United States
Department of
Agriculture

Statistical
Reporting
Service

**Research
Division**

**Staff Report
Number AGES 820309**

**Washington, D.C.
20250**

March 1982

Improving Grain Stocks Estimates from the June Enumerative Survey

Jack Nealon

IMPROVING GRAIN STOCKS ESTIMATES FROM THE JUNE ENUMERATIVE SURVEY. By Jack Nealon; Statistical Research Division; Statistical Reporting Service; U.S. Department of Agriculture; Washington, D.C. 20250; March, 1982. SRS Staff Report No. AGES820309.

ABSTRACT

An alternative data imputation method and four weighted estimators were evaluated for obtaining grain stocks estimates from the June Enumerative Survey. The alternative imputation method is recommended over the operational imputation method. Weighted estimators using total land or cropland as the weight produced large gains in precision over the operational estimator. Therefore, either of these weighted estimators would be beneficial to the estimation of grain stocks.

Keywords: Refined and zero imputation methods, open and weighted estimators.

CONTENTS

SUMMARY

The refined imputation method is recommended over the zero imputation method to impute for missing production and stocks data. The refined method classifies each nonrespondent into one of three categories and then imputes a mean value from the same stratum and replicate depending upon the category. The zero method imputes zeroes for the missing entries. The ratio estimates of stocks to production were generally similar for the two methods in the three states tested. However, only the refined method has the potential of supplying useful stocks estimates at a multi-state level.

Ratio estimates of grain stocks to production from the June Enumerative Survey (JES) were approximately twice as precise when weighted estimators based on total land or cropland were used rather than the operational open estimator. Therefore, either of these two weighted estimators would be advantageous to the estimation of grain stocks. However, neither weighted estimator will be recommended until additional, detailed research is completed on these two weights during the 1982 JES.

Weighted estimates based on the number of grain bins or total grain bin capacity were not as precise as the weighted estimates using total land or cropland. Therefore, these weighted estimators are not recommended.

Improving Grain Stocks Estimates from the June Enumerative Survey

Jack Nealon

INTRODUCTION

The Statistical Reporting Service (SRS) publishes quarterly estimates of on-farm grain stocks. The data used to generate these estimates are primarily obtained from nonprobability mail surveys. For the June grain stocks report, area frame estimates of stocks and the ratio of stocks to production are available for selected crops in fourteen states from the JES in addition to the quarterly mail survey. Twelve of the fourteen states are from the North Central Region of the country.

The area frame estimates have not become an integral part of the grain stocks estimation process for three reasons. First of all, the estimates are often imprecise at the state level. Secondly, high nonresponse rates in many states result in a considerable amount of missing data for production and stocks. Finally, the operational method used to impute for missing data has never been thoroughly evaluated.

Research was conducted in Minnesota, North Carolina and Ohio during the 1981 JES to address the problem of imprecise estimates and to evaluate the operational imputation technique. The issue of decreasing the amount of nonresponse was not part of this study.

The Survey Research Section decided to evaluate four weighted estimators that could potentially yield large increases in precision over the present JES estimator. In addition, a computer imputation method called the refined method, which is more appealing than the current imputation method, was explored. Analyses concentrated on the ratio estimates of stocks to production rather than the stocks estimates since the ratio estimates are of major interest to the commodity statisticians in SRS.

CURRENT ESTIMATION PROCEDURE

The open estimator, which is also referred to as the farm expansion, is presently used by SRS to provide estimates of stocks and ratios of stocks to production from the JES. This estimator requires the enumerator to obtain production and stocks information for the entire farm from each farm operation with headquarters in the sampled segments.

When confronted with missing data, the current imputation method requires the survey statistician to perform one of the following two actions:

(1) If data for either production or stocks is missing for a commodity (but not both), the statistician must enter a number for the missing item,
(2) If both production and stocks data are missing for a crop, the entries are left blank.

An objective procedure is recommended in the editing instructions to the survey statisticians on how to impute a number when only one entry is missing for a crop. When both entries are missing, the statistician is in effect imputing zeroes for production and stocks since a blank entry is synonymous with a zero entry. Zero imputations will bias the stocks estimates downward if any of the nonrespondents actually have stocks. However, the effect of imputing zeroes on the estimate of major interest -- the ratio of stocks to production -- may not be serious.

Imputation of zeroes is used more by the survey statistician than the objective procedure since both entries are generally missing rather than just one of the entries. This assertion is supported by data collected during the 1979 JES in the North Central states. The 1979 JES data showed that both entries were missing for one or more of the crops in 11.1 percent of the operations while only one entry was missing for 2.8 percent of the operations.

RESEARCH ESTIMATION PROCEDURE

The results from four weighted estimators will be compared with the open estimator. Data collection expenses are slightly higher with a weighted estimator because production and stocks data are obtained from all farm operations with land in the sampled segments rather than only farm operations with headquarters in the segments. Therefore, if only a modest improvement in precision is realized with a weighted estimator, this gain is probably nullified by the increase in data collection costs and respondent burden. However, a weighted estimator would be an improvement if there are large gains in precision and the weight does not introduce an unknown bias into the estimate.

