D}_1$, $\hat{D}_2, \ldots, \hat{D}_q$ be the differences $\hat{Y} - \hat{Z}$ for the q items of interest. Form the q x 1 column vector of differences $\hat{D} = (\hat{D}_1, \hat{D}_2, \ldots \hat{D}_q)^T$. Let W be the variance - covariance matrix of D, where the variance estimates for $\hat{D}$ lie on the main diagonal and the covariance estimates are the off-diagonal entries. Specifically,

$$\hat{var}(\hat{D}_\ell) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij.}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \left\{ d'_{\ell(ijk)} - \bar{d}'_{\ell(ij.)} \right\}^2 , \quad \ell = 1, 2, \ldots, q$$

where

S, P_i, r_{ij}, and e_{ijk} are defined as before.

$$\text{cov}\,(\hat{D}_{\ell}, \hat{D}_m) = \sum_{i=1}^{S} \sum_{j=1}^{P_i} \sum_{k=1}^{r_{ij}} \frac{\left(1 - \frac{1}{\bar{e}_{ij\cdot}}\right)}{\left(1 - \frac{1}{r_{ij}}\right)} \times$$

$$\left\{ d'_{\ell\,(ijk)} - \bar{d}'_{\ell\,(ij.)} \right\} \left\{ d'_{m(ijk)} - \bar{d}'_{m(ij.)} \right\} .$$

If W_{ij} is the entry in row i and column j in the matrix W then $W_{ii} = \text{var}\,(\hat{D}_i)$, i=1, 2, ..., q and

$$W_{ij} = W_{ji} = \hat{\text{cov}}\;(\hat{D}_i, \hat{D}_j)$$

i=1, 2, ..., q; j = 1, 2, ..., q, $i \neq j$.

Thus, W is a q x q symmetric matrix.

To test H_o: D is a zero vector us. H_A: at least one component of D is non-zero, compute

$$t^2 = \hat{D}^T W^{-1} \hat{D}.$$

Let $F = \left(\frac{r.. - P. - q + 1}{(r.. - P.)\, q} \right) t^2,$

where $r.. = \sum_{i=1}^{S} \sum_{j=1}^{P_i} r_{ij}$ is the number of segments in the state and $P. = \sum_{i=1}^{S} P_i$ is the number of paper strata. Then F is distributed as an F - statistic with degrees of freedom equal to (q, r.. - P. - q + 1). Reject H_o if F exceeds the tabular value of F. Tabular values of F exist in many statistical references. In case q=1, Hotelling's test reduces to the paired t-test explained earlier.

APPENDIX F

This appendix presents the formulas for the significance tests used in Table 7 when comparing the tract ratio of cropland to total land with the farm ratio.

Suppose $\hat{x}_1$ is the tract expansion for cropland and $\hat{y}_1$ is the tract expansion for total land. Let $\hat{x}_2$ and $\hat{y}_2$ be the corresponding farm expansions.

Let $\hat{r}_1 = \frac{\hat{x}_1}{\hat{y}_1}$ and $\hat{r}_2 = \frac{\hat{x}_2}{\hat{y}_2}$.

Then, according to Hansen, Hurwitz, and Madow (2), the estimated variance of $\hat{r}_1$ is given by

$$\hat{var}(\hat{r}_1) = \frac{1}{\hat{y}_1^2} \sum_{i=1}^{S} \sum_{j=1}^{P_i} \bar{e}_{ij.}\, r_{ij}\, (\bar{e}_{ij.} - 1)\, s_{ijz}^2 ,$$

where

S=number of land use strata in the state,

P_i= number of paper strata within land use stratum i,

r_{ij}= number of segments within paper stratum j within land use stratum i,

$\bar{e}_{ij.}$= the average expansion factor for paper stratum j within land use stratum i,

and $s_{ijz}^2 = s_{ijx_1}^2 + \hat{r}_1^2\, s_{ijy_1}^2 - 2\hat{r}_1\, s_{ijx_1y_1}$,

$$s_{ijx_1}^2 = \frac{1}{r_{ij} - 1} \sum_{k=1}^{r_{ij}} (X_{1ijk} - \bar{X}_{1ij.})^2 ,$$

where $X_{1ijk} = \sum_{\ell} X_{1ijk\ell}$ and $\bar{X}_{1ij.} = \frac{1}{r_{ij}} \sum_{k=1}^{r_{ij}} X_{1ijk}$, where $X_{1ijk\ell}$ is the value of the variable x_1 for tract ℓ within segment k within paper stratum j within land use stratum i.

$s_{ijy_1}^{\;2}$ is defined similarly, and

$$s_{ijx_1y_1} = \frac{1}{r_{ij}-1} \sum_{k=1}^{r_{ij}} (X_{1ijk} - \bar{X}_{1ij.})(Y_{1ijk} - \bar{Y}_{1ij.}).$$

$\hat{var}(\hat{r}_2)$ is defined similarly.

$$\hat{var}(\hat{r}_1-\hat{r}_2)=\hat{var}(\hat{r}_1) + \hat{var}(\hat{r}_2) - 2\,\hat{cov}(\hat{r}_1,\hat{r}_2).$$

$$\hat{cov}(\hat{r}_1,\hat{r}_2)= \frac{1}{\hat{y}_1\hat{y}_2} \sum_{i=1}^{S} \sum_{j=1}^{P_i} \bar{e}_{ij.}\, r_{ij}\, (\bar{e}_{ij.}-1)\, s_{ijz'}^{2},$$

where

$$s_{ijz'}^{2}= \frac{1}{r_{ij}-1} \sum_{k=1}^{r_{ij}} \Big[(X_{1ijk} - \bar{X}_{1ij.})(X_{2ijk} - \bar{X}_{2ij.})$$
$$+ \hat{r}_1\hat{r}_2 (Y_{1ijk} - \bar{Y}_{1ij.})(Y_{2ijk} - \bar{Y}_{2ij.})$$
$$- \hat{r}_1 (X_{2ijk} - \bar{X}_{2ij.})(Y_{1ijk} - \bar{Y}_{1ij.})$$
$$- \hat{r}_2 (X_{1ijk} - \bar{X}_{1ij.})(Y_{2ijk} - \bar{Y}_{2ij.})\Big].$$

To test $H_o: r_1 - r_2 = o$ vs. $H_A: r_1 - r_2 \neq o$, compute

$$t= \frac{\hat{r_1-r_2}}{\hat{var}(\hat{r_1-r_2})}$$

and reject H_o if t is too large in absolute value.

Tabular values of t exist in most statistical references.

☆U.S. Government Printing Office 1983 - 380-931/315