Tract and entire farm data were collected on four weights in order to evaluate the merits of weighted estimators. The weights were: (1) total land, (2) total cropland,

(3) number of permanent grain bins, and (4) the total capacity of these grain bins.

The total land weight was evaluated because this weight is presently used to provide weighted estimates of hog and cattle inventories and the number of farms in numerous states. The cropland weight was tested because it is currently being considered as a replacement for the total land weight. Concerns have been raised about the quality of the cropland weight from the 1981 JES in Minnesota, North Carolina and Ohio. (3) Therefore, caution should be exercised when evaluating the estimates based on this weight. Finally, the bin and bin capacity weights were investigated to determine their usefulness for estimating the ratio of stocks to production.

Comparisons of the imputation methods will focus solely on situations where both production and stocks data were missing. The computer imputation method being explored will be referred to as the refined method. The operational imputation method will be called the zero method since zeroes are entered for the missing entries.

The refined method is patterned after the imputation method recommended for livestock surveys by the Survey Research Section. (1) The mechanics of the refined method are:

(1) A nonrespondent is classified as a positive, zero or unknown nonrespondent for each production and stocks entry that is missing for an operation. Positive nonrespondents are operators who have the item, e.g. corn stocks, but refused to say how much they have. Zero nonrespondents are operators who do not have the item. Finally, unknown nonrespondents are operators who may or may not have the item.

(2) If the operator was classified as a positive nonrespondent, the average value based on respondents who have the item (positive respondents) from the same stratum and replicate was imputed. A value of zero was entered for zero nonrespondents, and the average value based on all respondents (positive and zero) from the same stratum and replicate was imputed for unknown nonrespondents. The imputations were applied to each replicate independently within a stratum so that unbiased variance estimates for production and stocks would be available. (2) The formulas for the refined method are given in Appendix A.

The classification process for each nonrespondent was based on the responses to two introductory questions asked prior to the questions on production and stocks data. These questions simply asked if each crop was produced the previous year and if each crop was being stored at the time of the interview. The introductory questions (questions 1 and 2) are shown in Appendix B. The refined method can distinguish between valid zeroes and missing data for production and stocks because of the introductory questions. The operational procedure does not have this benefit.

COMPARISON OF IMPUTATION METHODS

The survey statistician was instructed to enter an estimate if either production or stocks was missing for a crop. If both production and stocks were missing the data cells were to be coded with minus ones to signify nonresponses for research purposes. The JES summary system then converted the minus ones to zeroes so that the estimates were based on the zero imputation method.

The survey statistician entered a minus one rather than an estimate in many instances when only one of the entries was missing. These instances were converted to missing for both production and stocks since it was not feasible to guess the value that would have been entered. Overall this decision resulted in a 31 percent increase -- 2284 rather than 1744 -- in the instances classified as missing for both entries. <u>This situation can be avoided in the future if an edit check is incorporated in the Generalized Edit program to insure that a single entry cannot be missing</u>.

The percentage of times production and stocks data were missing or converted to be both missing is shown in Table 1 for each crop. Missing data occurred frequently in Minnesota and Ohio and at a higher rate for corn and soybeans than the small grains.

As mentioned previously, the stocks estimates from the zero method are biased downward if any nonrespondents have stocks. Due to the high rates of missing data in Minnesota and Ohio, the biases may be large in these states. However, bias in the ratio of stocks to production may very well not be large. On the other hand, bias in the stocks estimates from the refined method should not be nearly as severe since mean values are imputed for missing entries rather than zeroes. One would anticipate that the ratio estimates from the refined method would be preferred to the zero method since mean values rather than zeroes are imputed for production and stocks. For these reasons the refined method is intuitively a more appealing approach. The ratio and stocks estimates from the three

Table 1--Percentage of times production and stocks were missing or converted to be both missing for each crop.

State	Number of Farm Operations	Corn	Soybeans	Wheat	Oats	Barley
				%		
Minn.	1368	21.3	19.8	18.3	19.2	16.3
N.C.	1511	4.0	3.8	3.1	2.1	2.1
Ohio	1202	17.6	16.3	15.2	9.1	4.8
Three States Combined	4081	13.8	12.9	11.8	9.9	7.7

states combined were compared between the refined and zero methods to determine if the refined method is a worthy alternative.

The relative differences between the ratio estimates from the zero and refined methods are shown in Table 2 for the three states combined. Surprisingly, the relative differences were generally not large despite the contrasting imputation methods. The differences were always less than 3 percent for the open estimator, which was based on 1575 farm operations. The differences were generally less than 6 percent for the weighted estimators, which were based on 4081 farm operations. Overall, the relative differences were not consistently positive or negative.

Table 2--The relative difference between the ratio estimates from the two imputation methods for each estimator when the three states are combined.

Estimator	Relative Difference (%)				
	Corn	Soybeans	Wheat	Oats	Barley
Open	-2.3	-1.4	2.6	2.8	2.3
Weighted by:					
Total Land	-0.3	-6.1	-10.5	1.6	3.8
Cropland	0.7	-6.0	-10.1	2.5	4.5
Bins	3.1	-5.1	-12.0	-1.3	-15.5
Bin Capacity	3.4	-5.6	-12.7	-0.9	-14.4

$$\text{Relative Difference} = \frac{(\text{Zero-Refined})}{\text{Refined}} * 100$$

Table 3--The relative difference between the stocks estimates from the two imputation methods for each estimator when the three states are combined.

Estimator	Relative Difference (%)				
	Corn	Soybeans	Wheat	Oats	Barley
Open	-22.0	-24.5	-29.2	-19.2	-12.6
Weighted by:					
Total Land	-36.9	-37.7	-46.7	-26.4	-33.7
Cropland	-37.1	-38.6	-47.3	-26.8	-34.2
Bins	-36.7	-38.0	-41.8	-27.7	-37.2
Bin Capacity	-36.4	-38.5	-43.7	-27.4	-37.2

Shown in Table 3 are the relative differences between the stocks estimates from the two methods. The estimates from the zero method are always much below the refined method. An unknown part of this difference resulted from the survey statisticians not imputing a value when only stocks was missing. Analysis showed that the author's decision to convert cases to missing had a negligible effect on the stocks estimates from the two methods. For example, if the cases had not been converted, the relative differences between the stocks estimates shown in Table 3 for the estimator weighted by total land would have been unchanged for wheat and barley, -35.1 for corn, -36.6 for soybeans and -25.8 for oats. Therefore, regardless of the conversions the two imputation methods generated dissimilar stocks estimates.

To illustrate this fact the stocks estimates for the weighted estimator based on total land are presented in Table 4 for crops with a relative error less than 10 percent at the three state level. Also shown are the 99 percent confidence intervals for these estimates and the Crop Reporting Board's official estimates. The weighted estimator using total land was selected for this comparison because this estimator generally provided the smallest relative errors.

The stocks estimates from the zero method were much lower than the official estimates. The 99 percent confidence intervals did not even include the official estimates. The stocks estimates from the refined method were much closer to the official estimates for soybeans and oats and a little closer for corn. The 99 percent confidence intervals encompassed the official soybean and oats estimates but not the corn estimate. Although the stocks estimates from the refined method may not have much application at the state level, these estimates should be useful at a multi-state level such as the

Table 4--Stocks estimates using the total land weighted estimator and each imputation method for crops with relative errors less than 10 percent at the three state level.

Crop	Official Board Estimate	Zero Method		Refined Method	
		Stocks Est.	99% Confidence Interval	Stocks Est.	99% Confidence Interval
	(One Million Bushels)				
Corn	360	273	(233, 313)	433	(369, 497)
Soybeans	81	45	(38, 53)	73	(62, 84)
Oats	33	20	(16, 25)	28	(22, 34)

North Central Region for some of the crops. The usefulness of the stocks estimates will depend upon the precision of the estimates and the amount of imputing.

In summary, the refined method is preferred over the zero method. Despite the dissimilarity in the imputation methods, the ratio estimates were generally similar. To the contrary, the stocks estimates were much different for the two methods. The refined method has the potential of providing useful stocks estimates at a multi-state level. This is not feasible with the zero method.

COMPARISON OF ESTIMATORS

The survey statistician entered an estimate for total land and cropland whenever the tract and/or entire farm data were missing. On the other hand, data were not entered for missing bin and bin capacity entries since these imputation would have been very subjective. Therefore, the bin and bin capacity weights can be missing. Unlike the total land weight, the weights based on cropland, number of bins and bin capacity can be undefined. For example, if an operator has no cropland acres on the entire farm, the cropland weight is undefined. Whenever a weight was missing or undefined, the value of the total land weight was substituted so that the cropland, bin and bin capacity weighted estimates could be computed.

The ratio estimates of stocks to production from the open and four weighted estimators were compared when the refined imputation method was used. Results using the zero method were similar so will not be presented. The formulas for the five estimators are described in Appendix C.

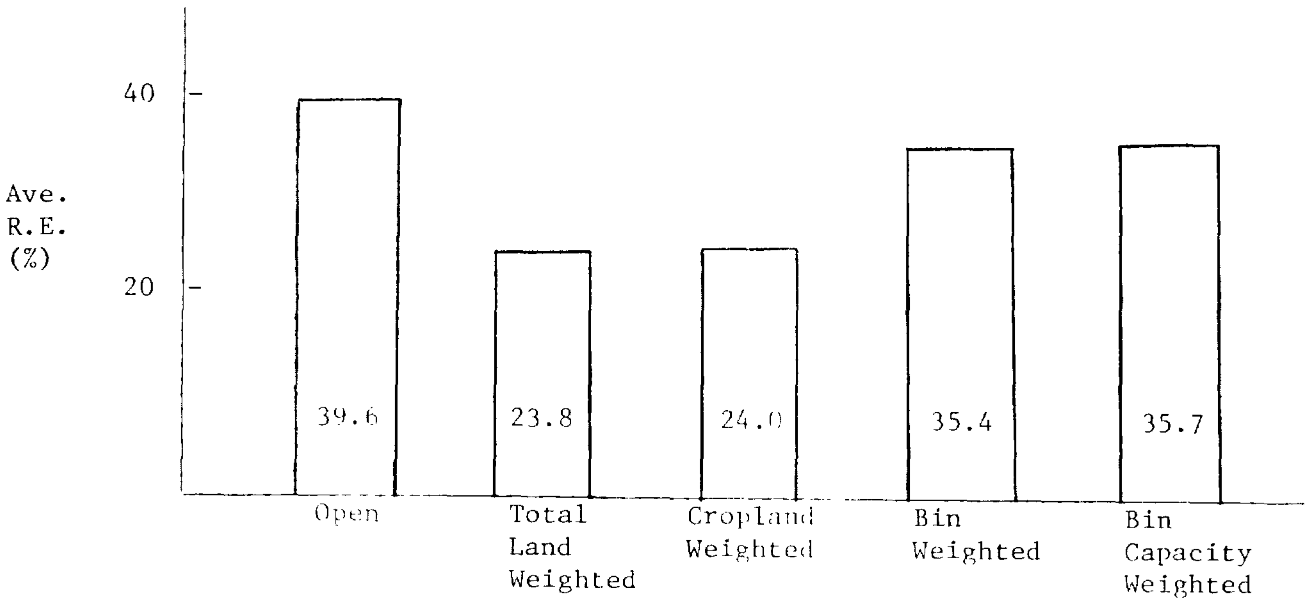
Table 5 gives the ratio estimates and relative errors for each crop and estimator. The relative error is the standard error of the estimate divided by the estimate. The ratio estimates and relative errors are presented in percentage terms. Figure 1 pictorially summarizes the results concerning the relative errors. Shown in this figure is the average relative error for each estimator based on all the relative errors in Table 5.

The bin and bin capacity weighted estimators were only slightly more precise than the open estimator overall. However, the weighted estimators based on total land and cropland showed large gains in precision over the open estimator. When the barley estimates in Ohio were excluded from the comparison, the total land and cropland weighted estimators were more than twice as precise as the open estimator. Pairwise paired

Table 5--The ratio of stocks to production and the relative error (R.E.) of this ratio for each crop and estimator in each state and the three states combined when the refined imputation method is used.

State	Estimator	Corn		Soybeans		Wheat		Oats		Barley	
		Ratio	R.E.	Ratio	R.E.	Ratio	R.E.	Ratio	R.E.	Ratio	R.E.
		%		%		%		%		%	
Minn.	Open	46.9	9.3	24.7	14.8	56.5	31.0	25.9	12.5	13.9	58.6
	Weighted by:										
	Total Land	43.3	5.4	30.0	7.9	44.8	12.0	26.6	5.9	36.4	26.4
	Cropland	43.3	5.3	30.3	8.0	44.7	11.9	26.6	5.9	35.0	26.4
	Bins	45.5	6.5	29.5	8.8	44.4	15.0	25.8	8.8	25.8	36.1
	Bin Capacity	45.1	6.2	29.1	9.2	43.2	15.5	25.3	9.2	27.1	35.5
N.C.	Open	11.5	31.5	17.7	29.4	0.5	98.3	16.0	50.2	1.4	113.6
	Weighted by:										
	Total Land	7.6	22.5	9.4	17.9	0.2	47.1	5.6	31.9	4.0	42.4
	Cropland	8.3	25.0	8.7	19.1	0.2	42.5	5.7	34.7	6.1	39.8
	Bins	6.6	45.1	10.6	27.5	0.1	88.8	2.2	90.4	1.5	103.3
	Bin Capacity	7.1	42.7	10.8	31.1	0.1	88.3	2.2	90.5	1.8	95.5
Ohio	Open	27.1	12.3	18.1	27.5	1.1	79.3	16.9	15.7	11.9	82.5
	Weighted by:										
	Total Land	19.2	7.6	15.5	11.9	1.4	32.4	14.2	14.3	1.1	135.9
	Cropland	19.3	7.9	15.5	12.0	1.3	35.1	13.8	15.9	1.0	136.7
	Bins	22.7	16.6	15.3	14.9	1.5	39.3	14.1	17.1	5.0	116.7
	Bin Capacity	22.8	16.1	15.6	16.0	1.7	31.9	14.0	17.8	1.3	134.9
Three States Com-bined	Open	35.2	8.1	21.3	13.2	42.1	35.2	24.6	11.3	12.9	57.4
	Weighted by:										
	Total Land	29.8	4.8	21.2	6.0	27.6	11.6	24.3	5.5	34.5	26.1
	Cropland	30.2	4.8	21.5	6.1	27.7	11.6	24.2	5.6	33.6	26.1
	Bins	32.2	7.1	21.4	7.2	30.0	14.7	23.4	8.4	23.9	35.8
	Bin Capacity	32.0	6.9	21.3	7.6	29.4	14.5	22.9	8.8	25.0	35.3

Figure 1--The average relative error across each state and crop for each estimator.



t-tests were calculated using the relative errors from each state and crop to determine if the relative errors were significantly different between estimators. The results were:

> The relative errors for the total land and cropland weighted estimators were not significantly different from one another, but were significantly smaller than the other three estimators.

The ratio estimates using the total land or cropland weighted estimator had high relative errors for all crops in North Carolina, for barley in each state and for wheat in Ohio. Sorghum grain estimates were not included in Table 5 because only a handful of operators in the three states produced or stored this crop. The high relative errors for certain crops were mainly the result of stocks being a rare item. This fact is illustrated in Table 6 which gives the percentage of responding farm operation in each state that did not have any stocks for each crop. The percentages were generally very high.

Table 6--The percentage of responding farm operations that had no stocks for each crop and state.

State	Corn	Soybeans	Wheat	Oats	Barley
			%		
Minn.	53.9	79.9	87.6	73.3	95.0
N.C.	85.5	91.5	99.4	98.7	99.1
Ohio	59.3	83.0	98.2	89.6	99.9
Three States Combined	68.4	85.5	95.4	88.4	98.1

A high relative error does not automatically imply that a ratio estimate is of little or no value. This is often true if the ratio estimate has a small value. For example, the ratio estimate from the total land weighted estimator in North Carolina was 7.6 percent with a relative error of 22.5 percent and a standard error of 1.7 percent. Despite the high relative error, the commodity statistician may find this ratio estimate very useful since the standard error is small. Therefore, some of the ratio estimates with high relative errors may still be beneficial for estimating grain stocks. However, state-level ratio estimates for some crops will continue to be of little value despite the increases in precision anticipated from weighted estimators. A multiple frame approach would be needed for certain crops in order to obtain ratio estimates that are useful at the state level.

Why were the estimates from the bin and bin capacity weighted estimators not as precise as the other two weighted estimators? Table 7 sheds light on this question. This table gives the distribution of the bin weight for each state and the three states combined. Notice that a considerable percentage of the bin weights have a value of zero or one. Recall that when the bin weight was undefined or missing the operational weight was substituted. Since the operational weight had a value of one on 17.5 percent of all reports, the bin weighted estimator actually had more zero and one values as its weight than are shown in Table 7. This kind of distribution for a weight yields weighted tract data that are quite variable because much of the farm data is either weighted to zero or given its entire farm value. If the weights are concentrated at zero and one, a weighted estimator behaves somewhat like an open estimator, which can be thought of as applying a weight of one to operations with headquarters in the segment and zero to operations with headquarters outside the segment. Therefore, the bin weight does not provide increases in precision nearly as large as the total land and cropland weights. The same is true for the bin capacity weight which is zero or one whenever the bin weight is zero or one.

Another problem with the bin and bin capacity weights is that they were often undefined. This occurrence, in conjunction with the unfavorable distributions of these weights when they were defined, lead to the recommendation to exclude the bin and bin capacity weighted estimators from further consideration.

The ratio estimates from the total land and cropland weighted estimators were generally very similar. The ratio estimates from the open estimator tended to be higher than the two weighted estimators particularly in North Carolina and Ohio. The difference in the weighted and open estimates may very well be caused by sampling variation and/or the imputation process. However, the possibility exists that the weighted ratio estimates may be biased. Plans are underway to examine the total land and cropland weights in more detail during the 1982 JES.

In summary, the weighted estimators based on total land and cropland were much more precise than the other three estimators. For open estimates that are marginally precise, e.g. 10 percent relative error, these weighted estimators should provide precise estimates.

Table 7--The distribution of the bin weight for each state.

State	Distribution of bin weight (%)				
	Zero	Between zero and one	One	Undefined	Missing
Minn.	33.5	5.4	17.4	27.6	16.1
N.C.	19.8	0.4	3.5	73.6	2.7
Ohio	31.5	2.8	11.7	44.1	9.9
Three States Combined	27.9	2.8	10.6	49.5	9.2

CONCLUSIONS AND RECOMMENDATIONS

The refined imputation method is recommended over the zero method that is currently used to impute for missing production and stocks data from the JES. Although the ratio estimates from these methods were generally similar, the refined method has the potential of providing useful stocks estimates at a multi-state level. Stocks estimates from the zero method are of little value.

Weighted estimators based on total land and cropland provided large gains in precision over the open estimator, bin weighted estimator and bin capacity weighted estimator. Therefore, the total land and cropland weighted estimators are preferred. Neither weight will be recommended until additional research on these weights during the 1982 JES is completed.

Analysis showed that imprecise state estimates will continue to exist for certain crops despite the proposed improvements. Estimates Division should either eliminate the crops and states where the benefits are negligible for the state and multi-state estimation program or implement a multiple frame approach to improve the precision of these estimates.

The editing instructions for the survey statisticians were not followed. There were 540 instances where a value was not imputed when a single entry was missing. Therefore, a computer edit check is needed to insure the proper use of the editing instructions.

The recommended changes will result in additional burden to the operational program. The major changes involve modifying the summary system to handle the refined imputation method, computing variances based on replicates, and collecting grain stocks data from all tract operations so that weighted estimates can be calculated. The impact of the recommendations needs to be assessed by Methods Staff.

REFERENCES

(1) Crank, Keith N. The Use of Current Partial Information to Adjust for Nonrespondents. Statistical Reporting Service, U.S. Department of Agriculture. April, 1979.

(2) Ford, Barry L., Kleweno, Douglas G. and Tortora, Robert D. A Simulation Study to Compare Procedures Which Impute for Missing Items on an ESS Hog Survey. Statistical Reporting Service, U.S. Department of Agriculture. December, 1980.

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

APPENDIX A

The formulas for the refined imputation method will be presented in this appendix. The refined method classifies each nonrespondent for a crop as a positive, zero or unknown nonrespondent for production and for stocks. A value is then imputed for each missing entry based on respondents in the same land use stratum and replicate who fall in the same classification. Let $x'_{ijk\ell}$ represent the missing stocks entry of a crop for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum. Then

$$x'_{ijk\ell} \begin{cases} = \bar{x}^{+}_{i\cdot k\cdot} & \text{if } x'_{ijk\ell} \text{ is a } \underline{\text{positive}} \text{ nonrespondent} \\ = 0 & \text{if } x'_{ijk\ell} \text{ is a } \underline{\text{zero}} \text{ nonrespondent} \\ = \bar{x}_{i\cdot k\cdot} & \text{if } x'_{ijk\ell} \text{ is an } \underline{\text{unknown}} \text{ nonrespondent} \end{cases}$$

where

$$\bar{x}^{+}_{i\cdot k\cdot} = \frac{\sum_{j=1}^{p_i} \sum_{\ell=1}^{m^{+}_{ijk}} e_{ijk} x_{ijk\ell}}{\sum_{j=1}^{p_i} m^{+}_{ijk} e_{ijk}}, \quad \bar{x}_{i\cdot k\cdot} = \frac{\sum_{j=1}^{p_i} \sum_{\ell=1}^{m_{ijk}} e_{ijk} x_{ijk\ell}}{\sum_{j=1}^{p_i} m_{ijk} e_{ijk}},$$

p_i = the number of paper strata in the i^{th} land use stratum,

e_{ijk} = the expansion factor for the k^{th} replicate in the j^{th} paper stratum and i^{th} land use stratum,

$x_{ijk\ell}$ = the entire farm stocks data reported for a commodity for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum,

$$m^{+}_{ijk} \begin{cases} = g^{+}_{ijk} & \text{if the open estimator is used,} \\ = f^{+}_{ijk} & \text{if the weighted estimator is used,} \end{cases}$$

$$m_{ijk} \begin{cases} = g_{ijk} & \text{if the open estimator is used,} \\ = f_{ijk} & \text{if the weighted estimator is used,} \end{cases}$$

g^{+}_{ijk} = the number of responding farm operations with headquarters in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum that were positive stocks respondents for the crop,

f^{+}_{ijk} = the number of responding farm operations with land in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum that were positive stocks respondents for the crop,

g_{ijk} = the number of responding farm operations with headquarters in replicate k, paper stratum j and land use stratum i,

f_{ijk} = the number of responding farm operations with land in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum.

The formulas are identical in form when imputing for $y'_{ijk\ell}$, the missing production entry for a crop. The only difference in terminology is that the g^{+}_{ijk} and f^{+}_{ijk} are based on positive production respondents rather than positive stocks respondents for the crop.

APPENDIX B

SECTION K – GRAIN PRODUCTION, STOCKS AND SALES

1. Did you produce any of the following grains in 1980?

Enter Code YES = 1 NO = 2

a. CORN	170
b. SORGHUM GRAIN	171
c. OATS, old crop	172
d. BARLEY, old crop	173
e. SOYBEANS	169
f. ALL WHEAT, old crop including DURUM	168

2. Are you storing any of the following grains on the total acres you operate? *(Include grain from 1980 and earlier years including grain for feed and seed.)*

Enter Code YES = 1 NO = 2

a. CORN	176
b. SORGHUM GRAIN	177
c. OATS, old crop	178
d. BARLEY, old crop	179
e. SOYBEANS	175
f. ALL WHEAT, old crop including DURUM	174

3. Now I would like to ask you about grain production and grain sold from the 1980 crop. Also, I need grain stored in bins, cribs, and granaries on the total acres you operate from both 1980 and earlier years' crops. Do not include any grain you own which is stored in commercial facilities such as local elevators and terminals.

	1	2	3	4
	FOR THE 1980 CROP ONLY			FOR 1980 AND EARLIER YEARS' CROP
CROP	How many bushels of the following grains were produced on the total acres you operated in 1980? *(Include landlord's share)*	Of the bushels produced in 1980 (Col. 1), how many bushels have been sold or will be sold?	Of the 1980 crop sold (Col. 2), how many bushels were sold directly to another rancher, farmer or feedlot operator?	How many bushels of grain are stored on the total acres you operate? *(Include grain for sale, for feed, and for seed. Exclude grain in commercial storage.)*
a. CORN .... 56 lb. shelled bushels	584	586	587	585
b. SORGHUM GRAIN .. 56 lb. bu. or Cwt.	588	590	591	589
c. OATS, old crop ..... 32 lb. bu.	592	594	595	593
d. BARLEY, old crop .. 48 lb. bu.	596	598	599	597
e. SOYBEANS ... . 60 lb. bu.	582			583
f. ALL WHEAT, old crop including DURUM .. 60 lb. bu.	580			581

	Column A On Total Acres Operated	Column B on Tract Acres
4. How many permanent bins *(excluding commercial bins)* that are used solely to store grain are located *(include bins now empty)*	164	165
5. What is the total capacity of these grain storage bins *(excluding commercial bins)*	166	167

APPENDIX C

The formulas at the state level for the area frame estimate of the ratio of stocks to production for a given crop will be presented in this appendix for each of the five estimators.

(1) OPEN ESTIMATOR:

The state estimate of total stocks can be expressed as:

$$\hat{X}_1 = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} \hat{x}_{ijk} = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} e_{ijk} x_{ijk},$$

where

$\hat{x}_{ijk}$ = the expanded data for stocks in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum,

e_{ijk} = the expansion factor for the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum,

r_{ij} = the number of sample replicates in the j^{th} paper stratum in the i^{th} land use stratum. In most states, $r_{ij} = r_i$,

p_i = the number of paper strata in the i^{th} land use stratum,

s = the number of land use strata in the state,

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

g_{ijk} = the number of farm operations with headquarters in replicate k, paper stratum j and land use stratum i, and

$x_{ijk\ell}$ = the stocks on the entire farm for a given commodity for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum.

The estimated variance of $\hat{X}_1$ expressed in terms of variability among sample replicates in a land use stratum can be very closely approximated by:

$$\hat{V}(\hat{X}_1) \doteq \sum_{i=1}^{s} \frac{r_i}{r_i - 1} \sum_{k=1}^{r_i} (\hat{x}_{i\cdot k\cdot} - \bar{\hat{x}}_{i\cdot\cdot})^2,$$

where $r_i = r_{ij}$,

$\hat{x}_{i\cdot k} = \sum_{j=1}^{p_i} \hat{x}_{ijk}$ is the expanded total for the k^{th} replicate in the i^{th} land use stratum,

$\hat{\bar{x}}_{i\cdot\cdot} = \sum_{k=1}^{r_i} \frac{\hat{x}_{i\cdot k}}{r_i}$ is the average expanded replicate total for the i^{th} land use stratum,

s and p_i are as previously defined.

The formula for the variance estimate is not an equality because the last paper stratum in each land use stratum often has a slightly different number of segments in the population. However, this difference is expected to have a negligible effect. The finite population correction factor is not included in the variance estimate because the sampling rates are very small within land use strata.

The state estimate of total production and the estimated variance of this estimate are identical in form to the formulas used for stocks. Using the label, Y_1, to denote production, it follows that:

$$\hat{Y}_1 = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} \hat{y}_{ijk} = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} e_{ijk}\, y_{ijk}$$

and

$$\hat{V}(\hat{Y}_1) \doteq \sum_{i=1}^{s} \frac{r_i}{r_i - 1} \sum_{k=1}^{r_i} (\hat{y}_{i\cdot k} - \hat{\bar{y}}_{i\cdot\cdot})^2 .$$

Finally, the estimate of interest, that is, the ratio of stocks to production using the open estimator is given by:

$$\hat{R}_1 = \hat{X}_1 / \hat{Y}_1 .$$

The estimated variance of $\hat{R}_1$ is:

$$\hat{V}(\hat{R}_1) \doteq \frac{1}{\hat{Y}_1^2} [\hat{V}(\hat{X}_1) + \hat{R}_1^2\, \hat{V}(\hat{Y}_1) - 2\hat{R}_1 \hat{C}ov\, (\hat{X}_1, \hat{Y}_1)],$$

where

$$\hat{Cov}(\hat{X}_1,\hat{Y}_1) \doteq \sum_{i=1}^{s} \frac{r_i}{r_i-1} \sum_{k=1}^{r_i} (\hat{x}_{i\cdot k} - \hat{\bar{x}}_{i\cdot\cdot})(\hat{y}_{i\cdot k} - \hat{\bar{y}}_{i\cdot\cdot}).$$

(2) TOTAL LAND WEIGHTED ESTIMATOR:

The sample estimate of the ratio of stocks to production, that is, $\hat{R}_2$, is given by:

$$\hat{R}_2 = \hat{X}_2/\hat{Y}_2,$$

where,

$$\hat{X}_2 = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} e_{ijk}x_{ijk},$$

$$\hat{Y}_2 = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} e_{ijk}y_{ijk},$$

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

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

s, p_i, r_{ij}, e_{ijk}, $x_{ijk\ell}$ and $y_{ijk\ell}$ are as previously defined,

f_{ijk} = the number of farm operations with land in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum,

$a_{ijk\ell}$ = the weight for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum. This weight is the ratio of the total acres in the tract to the total acres in the entire farm.

The estimated variance of $\hat{R}_2$ is:

$$\hat{V}(\hat{R}_2) \doteq \frac{1}{\hat{Y}_2^2} [\hat{V}(\hat{X}_2) + \hat{R}_2^2 \hat{V}(\hat{Y}_2) - 2\hat{R}_2 \hat{Cov}(\hat{X}_2, \hat{Y}_2)],$$

where the variance and covariance terms follow the same notation used for the open estimator.

(3) CROPLAND WEIGHTED ESTIMATOR:

The sample estimate of the ratio of stocks to production is:

$$\hat{R}_3 = \hat{X}_3 / \hat{Y}_3,$$

where

$$\hat{X}_3 = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} e_{ijk} x_{ijk},$$

$$\hat{Y}_3 = \sum_{i=1}^{s} \sum_{j=1}^{p_i} \sum_{k=1}^{r_{ij}} e_{ijk} y_{ijk},$$

$$x_{ijk} \begin{cases} = \sum_{\ell=1}^{f_{ijk}} b_{ijk\ell}\, x_{ijk\ell} \text{ if } f_{ijk} > 0, \\ = 0 \text{ if } f_{ijk} = 0, \end{cases}$$

$$y_{ijk} \begin{cases} = \sum_{\ell=1}^{f_{ijk}} b_{ijk\ell}\, y_{ijk\ell} \text{ if } f_{ijk} > 0, \\ = 0 \text{ if } f_{ijk} = 0, \end{cases}$$

$$b_{ijk\ell} \begin{cases} = c_{ijk\ell} \text{ if } c_{ijk\ell} \text{ is defined}, \\ = a_{ijk\ell} \text{ if } c_{ijk\ell} \text{ is not defined}, \end{cases}$$

$c_{ijk\ell}$ = the ratio of the tract cropland acres to the entire farm cropland acres for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum. If the entire farm cropland acreage is zero, $c_{ijk\ell}$ is undefined.

s, p_i, r_{ij}, e_{ijk}, f_{ijk}, $a_{ijk\ell}$, $x_{ijk\ell}$ and $y_{ijk\ell}$ are as previously defined.

The $\hat{V}(\hat{R}_3)$ can be expressed as:

$$\hat{V}(\hat{R}_3) = \frac{1}{\hat{Y}_3^2} [\hat{V}(\hat{X}_3) + \hat{R}_3^2 \hat{V}(\hat{Y}_3) - 2\hat{R}_3 \hat{Cov}(\hat{X}_3, \hat{Y}_3)],$$

where the variance and covariance terms use the same notation as the previous estimators.

(4) BIN WEIGHTED ESTIMATOR:

The formulas for this estimator are identical to the cropland weighted estimator except that $b_{ijk\ell}$ is defined differently. That is,

$$b_{ijk\ell} \begin{cases} = d_{ijk\ell} \text{ if } d_{ijk\ell} \text{ is defined,} \\ = a_{ijk\ell} \text{ if } d_{ijk\ell} \text{ is not defined.} \end{cases}$$

where

$a_{ijk\ell}$ is as previously defined, and

$d_{ijk\ell}$ = the ratio of the number of permanent grain storage bins in the tract to the number in the entire farm for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum.

(5) BIN CAPACITY WEIGHTED ESTIMATOR:

This estimator also differs from the cropland weighted estimator only in the definition of $b_{ijk\ell}$. The definition is:

$$b_{ijk\ell} \begin{cases} = h_{ijk\ell} \text{ if } h_{ijk\ell} \text{ is defined,} \\ = a_{ijk\ell} \text{ if } h_{ijk\ell} \text{ is not defined,} \end{cases}$$

where

$a_{ijk\ell}$ is defined as before, and

$h_{ijk\ell}$ = the ratio of the storage capacity of the permanent grain bins in the tract to the entire farm for the ℓ^{th} farm operation in the k^{th} replicate, j^{th} paper stratum and i^{th} land use stratum.

GOVERNMENT PRINTING OFFICE: 1982-360-931:SRS-506